Wavelet-Based Multiscale Window Transform and Energy and Vorticity Analysis

X. San Liang

May 19, 2002

Wavelet-Based Multiscale Window Transform and Energy and Vorticity Analysis

Abstract

A new methodology, *Multiscale Energy and Vorticity Analysis* (MS-EVA), is developed to investigate sub-mesoscale, meso-scale, and large-scale dynamical interactions in geophysical fluid flows which are intermittent in space and time. The development begins with the construction of a wavelet-based functional analysis tool, the *multiscale window transform* (MWT), which is local, orthonormal, self-similar, and windowed on scale. The MWT is first built over the real line then modified onto a finite domain. Properties are explored, the most important one being the property of marginalization which brings together a quadratic quantity in physical space with its phase space representation.

Based on MWT the MS-EVA is developed. Energy and enstrophy equations for the large-, meso-, and sub-meso-scale windows are derived and their terms interpreted. The processes thus represented are classified into four categories: transport, transfer, conversion, and dissipation/diffusion. The separation of transport from transfer is made possible with the introduction of the concept of *perfect transfer*. By the property of marginalization, the classical energetic analysis proves to be a particular case of the MS-EVA.

The MS-EVA developed is validated with classical instability problems. The validation is carried out through two steps. First, it is established that the barotropic and baroclinic instabilities are indicated by the spatial averages of certain transfer term interaction analyses. Then calculations of these indicators are made with an Eady model and a Kuo model. The results agree precisely with what is expected from their analytical solutions, and the energetics reproduced reveal a consistent and important aspect of the unknown dynamic structures of instability processes.

As an application, the MS-EVA is used to investigate the Iceland-Faeroe frontal (IFF) variability. A MS-EVA-ready dataset is first generated, through a forecasting study with the Harvard Ocean Prediction System using the data gathered during the 1993 NRV *Alliance* cruise. The application starts with a determination of the scale window bounds, which characterize a double-peak structure in either the time wavelet spectrum or the space wavelet spectrum. The resulting energetics, when locally averaged, reveal that there is a clear baroclinic instability happening around the cold tongue intrusion observed in the forecast. Moreover, an interaction analysis shows that the energy released by the instability indeed goes to the meso-scale window and fuel the growth of the intrusion. The sensitivity study shows that, in this case, the key to a successful application is a correct decomposition of the large-scale window from the meso-scale window.

Acknowledgements

I would like to take this opportunity to thank many people who have helped me during my graduate studies. First and foremost, I am deeply grateful to my advisor, Professor Allan R. Robinson. This work would not be possible without his scientific sponsorship and financial support. I have been shaped into what I am since he brought me under his wing. Professor Donald G. M. Anderson has spent numerous hours in guiding the development of the *multiscale window transform*, and helping me prepare for my academic career. From the challenging discussions with him I have learned how to formulate problems in a rigorous language. My other advisors, Dr. Kenneth Brink and Professor Brian Farrell, have carefully read this thesis and shared with me their insightful ideas on geophysical fluid dynamics. I wish to express my gratitude to Dr. Arthur Miller for agreeing to join my committee towards the end of this work, for his thoughtful comments, and for his final-minute help with the defense.

Members of the Harvard Ocean Group, past and present, were the key factors in the fulfillment of my graduate study. The discussions with Dr. Carlos Lozano have been very helpful in bridging abstract mathematics and physical oceanography. In fact, the term *multiscale window transform* was initially suggested by him. Mr. Wayne Leslie was always there when I needed a hand. His input in the Iceland-Faeroe Frontal variability forecasting is greatly acknowledged. Special thanks go to Drs. Patrick Haley, Jr., and Pierre Lermusiaux. They have shared my frustration and happiness, and offered numerous assistance during my thesis work. Besides, Pat has been coaching me in American English as well as the Harvard Ocean Prediction System (HOPS); Pierre's wit has helped me overcome many unexpected difficulties. I am also indebted to Professors Ching-Sang Chiu and Sukru Besiktepe, for their attention to my personal matters. I sincerely appreciate Ms. Gioia Sweetland's partnership, and her unique blending of caring and administration has made things neatly organized. Mrs. Peggy Zaldivar's kindness and humor are always impressive, and the friendship from Ms. Marsha Glass is sorely missed. The activities with Mr. Nick Dechman and Mr. Christian Casillas have made me fit into the American culture. I thank Dr. Quinn Sloan for familiarizing me with the facilities when I first came in the group. Dr. Dimitri Kroujiline's neat explanation on nonlinear mixing dynamics is very admirable, and Ms. Patricia Moreno's willingness to proofread the early version of this thesis is especially acknowledged. In formatting the chapters and many other documents, Dr. Lyon Lanerolle's template with LaTeX, and Dr. Jeff Dusenberry's installation of WordPerfect on the unix system have been of great help.

This work has also benefited from many people outside the group. I was first called to attention by Professor William Vetterling to wavelet analysis. Professor Gilbert Strang introduced to me the notion of multi-resolution analysis, on which many important concepts in this thesis are based. Dr. Joseph Pedlosky's critical thoughts have led to the whole section on how a transport may be separated from a transfer process. Professor Nadia Pinardi and Dr. Michael Spall, the two pioneers in EVA study, have given invaluable comments and suggestions about this thesis and the career future based on my work. I appreciate Dr. Rui Xin Huang's advice and friendship, and am particularly thankful to him for introducing to me opportunities in areas related to my interest. I also thank Dr. Joe LaCase, who spent time with me explaining the law of geostrophic turbulence over a slope. This work has benefited a lot from Dr. Ce Wang's insight on human wisdom and artificial intelligence. His help has made the thesis well prepared before the due date. I owe my debts to Professors Roger Brockett and Howard Stone, who afforded much time to help me in seeking for appropriate topics for my Ph.D. thesis. Dr. Fred Nicolier of the University of Burgundy, France, provided the English translation of a part of a French-written wavelet book, and the email exchange with Professor Hernan Arango was helpful in bringing to an end the real-time forecasting of the Iceland-Faeroe Frontal variability.

I am grateful to the Division of Engineering and Applied Sciences for allowing me to have a chance to expand my knowledge. I would like to acknowledge the support of the Ernst Habicht Fellowship, and the skillful administration of Ms. Susan M. Wieczorek. This thesis was supported in part by the Office of Naval Research under grants N00014-95-1-0371 to Harvard University, through the generosity of my advisor.

I am greatly indebted to Professor Kuo-Chuin Wong, for his understanding when I decided to make a switch. I will never forget the help I received during my early days in the States from Dr. Jianhua Ye, Dr. Yuguang Liu, Mr. Bill Oosterman, Dr. Yun He, to name but a few. The generosity and friendship from Dr. Xudong Jiang during my hard times is especially acknowledged.

Finally, I wish to express my most heartfelt thanks to my family, for their forbearance and support. My mother-in-law, Mrs. Shen Qinfen, traveled from China to take care of my daily life during September through December, 2001. The affectionate encouragement and unconditional moral backing from My wife, Feiyan Zhao, has helped me out at many critical moments. I thank my late grandmother, Mrs. Zeng Dechu, who had taught me all the basic principles as I grew up. I wish she could read these words.

Contents

List of Figures

List of Tables

Glossary of symbols and acronyms

$\mathbb{N}$: Set of natural numbers (positive integers)

$\mathbb{R}$: Field of real numbers

$\mathbb{T}_\varrho$: Torus with period ϱ

$\mathbb{Z}$: Ring of integers

$\mathbb{Z}_+$: Set of nonnegative integers

j_0: Upper bound of scale level for large-scale window

j_1: Upper bound of scale level for meso-scale window

j_2: The largest scale level for a given series

$p^{\sim 0}$: Large-scale part of p

$p^{\sim 1}$: Meso-scale part of p

$p^{\sim 2}$: Sub-mesoscale part of p

$\bar{p}$: Mean of p ($\int_0^1 p(t)\ dt$ for a scaled t)

p': $p - \bar{p}$

$\widehat{p}_n^{\sim 0}$: Large-scale (window) transform of p

$\widehat{p}_n^{\sim 1}$: Meso-scale (window) transform of p

$\widehat{p}_n^{\sim 2}$: Sub-mesoscale (window) transform of p

ζ: Vertical component of vorticity

$\underline{\zeta}^*$: Horizontal components of vorticity

ρ: Density anomaly

ϱ: Period of the boundary extension

ϕ: Scaling function

ϕ_n^j: $2^{j/2}\phi(2^j t - n)$

$\phi_n^{\varrho,j}$: Periodized ϕ_n^j with period ϱ

ψ: Wavelet function

ψ_n^j: $2^{j/2}\psi(2^j t - n)$

$\psi_n^{\varrho,j}$: Periodized ψ_n^j with period ϱ

Ψ: Streamfunction

A: Available potential energy $(\frac{1}{2}\frac{g^2}{\rho_0^2 N^2}\rho^2)$

b: Buoyancy conversion rate

BC: Transfer interaction analysis term that accounts for the baroclinic instability

BT: Transfer interaction analysis term that accounts for the barotropic instability

K: Kinetic energy $(\frac{1}{2}\underline{\mathbf{v}} \cdot \underline{\mathbf{v}})$

N^2: Buoyancy frequency squared

$\mathcal{N}_\varrho^j$: The set $\{0, 1, 2, ..., 2^j \varrho - 1\}$

$\mathcal{M}_n$: Operator of the (finite domain) marginalization

$\underline{\underline{\mathbf{M}}}$: dilation (sampling) matrix

P: Pressure

$\underline{\mathbf{v}}$: Horizontal components of velocity (u, v)

V_j: Multi-resolution approximation of $L_2(\mathbb{R})$

$V_{\varrho,j}$: Multi-resolution approximation of $L_2(\mathbb{T}_\varrho)$

$\mathcal{V}_j$: Voronoi cell with respect to dilation matrix $\underline{\underline{\mathbf{M}}}^j$

w: Vertical component of velocity (also window index)

Z: Enstrophy

ACC: Anomaly Correlation Coefficient

ADCP: Acoustic Doppler Current Profiler AIW: Arctic Intermediate Water

AVHRR: Advanced Very High Resolution Radiometry

AXBT: Airborne Expendable Bathythermograph

CFvN: Charney-Fjörtoft-von Neumann

CTD: Conductivity-Temperature-Depth

EICW: East Iceland and Current Water

EIW: East Icelandic Water

EOF: Empirical Orthogonal Function

EVA: Energy and Vorticity Analysis

GFD: Geophysical Fluid Dynamics

GIN Seas: Greenland Sea, Iceland Basin, and Norwegian Sea

HCF: Hydrostatic Consistency Factor

HOPS: Harvard Ocean Prediction System

IFF: Iceland-Faeroe Front

IFR: Iceland-Faeroe Ridge

MNAW: Modified North Atlantic Water

MS-EVA: Multi-scale Energy and Vorticity Analysis

NAW: North Atlantic Water

NNAW: Norwegian North Atlantic Water

NSDW: Norwegian Sea Deep Water

OA: Objective Analysis

OI: Optimal Interpolation

PCA: Principal Component Analysis

PE model: Primitive Equation model

QG model: Quasi-Geostrophic model

RMSE: Root-Mean-Square Error

SACLANTCEN: Supreme Allied Commander, Atlantic Undersea Research Center

SST: Sea Surface Temperature

XBT: Expendable Bathythermograph

XCTD: Expendable Conductivity-Temperature-Depth

Chapter 1

Introduction

1.1 An example problem

We begin the introduction with the well-known stability problem for a zonal jet stream, whose concepts have been widely applied in geophysical fluid flow studies. Let t be the time, and (x, y, z) denote the coordinates for a right-hand system, with x, y toward the east and north, respectively. The stability for a quasi-geostrophic jet, $\bar{u} = \bar{u}(y, z)$, which is confined within a zonal channel Ω, is governed by the following equation of evolution for the nondimensionalized perturbation energy[1] $\left\langle E^M \right\rangle^\Omega$, which measures the strength of the disturbances (the eddy velocity (u', v') and eddy density ρ'):

$$\frac{\partial \left\langle E^M \right\rangle^\Omega}{\partial t} = -R_0 \left\langle \overline{u'v'} \frac{\partial \bar{u}}{\partial y} \right\rangle^\Omega - R_0 \left\langle C \ \overline{v'\rho'} \frac{\partial \bar{u}}{\partial z} \right\rangle^\Omega, \tag{1.1}$$

where R_0 is the Rossby number, and $C = C(y, z)$ a time invariant factor depending only on the planetary rotation and the stratification of the background density field. The overbar is an ensemble mean, whereas the angle bracket with the superscript Ω signifies a spatial average over the whole domain Ω. The right hand side of this equation corresponds to two important processes in geophysical fluid dynamics (GFD): the barotropic instability and the baroclinic instability. It provides quantitative information about the energy source for the growth of disturbances.

If examined carefully, Eq. (1.1) involves:

[1] Refer to §4.2 for details.

(1) An ensemble mean, which in practice is usually replaced by a time or space average over a scale order(s) of magnitude larger than the scale of the event of concern,

(2) An integration/average over the whole basin Ω.

For convenience, we will temporarily refer to (1) and (2) respectively as a "bar average" and a "bracket average". Eq. (1.1) tells us, that the energetics provided by the theory are all averaged over a region global in space and/or time. While these averaged quantities may be appropriate for a spatially and temporally homogeneous process, it is hard to believe that they can represent those episodic events with significant energy bursts. The local information tends to be disguised or even obliterated if averaged with irrelevant events. Ironically, the original purpose of GFD stability theory is to study those processes like weather in the atmosphere or meso-scale eddies in the ocean, whose occurrence is accompanied with intermittence and is hence in general not at all homogeneous in space, nor stationary in time. Apparently, there exists a gap between the theory and the reality, and the usefulness of Eq. (1.1) is therefore limited.

Historically, the difficulty in making applications of Eq. (1.1) has long been recognized, and efforts have been made in an attempt to circumvent the problem. The process identification by buoyancy conversion, for example, is just such an effort. By *buoyancy conversion* we mean the rate of conversion from potential energy to kinetic energy due to the buoyancy force. For a one-dimensional problem, say the oscillation of a pendulum, a positive conversion implies a release of potential energy for motion, which corresponds to an instability (more specifically, a baroclinic instability). Buoyancy conversion is a field function, defined completely in a local fashion and easy to calculate. As a matter of fact, many past stability analyses with observations use buoyancy conversion as an identifier (e.g., Willbrand and Meincke, 1980).

A more complicated (though not necessarily more sophisticated) effort is the parcel stability analysis, which, by neglecting the pressure and dissipation effects, treats a fluid flow as a cluster of individual parcels. The instability for each parcel is thus local, either in time or in space. This approach has been used to study symmetric instability in the atmosphere (e.g., Holton, 1992).

Both the buoyancy conversion identification and the parcel stability analysis, however, can only be used with caution under certain circumstances, since neither of them contains the complete physics for a real problem. As we know, geofluid flows are of infinite dimensionality. A pendulum-like system is definitely too crude to model them. For the parcel stability analysis, the pressure influence is dropped, totally excluding the possibility of planetary wave propagation, which is of fundamental importance in atmospheric or oceanic dynamics. It is hard to believe such a model can tell us much about the physics for a real flow. We have no choice but to face Eq. (1.1).

In Eq. (1.1), it is those averages (the "bar average" and "bracket average" as we have called them) that give rise to problems. But what is the reason to have them there? Wouldn't it be okay to replace the equation with a non-averaged or locally averaged one, so that the energetics appear in their regional and/or instantaneous forms?

To answer the question, look at the bracket average first. Mathematically it is an inner product defined as a measure of the system. (See §2.1 for the definition of inner product.) But from a physical point of view what really matters is that a basin average/integration helps remove the boundary fluxes, which bear no relevance to the scale-scale energy transfer. If the transport processes can be reasonably separated from the transfer processes,[2] then a replacement of the bracket with some localized average will be just fine.

For the bar average, things are far more complicated. In contrast to its bracket counterpart, it is not just an "average" in the normal sense. In fact, it is the simplest form of an analytical device (a transform or a synthesis, depending on the context) which is used to decompose a field according to scales. The bars appearing in Eq. (1.1) result from the scale decomposition, and hence cannot be removed or simply changed. In fact, as we will know soon, none of the classical methods of decomposition will yield a multiscale energy which is local in nature. A "localization" of Eq. (1.1) thus requires some analytical method conceptually different from what we already have. This is why this example problem is so difficult and for so long a period the theory-reality gap has not been filled.

[2] Loosely speaking, a transport is a process of energy or matter exchange related to advection, while a transfer process regards the energy exchange across different scales. For a strict definition, refer to §3.6, particularly §3.6.1.

1.2 Regional ocean modeling and energy and vorticity analysis

The example problem actually belongs to a much larger set of problems in geophysical fluid dynamics. Classically, many theories are developed globally in the sense that the system of concern is treated as a whole. The conservation of energy and enstrophy, the Fourier analysis-based wave theories, as well as the above instability problem, are all contained in this category. While the physical robustness of these theories is not in question, it is usually hard to reconcile them with real regional datasets because of the reasons described above. This regional ocean process interpretation problem used to receive little attention, however, as observations were scarce and people usually focused more on homogeneous phenomena on larger scales. During the 80's, this problem began to surface, and has become more and more pressing, particularly after the advent of regional ocean modeling.

Regional ocean modeling came into existence only about two decades ago. It flourished in company with the development of the Harvard Ocean Prediction System (HOPS). Since the birth of the HOPS, there has been an upsurge of interest in real-time forecasting (e.g., Robinson, 1992, 1998, 1999; Robinson et al., 1996, 1999, 2000), which is regional and synoptic in nature. Real-time forecasting differs from global ocean modeling in that it chooses a local block of area so as to achieve as fine as possible a resolution for the processes, with the connection to outside ocean parameterized as boundary conditions. Regionally modeled results are rarely in statistical equilibrium. They are in general characterized by spatial inhomogeneity and temporal nonstationarity. Interpretation of a dataset thus produced is usually difficult in terms of the processes conceptualized with theories developed in a global sense. Some new methodology is therefore needed, in order to bridge the gap between the global theories on one side, and the highly localized events on the other.

Energy and Vorticity Analysis (EVA) is just such a methodology. It aims to understand the physics of a given region of ocean (or atmosphere) in a uniform, generic, and most importantly, local, approach. Specifically, the EVA probes oceanic processes with energy and vorticity equations which are localized in nature, terms decomposed into dynamically meaningful units (Pinardi and Robinson, 1986; Spall, 1989). Here "vorticity" is referred peculiarly to the curl of horizontal velocity, i.e., the upward component of vorticity in the usual meaning. As for energy, it comes in many forms. In the EVA we consider only mechanical energy, which includes the energy of motion (kinetic) and the energy of position (potential).[3] For a geofluid, potential energy is not a good measure of work capacity, as a predominant part of it cannot be released for motion. A simple but typical estimation given by Lorenz (see Lorenz, 1967) shows that only 5% of the total is available. In this sense, the atmosphere or ocean is a rather inefficient heat engine, and for this reason, geophysical dynamists usually adopt another measure, *available potential energy*, to represent the energy storage. Available potential energy (APE) is also an energy of position, but with a huge inert part removed. It constitutes, together with kinetic energy (KE) and vorticity, the group of diagnostic fields for the EVA package.

The choice of energy and vorticity for the process diagnosis is natural. Energy is a first integral for a frictionless dynamical system. It is the capacity to move fluid parcels to make a flow. Vorticity derives itself from velocity. It is a measure of rotation, obviously important to rotating flows. In oceanic or atmospheric dynamics, the importance of vorticity can never be overemphasized. Virtually all the conservation laws peculiar to this field found thus far are related to vorticity. These laws, with vorticity, velocity, and position denoted respectively by ζ, $\underline{\mathbf{v}}$, and $\underline{\mathbf{x}}$, include

- Conservation of (Ertel) potential vorticity for inviscid flows, (Pedlosky, 1979)
- Conservation of $I_k = \int \zeta^k(\underline{\mathbf{x}})d\underline{\mathbf{x}}$, $(k = 1, 2, 3, ...)$ for 2-D inviscid flows, (Chorin, 1994)
- Conservation of helicity $\mathcal{H} = \int \underline{\zeta} \cdot \underline{\mathbf{v}} d\underline{\mathbf{x}}$ for ideal 3-D flows, (Chorin, 1994. P11)
- Conservation of impulse $\mathcal{I} = \int \underline{\mathbf{x}} \wedge \underline{\zeta} d\underline{\mathbf{x}}$ for any real 3-D flows, (Chorin, 1994. P12)
- Conservation of enstrophy (vorticity squared) for a 2-D quasi-geostrophic flow with vanishing boundary conditions, (Pedlosky, 1979)
- Conservation of $\iint F(q(x, y, z, t))dxdy$ for a 3-D quasi-geostrophic flow rigidly bounded to the north and south, and cyclic in the east-west direction. Here $F(q)$ is an *arbitrary* function of the quasi-geostrophic potential vorticity q, (Salmon, 1998, P280; Majda and DiBattista, 2000)

[3]For a hydrostatic atmosphere, the internal energy is proportional to the potential energy. In this case, the thermal energy is actually also in consideration, as it can be combined with the potential energy into a single term. The combined energy is called the *total potential energy*. See Holton (1992).

where all the integrations are over the whole domain. Some of these laws, together with the conservation of energy, have enabled reproduction of certain type of geophysical turbulence with energy-carrying unit vortices as the building blocks. (e.g., Chorin, 1994; Salmon, 1998; Majda and DiBattista, 2000). The success of this kind of energy-vorticity related model corroborates, from another aspect, our choice of energy and vorticity for the dynamical process interpretation.

Using energy and vorticity to analyze processes is a common practice in geophysical fluid dynamics, but until Pinardi and Robinson (1984), the energetic studies before had always been basin-integrated and/or time-averaged. As a result, they cannot be employed to handle the inhomogeneous and nonstationary events with the HOPS outputs. In an attempt to overcome this difficulty, Pinardi and Robinson introduced the concept of EVA, which aims at tackling problems with local dynamics. EVA has been since developed with a quasi-geostrophic (QG) model (Pinardi and Robinson, 1986) and a primitive equation (PE) model (Spall, 1989). Applications have been successfully made to interpret the evolution of those fields in the POLYMODE and OPTOMA regions (Pinardi and Robinson, 1986; 1987), the dynamics of the California Current system (Robinson et al., 1986), the formation of rings in the Gulf Stream region (Robinson et al., 1988), the jet formation and evolution with the POLYMODE Mark II data set (Spall, 1989), the dynamical structure of the Rhodes Gyre System (Millif and Robinson, 1992), the merger of two idealized vortices (Masina et al., 1993), and the cold-core intrusion in the QG model forecasting of Iceland-Faeroe Frontal variability with the 1992 dataset (Miller et al., 1995).

The classical EVA, however, does not provide information on how energy is redistributed over the power spectrum. It lacks the multiscale representation in the earlier energetic studies. This is unfortunate, as the importance of multiscaling has been well recognized for geofluid flows, and problems such as the above mentioned barotropic and baroclinic instabilities are directly formulated with a two-scale decomposition. A major task of this study, therefore, is to re-formulate the EVA such that it is capable of multiscale energetic analyses. To distinguish this re-formulated methodology from its classical counterpart, we will call it a **Multiscale Energy and Vorticity Analysis**, or MS-EVA for short. As the vorticity analysis has been carefully investigated by Spall (1989), when we refer to the formulation of the MS-EVA, we actually mean the development of its multiscale energetics and multiscale enstrophy part. For the same reason, sometimes the name MS-EVA may be used identically as the multiscale energetic study without further clarification. (Enstrophy can be viewed as a kind of generalized energy.)

What characterize the MS-EVA are the multiscale representation and the localization. As mentioned in the opening section, it is a challenging business to decompose a field by scale while keeping the decomposed energy localized. (We will see soon why it is so difficult) A decomposition device which is conceptually different from the classical ones is needed to fulfill this mission. In the following, we first look at what is characteristic of the multiscale GFD processes, then give a brief review of the multiscale energetic approaches to these processes to date, which eventually leads to the strategy to tackle our problem.

1.3 Multiscale phenomena in geofluid flows

In order to develop a tool for some process, one first needs to have a general knowledge of that process. In this chapter, we give a brief overview of the multiscale phenomena in geophysical fluid flows.

It has long been observed that oceanic or atmospheric flows are notoriously rich in scales. In a seemingly quiescent ocean, for example, there could exist millions of events, if ordered by size, from capillary waves to planetary waves. These events, large or small as they are, could be directly driven by the external forcing, as is most understood, or could be from interactions between the existing processes, or spawned by a parental system with another scale. The meandering and filamenting of intense currents like Gulf Stream, the sheetification of an otherwise circular vortex, the sudden event bursts in boundary layers, the absorption of eddies back into the mean current system, to name but a few, are all good examples.

A major observation from multiscale phenomena is the self-similarity symmetry. By a symmetry for a dynamical system we mean a transformation that, if applied, will not alter the system itself. Loosely speaking, a symmetry is just a kind of invariance. A self-similarity symmetry is a scale invariance against change. That is to say, some features of a process repeat themselves upon rescaling. In physics, a symmetry of a system usually corresponds to some underlying laws, and in a Lagrangian formalism, these laws are expressed as first integrals. In particular, the time symmetry implies the conservation of energy, the invariance under translation gives the conservation of momentum, and the isotropicity is responsible for the conservation of angular momentum. In geophysical fluid dynamics, the most important conservation law, the conservation of Ertel potential vorticity, can also be related to a kind of symmetry. It is an invariance under arbitrary particle-relabeling (Salmon, 1998; also seen in Ripa, 1980). Another symmetry, which is fundamental in the study of wave-mean current interaction, relates the conservation of wave action to the invariance under phase translation (Salmon, 1998).

In distinction to the above known symmetries, the self-similarity symmetry has not received much attention until recent decades. But its observation can be dated back to a long time ago in a vast variety of fields. In the area of turbulence study, a widely cited sonnet by Richardson in 1922 (see Khac, 1997) provides a testament of this long-existing account:

> Large vortices beget small vortices
> Which feed off their velocity
> And the small vortices beget smaller vortices
> And so on down to viscosity

though ingredients of it had appeared far before. Another excellent testament of the self-similarity symmetry is Kolmogorov's $-\frac{5}{3}$ power law of energy spectrum (e.g., Chorin, 1994). It states that energy for three-dimensional isotropic turbulence of universal equilibrium in the inertial range (meso-scale eddies, for instance) is distributed over the wavenumber (k) spectrum in a form

$$E(k) = C\epsilon^{2/3}k^{-5/3},$$

where ϵ is the rate of energy dissipation with viscosity ν and C a dimensionless constant. Thought not derived rigorously from a dynamical equation, this law has been well reproduced either in physical and numerical experiments. A self-similarity symmetry, of course, need not lead to a power law; a power law, however, does result in a self-similarity symmetry. Kolmogorov's $-\frac{5}{3}$ law tells us that, if the dimension is re-scaled (here a scale is taken to be the reciprocal of k) by a factor r, then the spectrum will be still the same up to a multiplicative factor $r^{5/3}$.

Scale invariance or self-similarity symmetry does not necessarily mean a lack of characteristic scales, in which case life would turn out to be boring. In a process-rich geophysical fluid system, scale invariance usually has its limiting bounds. Beyond these bounds different laws (power laws, for example) may apply. For an isotropic turbulence of universal equilibrium, Kolmogorov's $-\frac{5}{3}$ law holds only in a meso-scale range called *inertial range* on the power spectrum. This inertial range is bounded above and below respectively by the *dissipation range* and the *energy range*, as schematized in Fig. 1.1 (McComb, 1991). In this case, the energy range, inertial range, and dissipation range form a partition for the power spectrum. In a more general case, a power spectrum (either over wavenumber or over frequency) can be divided into three distinct regimes, i.e., large-scale, mesoscale, and sub-mesoscale regime in succession. A regime does not correspond to a single scale, but a range or a *window* of scales. This phenomenon is called **scale windowing**. The rigorous definition for a scale window is postponed to the next chapter. The size of a window is dependent on many factors, and may vary with the local dynamics. On a Kolmogorov spectrum, the inertial range could be large when the Reynolds number Re is large, for the energy dissipation, which is proportional to Re^{-1} (k the typical wavenumber), comes to reign only when k is large enough to make the product $k^2 Re^{-1}$ significant (see Chorin, 1994).

To summarize, multiscale phenomena are ubiquitous in geophysical fluid processes. The self-similarity symmetry and scale windowing, among others, are two prominent observations whose importance has been gradually recognized. A self-similarity symmetry is a scale invariance, while the scale windowing marks off the limits for the invariance. Just as other symmetries do in physics, the self-similarity symmetry and its limit demarcating may offer a clue to those complex phenomena such as turbulence, shedding light on the darkest corner of geophysical fluid dynamics. A satisfactory multiscale energetic study tool, therefore, should be able to represent faithfully these two characteristics in its formulation.

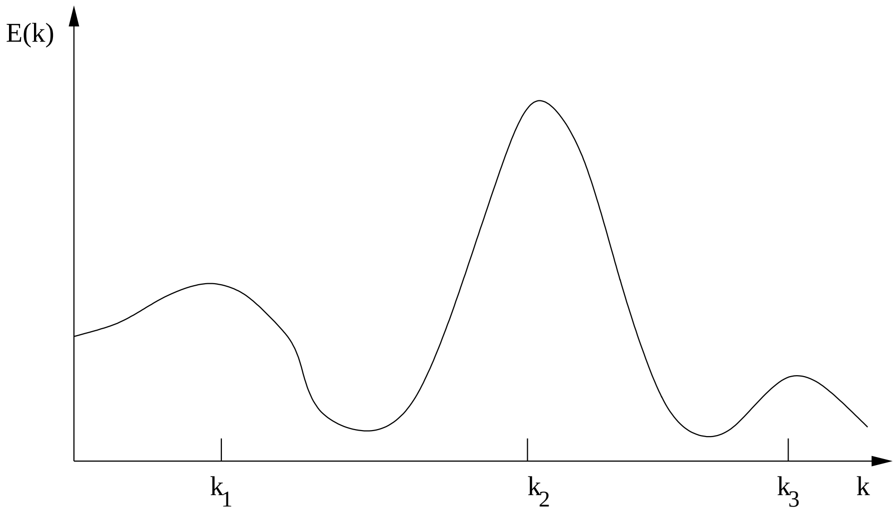

Figure 1.1: A cartoon of the three regimes on a typical power spectrum of an energetic geofluid flow system. Here the abscissa k could be either wavenumber or frequency. A similar paradigm is also seen in the experiment output shown in McComb (1991) (Fig. 3.6).

1.4 Fourier- and Reynolds-type multiscale energetics

A new finding is based on the knowledge of the past. Historically there have been many kinds of formulations for energetic studies. Different in form as they may be, these formulations can be generally classified into two categories: the Fourier-type and the Reynolds-type. All others may be viewed as varieties or derivatives of them. In this section, we give these two categories a brief introduction with a simple model using the decomposition methods they originally adopt. Here by a decomposition method we mean an approach to the scale analysis or decomposition. (For a more strict definition, see Chapter 2.) The simple model is a two-dimensional hyperbolic problem, together with an incompressibility assumption,

$$\frac{\partial \underline{\mathbf{v}}}{\partial t} + \underline{\mathbf{v}} \cdot \nabla \underline{\mathbf{v}} = 0, \tag{1.2}$$

$$\nabla \cdot \underline{\mathbf{v}} = 0 \tag{1.3}$$

where $\underline{\mathbf{v}} = [u(t;x,y),\ v(t;x,y)]$ is the velocity. For convenience, the domain is assumed to be periodic in x.

The simple model might be made even simpler by reducing the dimensionality by one. In this case, however, the problem must be compressible, otherwise it admits only trivial solutions. Considering incompressibility is generally presumed for geophysical fluid flows, we keep its present 2-D form. Define a dyad for vectors $\underline{\mathbf{A}}$ and $\underline{\mathbf{B}}$, $\underline{\mathbf{AB}}$, such that

$$\nabla \cdot \underline{\mathbf{A}}\ \underline{\mathbf{B}} = (\nabla \cdot \underline{\mathbf{A}})\underline{\mathbf{B}} + (\underline{\mathbf{A}} \cdot \nabla)\underline{\mathbf{B}}. \tag{1.4}$$

Eq. (1.2) then can be written, with Eq. (1.3) substituted in, as

$$\frac{\partial \underline{\mathbf{v}}}{\partial t} + \nabla \cdot \underline{\mathbf{v}}\ \underline{\mathbf{v}} = 0. \tag{1.5}$$

1.4.1 Fourier-type formulation

The Fourier-type formulation tackles the energetics problem by transforming the model equation with an orthonormal basis (cf. Chapter 2) over the periodic direction x. Originally the basis used is harmonic, and that is why the Fourier-type gains its name. Let

$$\underline{\mathbf{v}}(t; x, y) = \sum \hat{\underline{\mathbf{v}}}_m(t; y) e^{imx}, \tag{1.6}$$

where $i = \sqrt{-1}$, and

$$\hat{\underline{\mathbf{v}}}_m(t; y) = \mathcal{F}_m \left[\underline{\mathbf{v}}(t; x, y)\right]. \tag{1.7}$$

The model equation in the Fourier domain is then

$$\frac{\partial \hat{\underline{\mathbf{v}}}_m}{\partial t} + \mathcal{F}_m \left[\nabla \cdot \underline{\mathbf{v}}\ \underline{\mathbf{v}}\right] = 0. \tag{1.8}$$

Denote

$$K_m = \frac{1}{2} \hat{\underline{\mathbf{v}}}_{-m} \cdot \hat{\underline{\mathbf{v}}}_m. \tag{1.9}$$

K_m is the kinetic energy for wavenumber m (cf. Chapter 2). Eq. (1.8) dotted with $\hat{\underline{\mathbf{v}}}_{-m}$ followed by taking the real part gives

$$\frac{\partial K_m}{\partial t} = -\hat{\underline{\mathbf{v}}}_{-m} \cdot \mathcal{F}_m \left[\nabla \cdot \underline{\mathbf{v}}\ \underline{\mathbf{v}}\right]. \tag{1.10}$$

This is the evolution equation of the kinetic energy for the wavenumber m. The dependencies on y and t are kept intact. For a fixed y and t, the right hand side describes the transfer of kinetic energy to wavenumber m from all other wavenumbers (see Chorin, 1994; Hansen, 1981).

1.4.2 Reynolds-type formulation

Reynolds-type formulation originally approaches the problem statistically. An ensemble is needed for each field variable in study. Consider again the toy model equation (1.5), and suppose now we have at hand an ensemble of realizations for $\underline{\mathbf{v}}$. Let $\bar{\underline{\mathbf{v}}}$ be the mean of $\underline{\mathbf{v}}$ over the ensemble, and $\underline{\mathbf{v}}'$ the instantaneous deviation from it, $\underline{\mathbf{v}}' = \underline{\mathbf{v}} - \bar{\underline{\mathbf{v}}}$. With this decomposition, the mean energy and eddy energy are defined as $K_{\text{mean}} = \frac{1}{2}\bar{\underline{\mathbf{v}}} \cdot \bar{\underline{\mathbf{v}}}$ and $K_{\text{eddy}} = \frac{1}{2}\overline{\underline{\mathbf{v}}' \cdot \underline{\mathbf{v}}'}$, respectively (cf. Chapter 2). It is easy to check that $K \equiv \frac{1}{2}\overline{\underline{\mathbf{v}} \cdot \underline{\mathbf{v}}} = K_{\text{mean}} + K_{\text{eddy}}$. The Reynolds equation of (1.5) is

$$\frac{\partial \bar{\underline{\mathbf{v}}}}{\partial t} + \nabla \cdot \bar{\underline{\mathbf{v}}}\ \bar{\underline{\mathbf{v}}} = -\nabla \cdot \underline{\underline{\mathbf{T}}}, \tag{1.11}$$

where

$$\underline{\underline{\mathbf{T}}} = \begin{bmatrix} \overline{u'u'} & \overline{u'v'} \\ \overline{v'u'} & \overline{v'v'} \end{bmatrix} \tag{1.12}$$

is the 2-D Reynolds stress. The eddy field equation results from a subtraction of the Reynolds equation from (1.5),

$$\frac{\partial \underline{\mathbf{v}}'}{\partial t} + \nabla \cdot (\underline{\mathbf{v}}\ \underline{\mathbf{v}})' = 0. \tag{1.13}$$

In order to get the equation for the mean energy K_{mean}, take a dot product of $\bar{\underline{\mathbf{v}}}$ with (1.11). This gives

$$\frac{\partial K_{\text{mean}}}{\partial t} + \bar{\underline{\mathbf{v}}} \cdot \nabla \bar{K} = -\bar{\underline{\mathbf{v}}} \cdot \nabla \cdot \underline{\underline{\mathbf{T}}}. \tag{1.14}$$

The evolution law for K_{eddy} can be obtained similarly, this time by dotting $\underline{\mathbf{v}}'$ to (1.13) followed by an application of the mean operator. After some manipulation, the resulting equation becomes

$$\frac{\partial K_{\text{eddy}}}{\partial t} + \overline{\underline{\mathbf{v}} \cdot \nabla \cdot \underline{\mathbf{v}}'\underline{\mathbf{v}}'} = -\overline{\underline{\mathbf{v}}'\underline{\mathbf{v}}'} : \nabla \bar{\underline{\mathbf{v}}}, \tag{1.15}$$

where the colon operator is defined, for two dyads $\underline{\mathbf{AB}}$ and $\underline{\mathbf{CD}}$, such that

$$\underline{\mathbf{AB}} : \underline{\mathbf{CD}} = (\underline{\mathbf{B}} \cdot \underline{\mathbf{C}})\ (\underline{\mathbf{A}} \cdot \underline{\mathbf{D}}). \tag{1.16}$$

In either Eq. (1.14) or Eq. (1.15), the second term on the left hand side represents some advection-related effect. The right hand side describes a different mechanism. One may call it the interaction between the mean flow and the Reynolds stress (see Goldstein, 1965; Monin and Yaglom, 1971; Harrison and Robinson, 1978; McComb, 1991). But here the advection effect has not been completely separated from the Reynolds stress term. As we have mentioned before, a unique separation is one of the difficulties for a localized energetic formulation.

It should be pointed out that, the statistical ensemble needed for this Reynolds formalism is usually not available. A common practice is to sample $\underline{\mathbf{v}}$ in time or space and convert (1.14) or (1.15) into a time meaned or space averaged energy equation. (Harrison and Robinson, 1978; Hansen, 1981). When there is a sizable gap on the power spectrum such that the meso scale is order(s) smaller in magnitude than the large scale, either the time mean or space average can preserve the eddy energy equation from deviating from its original Reynolds form by a simple argument based on the multiple scale perturbation analysis. In general, however, spectra are continuous, and in performing the time or space sampling, two energy equations different from (1.14) and (1.15) result. If sampled spatially in one direction (say, x), the derivative terms involved in that direction will be gone, and hence terms like those with Reynolds stress will no longer be retained. In this case, the resulting energetics are essentially as those of a special Fourier formalism, with harmonics with $|m| > 0$ all assembled together to form an eddy energy equation. If sampled in time, for the same reason, the time derivatives have to be changed (Harrison and Robinson, 1978). This change, however, affects only the left hand sides of Eqs. (1.14) and (1.15). The rest of the terms are all kept in their original forms. Therefore, in the light of practicality, only time direction sampling keeps the essence of this type of energetic study. *We henceforth refer to as Reynolds-type only those time-sampled formulations, to which the statistical notion is no longer attached.* (Other formulations from transforms in time are also classified into this category, since they can be reduced to a form similar to the time-sampled ones.)

Historically Fourier-type and Reynolds-type energetic studies have been developed in parallel. In their original formulation, the Fourier-type disguises information of spatial localities because of the employment of trigonometric functions, while the Reynolds-type loses temporal localization after applying the time averaging. In other words, from whichever direction the multiscaling is gained, the localization is lost in that direction. This is the reason why none of the classical multiscale energetic studies are capable of the EVA mission, and it is understandable that the previous EVA (Pinardi and Robinson, 1986; Spall, 1989) keeps both time and space from being sampled to gain the localization at the loss of the multiscaling. The dilemma between the multiscaling and localization is just where the difficulty of the MS-EVA lies. In principle, none of the classical analytical tools can make the two achieved simultaneously.

1.5 A comparison of the available multiscale energetic formulations

The purpose of this thesis is getting clearer now. But before proceeding, it is instructive to make a comparison between the available multiscale formulations. These formulations of course are all within the above two categories, except for different decomposition methods.

An ideal MS-EVA formulation should be able to reflect the two essential features from multiscale observations, as well as to meet the EVA needs. EVA requires localization, both in time and in space, while multiscaling expects a faithful representation of the self-similarity and scale windowing. These requirements form the basis for us to evaluate the energetic formulations available so far.

The Fourier transform is undoubtedly one of the most commonly used approaches to scale decomposition. Its global nature in amplitude and wavenumber (frequency), and hence energy, however, basically excludes it from the candidate list for our MS-EVA purposes. In fact, for a nonstationary and nonlinear process, any similar transforms with global orthonormal bases (cf. Chapter 2) turn out to be hopeless. When applied in one direction, localities of events are lost in that direction. A particular tool which has been widely used in atmospheric and oceanic sciences is the principal component analysis (PCA, see Appendix A5.1), or empirical orthogonal function (EOF) analysis as preferred by oceanographers. It is global in nature in either time or space, depending on which direction it is applied, and as a result, the energy represented with it is not appropriate for our purposes.

The time average is another widely adopted approach for the multiscale energetic problem. As already mentioned above, the Reynolds-type formulation thus obtained doesn't preserve information of time localities. Improvement of this approach leads to the running time average, which has been proposed in Harrison and Robinson (1978) and many others. While preserving the simplicity of ensemble averages, the running average introduces localized information in time direction. Of course, such a time localization is usually too crude. Partitioning the time span into more sub-intervals to increase the resolution may be partially a solution, but this can give only too small an ensemble for each sub-interval, and hence the adequacy of the decomposition is questionable.

Implementation of running-averages leads to filters. A low-pass filter is actually a weighted running average. Localization is hence there. However, we haven't seen any energetic formulation with filters so far. As will be established in the next chapter, not all filters can be employed for energetic studies. The widely-used objective analysis (OA) (see Appendix A5.2 of for definition), for example, turns out to be NOT energy conservative (cf. p. 40), and its eligibility is thus negated.

Nonetheless, another analysis tool, which is closely related to filters (or filter banks, to be precise), has received much attention during the recent years. This is the wavelet analysis.

Wavelet analysis is a newly invented technique which has been broadly applied in virtually all the fields where previously the Fourier analysis is used. There is no exception in the field of energetic studies. The localized nature of wavelets provides an excellent choice of *basis* (refer to §2.1.5 for definition) for the nonstationary field decomposition, and multiscale energetic studies with wavelet bases indeed meet the localization requirement imposed by the MS-EVA (e.g., Farge, 1992; Iima and Toh, 1995; Khac, 1997; Fournier, 1999). Usually in these studies the formulations are of Fourier-type, but there is no difficulty to extend them to Reynolds-type. The problem is, as shown schematically in Fig. 1.2, the resolution for large scales is too low, and the lower the scale level, the lower the resolution (refer to §2.2.5). At scale level 0, only one energy value (or two, depending on the extension scheme; see next chapter) is available. From an MS-EVA point of view, this is obviously not tolerable.

We have examined the localization aspect for formulations with Fourier analysis, principal component analysis, time average, running time average, and wavelet transform. The other two factors are the scale invariance and the scale windowing. For the scale invariance, the Fourier analysis and wavelet analysis have built-in self-similarity structures, but the others generally don't. For the scale windowing, it is well represented only in average-based formula-

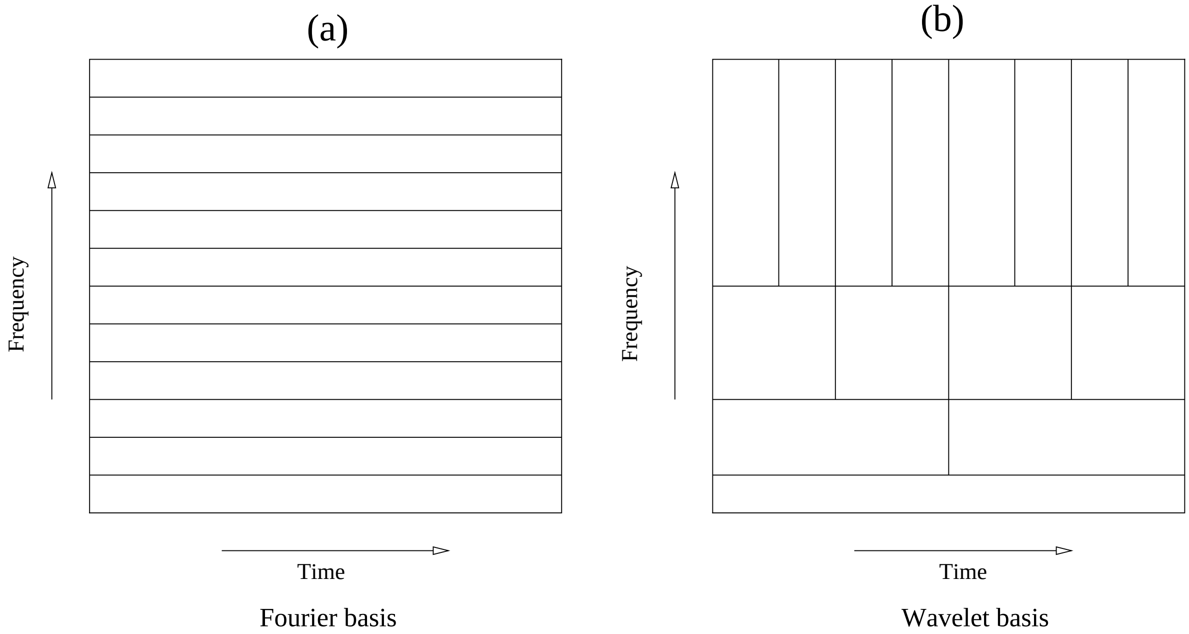

Figure 1.2: Schematic of the time-frequency plane decomposition using a global (Fourier) basis (a) and a local (wavelet) basis (b). For a Fourier case, the transform is a function of frequency only. There is no variation in the time direction. For the wavelet case, information is retained in the transform either in time or in frequency, but the time resolution varies with frequency. The lower the frequency, the coarser the resolution. (Adapted from Kumar and Foufoula-Georgiou, 1997, and Strang and Nguyen, 1997).

tions. The others lack this representation. They treat scales individually. This is unfortunate, as a scale window does not need to be just an accumulation of individual scales. In language of modern dynamics, the complexity of a system of units does not need to be the sum of the complexities of those individual units. In the presence of nonlinearity, it could be simpler, or could be more complex. This is the so-called problem of *emergence* (see Bar-Yam, 1997). Nonetheless, one may achieve the windowing by a summation of energy over the scale indices within the window of concern (cf. Chapter 2), except with the wavelet-based formulations. In that case, no cross-scale summation is permissible. The reason is best seen with the aid of Fig. 1.2b. On a wavelet spectrum, energy is defined discretely on different points (because of the different resolution) for different scale levels. They cannot be summed together to get the energy for a designated location.

To close this section, we tabulate and comment briefly in Table 1.1 all the aforementioned formulations. In the column of examples, it is not purported that an exhaustive list of literature is provided. Only typical ones are shown. In the Reynolds category, most of the formulations tabulated actually have not been seen as yet. We put them in there just for comparison. Other analysis approaches may offer new formulations. For example, the Hilbert transform has been evidenced to produce better localized nonstationary power spectra than wavelets do (Huang et al., 1999). But the results thus produced are correct only approximately if the data has a broad spectrum of frequency (Long et al., 1993). Actually, to my knowledge, so far no energetic studies have been conducted with the Hilbert transform.

1.6 Strategy of the MS-EVA development and an outline of this thesis

Look back to the example problem. We need to develop a localized multiscale energetic analysis, with the transport and transfer processes distinctly separated, and a localized average which is independent of the study domain. The latter is for the instability study only and we will introduce it where needed. The heart of the problem is the former. It is purposed in this thesis that this localized analysis be formulated, validated, and applied with a typical ocean dataset. To start, we first need to choose from the two available types of formulation, the Fourier and the Reynolds, one type that serves our purposes best.

The formalism choosen is based entirely on the MS-EVA problem itself. We first expect the new formulation to take keeps a form similar to that of the classical EVA. In the Fourier-type formalism, the transport mechanism is converted into interactions between particle-like entities, a form completely different from what we already have in the EVA. But this is still not much of a problem. A major difficulty with the Fourier-type formalism is regarding the field scaling, which is merely the scale window definition in this context. There is no reason to adopt the same spatial scaling for different fields.[4] For example, the usual combination of prognostic variables for a primitive equation model includes zonal velocity u and density anomaly ρ. If only horizontal scales are considered, by the thermal wind relation, u should bear the same scale as that of $\frac{\partial \rho}{\partial y}$, instead of ρ. As we know, a field and the gradient of that field may have completely different multiscale structures, even in the same process. Barotropic turbulence is an excellent example. While on a streamfunction (Ψ) map energy gets transferred to larger

[4]Nadia Pinardi, private communication.

Table 1.1: A comparison of the available multiscale energetic formulations. Some non-studied formulations are also listed for comparison.

Forumulation type	Decom-position	Time localization	Space localization	Scale windowing	Self-similarity symmetry	Examples
Fourier	Fourier transform	Yes	No	No	Yes	Saltzman (1957), Hansen (1981), Maltrud and Vallis (1991)
	PCA	Yes	No	Maybe	Generally no	Aubry et al (1988), Wilson and Wyngaard (1996)
	Wavelet transform	Yes	Yes, but w/ resolution problem	No	Yes	Iima and Toh (1995), Khac (1997), Fournier (1999)
Reynolds	Time average	No	Yes	Yes	No	Holland and Lin (1975), Harrison and Robinson (1978)
	PCA	No	Yes	Maybe	Generally no	
	Running time average	Yes but poor	Yes	Yes	No	
	Wavelet transform	Yes, but w/ resolution problem	Yes	No	Yes	

and larger scales, the opposite happens on the map of vorticity, which is the Laplacian of Ψ (e.g., Shepherd, 1990; Salmon, 1998; Chorin, 1994). It is thus very hard to obtain for each variable a consistent scaling, and hence a consistent scale windowing, if a Fourier-type formulation is adopted.

We therefore have to make the other choice, the Reynolds-type formalism. In a Reynolds-type MS-EVA, decomposition is performed in the time direction. Application of spatial gradients therefore won't alter the multiscale structures thus defined. The relative simplicity of a 1-D transform over a 2-D or 3-D transform, which correspond respectively to these two types of formalism, also favors our choice.

A Reynolds-type MS-EVA bases its formulation on the time scale analysis or time transform. From the preceding section, none of the existing approaches is appropriate for this purpose. A central issue to this thesis is, therefore, to find a transform with respect to time to meet the requirements of time localization (space localization is not a problem here) with a natural structure of self-similarity and scale windowing. This is the theme of the next chapter, and the resulting transform will be called a **multiscale window transform**. We shall first introduce the concepts and begin the construction with a multi-resolution analysis, the starting point from which a large family of wavelets are built. The MS-EVA is formulated in Chapter 3. Equations of multiscale energy and enstrophy are derived, and the processes the energetic terms represent are classified with clear interpretations. The physically reasonable separation of the transport process from the transfer process provides the basis for a successful solution of the example problem presented in the beginning of this thesis, and the success in handling this problem (with two idealized models) in turn serves to validate our methodology (Chapter 4). In that chapter, we will also show how the MS-EVA can be applied in process identification, and how the two classical instability processes, barotropic and baroclinic instabilities, are structured in terms of the MS-EVA analysis. This validated MS-EVA, together with the collected dynamical signatures for the two instabilities, is then taken to tackle a real problem: the study of the Iceland-Faeroe Frontal variability (IFF). We first launch a real-time forecasting, with the hydrographic data acquired during the 1993 R/V *Alliance* cruise near the IFF region (Chapter 5). The forecast output is then re-arranged into a MS-EVA-ready format, and the MS-EVA is performed with it as an input. The diagnostic results are presented in Chapter 6. From the maps shown, the energetics are characterized and the governing mechanisms of the frontal variability are investigated. In the last chapter, we summarize all the work that has been done, including the design of the mathematical tool, the development of the multiscale energetic analysis, the validation of the MS-EVA, the forecasting of the IFF variability, and the MS-EVA application to the frontal study. Following this summary is a brief discussion of further research directions, which marks the coda of the whole thesis.

Bibliography

[1] Aubry, N., Holmes, and Stone, 1988: Dynamics of coherent structures in the wall region of a turbulence boundary layer *J. Fluid Mech.*, 192, 115-173.

[2] Bar-Yam, Yaneer, 1997: *Dynamics of Complex Systems.* Perseus Books, Reading, Mass.. 848 pp.

[3] Chorin, Alexandre J., 1994: *Vorticity and Turbulence.* Springer. 174pp.

[4] Fournier, Aimé, 1999: Atmospheric energetics in the wavelet domain I: Governing equations and interpretation for idealized flows. Ph.D. Thesis, University of Maryland.

[5] Goldstein, S., 1965: *Modern Developments in Fluid Dynamics*, Vol. II. Dover, New York, 371 pp.

[6] Hansen, A. R., 1981: A diagnostic study of the spectral energetics of blocking. Ph.D. dissertation, Iowa State University, 183pp.

[7] Harrison, D. E., 1979: Eddies and the general circulation of numerical model gyres: an energetic perspective. *Rev. Geophys. Space Phys.*, 17: 969-979.

[8] Harrison, D. E, and A. R. Robinson, 1978: Energy analysis of open regions of turbulent flows-mean eddy energetics of a numerical ocean circulation experiment. *Dyn. Atmos. Oceans*, 2: 185-211.

[9] Holland, W. R., 1973: Energetics of baroclinic oceans. *Proceedings of Conference Numerical Modeling of the Oceans.* Washington, D. C., Nat'l Acad. Sci..

[10] Holland, W. R., and L. B. Lin, 1975: On the generation of mesoscale eddies and their contribution to the oceanic general circulation. I. A preliminary numerical experiment. *J. Phys. Oceanogr.*, 5, 642-657.

[11] Holland et al., 1983: In *Eddies in Marine Sciences*, A. R. Robinson, ed. (p391).

[12] Holton, James R., 1992: *An Introduction to Dynamic Meteorology.* Academic Press. 496pp.

[13] Huang, Norden E., Z. Shen, and S. R. Long, 1999: A new view of nonlinear water waves: The Hilbert spectrum. *Annu. Rev. Fluid Mech.*, 31, 417-57.

[14] Iima, Makoto, and Sadayoshi Toh, 1995: Wavelet analysis of the energy transfer caused by convective terms: Application to the Burgers shock. *Phys. Rev.*, E, 52:6, 6189-6201.

[15] Khac, Minh Do, 1998: *Use of Orthogonal Wavelets in the Analysis of Two- and Three-Dimensional Inhomogeneous Turbulent Flows.* European Space Agency Technical Translation of ONERA NT 1997-04.

[16] Kumar, Praveen, and Efi Foufoula-Georgiou, 1997: Wavelet analysis for geophysical applications. *Rev. Geophys.*, 35, 4, 385-412.

[17] Lesieur, M., 1990: *Turbulence in Fluids: Stochastic and Numerical Modelling.* 2nd ed., Kluwer Academic Publishers, 400 pp.

[18] Long, S. R., N. E. Huang, C. C Tung, M. L. Wu, and R. Q. Lin, 1993: The Hilbert techniques: an alternate approach for non-steady time series analysis. *IEEE Geosci. Remote Sensing Soc. Lett.*, 3: 6-11.

[19] Lorenz, E. N., 1967: *The natural and theory of the general circulation of the atmosphere.* World Meteorological Organization, Geneva.

[20] Majda, A. J., and M. Dibattista, 2000: An equilibrium statistical theory for large-scale features of open-ocean convection. *J. Phys. Oceanogr.*, 30, 1325-1353.

[21] Maltrud, M. E., and G. K. Vallis, 1991: *Phys. Fluids* A 3, 1760.

[22] Masina and Pinardi, 1993: *J. Phys. Oceanogr.*, 23, 1618-1637.

[23] Miller, A. J., H. Arango, A. R. Robinson, W. G. Leslie, P.-M. Poulain, and A. Warn-Varnas, 1995a: Quasigeostrophic Forecasting and Physical Processes of Iceland-Farore Frontal Variability. *J. Phys. Oceanogr.*, 25, 1273-1295.

[24] McComb, W. D., 1991: *The Physics of Fluid Turbulence.* Oxford Science Publications. 572 pp.

[25] Millif, R. F., A. R. Robinson, 1992: Structure and dynamics of the Rhodes Gyre System and dynamical interpolation for estimates of the mesoscale variability. *J. Phys. Oceanogr.*, 22, 317-337.

[26] Monin, A. S., and A. M. Yaglom, 1971: *Statistical Fluid Mechanics*, Vol. 1. MIT Press, Cambridge, Mass., 769 pp.

[27] Pedlosky, J., 1979: *Geophysical Fluid Dynamics.* Springer-Verlag.

[28] Pinardi, N., and A. R. Robinson, 1986: Quasigeostrophic energetics of open ocean regions. *Dyn. Atmos. Oceans* 10, 185-219.

[29] Pinardi, N., and A. R. Robinson, 1987: Dynamics of deep thermocline jets in the POLYMODE region. *J. Phys. Oceanogr.*, 17, 1163-1188.

[30] Robinson, A. R., 1983: *Eddies in Marine Science*, Springer-Verlag, 609 pp.

[31] Robinson, A. R., et al., 1986: *J. Phys. Oceanogr.*, 16, 1561-1579.

[32] Robinson, A. R., 1992: Shipboard prediction with a regional forecast model. *Oceanography*, 5, 42-48.

[33] Robinson, A. R., 1999: Forecasting and simulating coastal ocean processes and variabilities with the Harvard Ocean Prediction System (HOPS). In: *Coastal Ocean Prediction* (C. N. K. Mooers, ed.), AGU Coastal and Estuarine Studies Series, 77-100, American Geophysical Union.

[34] Robinson, A. R. and the LOOPS Group, 1999: *Real-time forecasting of the multidisciplinary coastal ocean with the littoral ocean observing and predicting system (LOOPS)*, Preprint volume of the third conference on coastal atmospheric and oceanic prediction

and processes, 3-5 November 1999, New Orleans, LA, American Meteorological Society, Boston, MA.

[35] Robinson, A. R., and J. Sellschopp, 2000: Rapid assessment of the coastal ocean environment. *Ocean Forecasting: Conceptual Basis and Applications.* N. Pinardi and J. D. Woods (eds.), Springer, 203-232.

[36] Salmon, Rick, 1998: *Lectures on Geophysical Fluid Dynamics.* Oxford University Press. 378pp.

[37] Saltzman, B., 1957: Equations governing the energetics of the larger scales of atmospheric turbulence in the domain of wave number. *J. Meteor.*, 14, 513-523.

[38] Schroeder, Manfred, 1991: *Fractals, Chaos, Power Laws.* W. H Freeman and Company, New York. 429pp.

[39] Shepherd, T. G., 1990: Symmetries, conservation laws, and Hamiltonian structure in geophysical fluid dynamics. *Adv. Geophys.*, 32, 287-338.

[40] Spall, M., 1989: Regional primitive equation modeling and analysis of the POLYMODE data set. *Dyn. Atmos. Oceans*, 14, 125-174.

[41] Willebrand, J., and J. Meincke, 1980: Statistical analysis of fluctuations in the Iceland-Scotland frontal zone. *Deep-Sea Res.*, 27A, 1047-1066.

[42] Wilson, D. K., and J. C. Wyngaard, 1996: Empirical orthogonal function analysis of the weakly convective atmospheric boundary layer. Part II: Eddy energetics. *J. Atmos. Sci.*, 53, 824-841.

Chapter 2

Multiscale window analysis

In multiscale energetic studies, energy is a concept defined with respect to some orthonormal analysis. The objective of this chapter is to design such an analytical tool, called the *multiscale window transform* (MWT) or *multiscale window analysis*.[1] It is localized, self-similar, and windowed on scales, as required by MS-EVA problems.

The layout of this chapter is as follows. We first develop the MWT with a one-dimensional (1-D) field (§2.1-§2.6), then have it extended to a multi-dimensional, particularly two-dimensional (2-D) case (§2.7). The first section is an overview of some facts from functional analysis. It is followed by an introduction to the concept of multi-resolution analysis, based on which we define and construct the *scale window*, multiscale window analysis, and energy on scale windows. Originally all these definitions are made in some space over $\mathbb{R}$, the whole of the real line. In practice, however, only signals of finite length are available. Extensions have to be performed to fit them into the theory, and the original definition and construction must be modified accordingly. Section 2.3 is on how this modification is made. With the two extension schemes that we will use throughout the thesis, the extensions by periodization and reflection, a system of finite domain multiscale window analysis is established and its properties investigated. Among these properties are marginalization and energy representation,

[1]The MWT to be built has nothing to do with the "windowed" or Gabor transform (e.g., Gabor, 1946; Bastiaans, 1980; Hallastchek and Zilber, 1998). In a Gabor transform, a "window" is a localized function of the physical space independent variable, while an MWT regards "windows" with respect to scales, as will be defined soon.

which will play important roles in the later MS-EVA formulation. Note in this analysis, all the functions are assumed to belong to some proper subspace of a space of square integrable functions defined on a circle. This assumption is justified with real problems in §2.4. The theoretical development stops here and in §2.5 and §2.6 the task is to find a proper scaling basis, realize the transform on computers, and validate it with testing examples. Once this is done, the 1-D MWT development is completed. Note §2.5 may be skipped over if read for the first time. One needs only to assume that he already has at hand an adequate scaling function, and then the reading wouldn't be interrupted. In §2.7, we extend the development to a multi-dimensional case, particularly a 2-D case, to build a 2-D MWT. The whole chapter is summarized in §2.8.

It should be pointed out that the word *analysis* is used differently in different settings. In the terms *multiscale analysis* (§2.2.2), *multiscale window analysis* (§2.2.5), and *multi-resolution analysis* (§2.2.1), it is rigorously defined, but in those like *energy/vorticity analysis*, it is used in a rather loose sense. This usage of terminology might cause confusion but we have to live with it. Another thing that merits mentioning is the dimensionality of the definition domain. For simplicity, this whole chapter deals with only the univariate case. But all the results obtained here can be extended straightforwardly to a multi-dimensional situation (e.g., functions defined on a plane) without difficulty. This fact won't be repeated when multivariate problems are encountered in later chapters.

A final note on the notation of this chapter. Except in the first section, subscripts are used to index locations for quantities like bases, energy, expansion coefficients, and so forth, while superscripts usually signify scale levels, unless otherwise clarified. This makes it easy to distinguish between indices of location and scale. A superscript, however, should not be confused with a an indexed quantity raised to a power. In case that potential confusion does arise, parentheses will be exploited to help clarify. Fortunately in this chapter only the power 2 will appear, as only squared (energy-like) quantities are of concern. In general, it should be clear in the context when a superscript and a power coexist.

2.1 Preliminaries from functional analysis

This section is an introduction of some terminologies and basic facts that will be used later from functional analysis. The first three subsections may trace their origins from Kreyszig (1989), Anderson (1996), and Young (1980). Their conventions of notation are therefore followed, perhaps with slight modification to ensure consistency.[2] The remaining four parts, which deal with infinite-dimensional spaces, are based mainly on Anderson (2001). Other literature may also be referred to but is not essential. Terms showing up but not introduced in advance can be found in the appendices of Kreyszig (1989).

Throughout this thesis, all the functions will be studied only in the setting of some Hilbert space. However, we begin this section with an introduction of Banach space (complete normed space), considering the concept of norm is needed in the following chapters. As a Hilbert space is also Banach space, results obtained in the latter certainly apply to the former. This fact is assumed whether specific clarification is made or not.

2.1.1 Banach space

Let $\mathbb{R}$ denote the field of real numbers. Suppose X is a vector (or linear) space over $\mathbb{R}$. A nonnegative mapping $\|\cdot\|$: $X \to \mathbb{R}$ defines a **norm** on X if, for any $x, y \in X$, and $\alpha \in \mathbb{R}$, it satisfies the following axioms:

1) $\|x\| \geq 0$.
2) $\|x\| = 0 \iff x = 0$.
3) $\|\alpha x\| = |\alpha| \cdot \|x\|$.
4) $\|x + y\| \leq \|x\| + \|y\|$ (triangle inequality).

A vector space endowed with a norm is called a **normed space**. In this context, it is denoted as a pair $(X, \|\cdot\|)$, or simply X if it is clear in the context.

[2] In this chapter, x will be used to denote a vector in a function space, and t the independent variable of x. That is to say, x could be velocity, density, or any state variable, while t could be time, or any dimension of the physical space, if applied to real geophysical fluid flow problems. In short, x and t in this chapter should not be confused with the physical space and time variables to appear later.

A function defined on the set of natural numbers, $\mathbb{N}$, is called a **sequence**. It is denoted in the form $\{x_n\}_{n\in\mathbb{N}}$, but the index set is often dropped for notational convenience. Later on in this thesis we also need to consider **bi-infinite sequences**, which are functions defined on the ring of integers, $\mathbb{Z}$. But until p. 32, we will restrict the indexing on $\mathbb{N}$.

A sequence $\{x_n\}$ in a normed space $(X, \|\cdot\|)$ is said to be **Cauchy** if for any $\varepsilon > 0$, there exists an integer $N_0 = N_0(\varepsilon) \in \mathbb{N}$, such that

$$\|x_m - x_n\| < \varepsilon, \qquad \forall\, m, n > N_0.$$

A normed space $(X, \|\cdot\|)$ is said to be **complete** if, for every Cauchy sequence $\{x_n\}_{n\in\mathbb{N}} \subset X$, there is an $x \in X$ such that

$$\lim_{n\to\infty} \|x_n - x\| = 0.$$

In other words, every Cauchy sequence is a convergent sequence, with x being the limit of the sequence: $x_n \to x$. A complete normed space is called a **Banach space**.

Any incomplete normed space can be completed. This is guaranteed by the Theorem of Completion which is disscussed on p. 69 of Kreyszig (1989).

2.1.2 Hilbert space

Let X be a vector space over the field $\mathbb{R}$. A mapping

$$g : X \times X \to \mathbb{R} \tag{2.1}$$

is called a **bilinear form** if, for all vectors x, y, $z \in X$, and α, $\beta \in \mathbb{R}$,

$$\begin{aligned} &1) \quad g(x+y,z) = g(x,z) + g(y,z) \\ &2) \quad g(x,y+z) = g(x,y) + g(x,z) \\ &3) \quad g(\alpha x, \beta y) = \alpha\beta \cdot g(x,y) \end{aligned}$$

An **inner product** on a vector space X over $\mathbb{R}$ is a positive definite and symmetric bilinear form. Denoted as $\langle\cdot,\cdot\rangle$, it is a mapping

$$\langle\cdot,\cdot\rangle : X \times X \to \mathbb{R},$$

satisfying the above axioms and having the following two additional properties:

$$4)\quad \langle x, y\rangle = \langle y, x\rangle$$

$$5)\quad \langle z, z\rangle > 0$$

for all $x, y, z \in X$, and $z \neq 0$. A vector space X over $\mathbb{R}$ equipped with an inner product $\langle\cdot,\cdot\rangle$ is called an **inner product space**. It is denoted as the pair $(X, \langle\cdot,\cdot\rangle)$, but a simple X is preferred if the context makes it clear.

An inner product space X is also a normed space, with the norm induced by the inner product through

$$\|x\| = [\langle x, x\rangle]^{1/2}, \quad \forall\, x \in X.$$

The concept of completeness introduced above then also applies here. A complete inner product space is called a **Hilbert space**. Apparently, if a space is Hilbert, it is also Banach.

An inner product space is not just a normed space, however. It is endowed with more geometric notions. Consider $(X, \langle\cdot,\cdot\rangle)$, an inner product space over $\mathbb{R}$. For $x, y \in X$, $x, y \neq 0$, we have the following **Cauchy-Schwarz Inequality**:

$$-1 \leq \frac{\langle x, y\rangle}{\|x\| \cdot \|y\|} \leq 1.$$

This allows a definition of an angle θ between vectors x and y, the cosine of which is measured by $\langle x, y\rangle / \|x\| \|y\|$. When $\theta = \pm 90^{\circ}$, or equivalently,

$$\langle x, y\rangle = 0, \qquad (x, y \text{ could then be zero})$$

x and y are said to be **orthogonal**.

One example of inner product space is $\mathbb{R}^n$. It is spanned (cf. p. 31) by n linearly independent vectors and hence has a finite dimension n. The inner product is defined as

$$\left\langle \underline{\mathbf{x}}, \underline{\mathbf{y}} \right\rangle = \sum_r x_r y_r = \underline{\mathbf{x}}^T \underline{\mathbf{y}}, \tag{2.2}$$

for any

$$\underline{\mathbf{x}} = (x_1, ..., x_n)^T, \underline{\mathbf{y}} = (y_1, ..., y_n)^T \in \mathbb{R}^n.$$

$\mathbb{R}^n$ is complete. It is a Hilbert space.

Another example of inner product space is $C_2[a, b]$, $a < b$, which consists of all real-valued continuous functions $x(t)$ defined on $[a, b]$, and equipped with the inner product

$$\langle x, y \rangle = \int_a^b x(t) y(t) \, dt, \tag{2.3}$$

for all $x(t)$ and $y(t)$ in it. $C_2[a, b]$ is of infinite dimensionality. It is not complete. The completion (Kreyszig, p. 41) is called $L_2[a, b]$. When the definition domain is extended to the whole $\mathbb{R}$, we obtain a space called $L_2(\mathbb{R})$. It contains all those functions x such that $\int_{\mathbb{R}} |x(t)|^2 \, dt$ converges. In other words, $L_2(\mathbb{R})$ is a Hilbert space where lie functions square integrable over the real line $\mathbb{R}$. Note the real line may be replaced by more general a manifold. Sometimes we shall need to consider the square integrable functions defined on a torus. A **torus** $\mathbb{T}_\varrho$, $\varrho > 0$, is the set of equivalence classes of $\mathbb{R}$ with two points on $\mathbb{R}$ identified if they differ by $\varrho \ell$, for any $\ell \in \mathbb{Z}$. (In this 1-D case, a torus can be identified with a circle with radius ϱ.) The function space here is denoted as $L_2(\mathbb{T}_\varrho)$.

2.1.3 Subspace

Given a normed space X, a **subspace** Y is a nonempty subset which, when equipped with a norm inherited from X (i.e., obtained by restricting the norm of X to it), also constitutes a normed space. The subspace of an inner product space can be defined similarly, except that now it is the inner product that is to be restricted.

Subspaces of a complete normed (Banach) or inner product (Hilbert) space need not be complete by themselves. The completeness requires the concept of closedness.

Given a normed space $X = (X, \|\cdot\|)$, a subset $Y \subset X$ is said to be **closed** in X if it contains limits of all convergent sequences. Y as a subspace is closed if it is closed as a subset in X. The smallest closed set that contains Y is called the **closure** of Y. It is written as $\mathrm{Cl}\, Y$. Y is closed if and only if $Y = \mathrm{Cl}\, Y$. A subspace of a Banach/Hilbert space is complete if and only if it is closed (see Kreyszig, 1989. pages 67 & 140).

Let Y and Z be two subspaces of a Hilbert space $X = (X, \langle\cdot,\cdot\rangle)$. They are said to be orthogonal (denoted as $Y \perp Z$) if for each $y \in Y$, $z \in Z$,

$$\langle y, z \rangle = 0.$$

The **orthogonal complement** of subspace Y in X, written $Y^\perp$, is

$$Y^\perp = \{z \in X \mid z \perp Y\}$$

If Y is closed, X can be decomposed into the **direct sum**[3] of Y and its orthogonal complement, *i.e.*,

$$X = Y \oplus Y^\perp$$

by the theorem of direct sum (Kreyszig, p. 146): that is, for any $x \in X$, there is a unique representation

$$x = y + z, \qquad \text{with } y \in Y \text{ and } \ z \in Y^\perp.$$

In this case, $y \in Y$ is called the **orthogonal projection** of $x \in X$ on $Y \subset X$.

2.1.4 Totality and orthonormality

Given a normed space $X = (X, \|\cdot\|)$, a set Y_ε is called an **ε-net** for $Y \subset X$ if for each $y \in Y$, there exists a $y_\varepsilon \in Y_\varepsilon$, such that $\|y - y_\varepsilon\| < \varepsilon$. A subset $M \subset X$ is said to be **dense** in X iff for all $\varepsilon > 0$, M is an ε-net of X.

Given a nonempty set M, we may define a **span** of it, denoted as span$M$. If M is finite, its span is made by all the linear combinations of vectors in M; if M is infinite, span$M$ is defined by the union of all the spans of its finite subsets.[4] A vector space X is said to be spanned or generated by $M \in X$ if $X = \text{span}M$. A **total** set in a normed space X is a subset $M \subset X$ such that the span of it is dense in X, i.e., Cl (span$M$) $= X$.

For an inner product space X, $M \subset X$ is said to be orthonormal if, for every pair $x, y \in M$,

$$\langle x, y \rangle = \begin{cases} 1, & x = y \\ 0, & x \neq y \end{cases}$$

It is called a **total orthonormal set** if span$M$ is dense in X. Every Hilbert space $X \neq \{0\}$ has at least one such total orthonormal set/system M (Yosida, 1974, p. 86). If M is countable, it can be made into an indexed family $\{x_j\}_{j \in \mathbb{Z}} \subset X$, or a sequence in X. Totality and orthonormality also apply to sequences.

[3] A direct sum of two subspaces X and Y does not require orthogonality between them, however.

[4] This definition doesn't invoke any issues of convergence. See Anderson, 2001.

2.1.5 Basis

A finite set $M = \{v_0, v_1, ..., v_n\}$ is said to be **linearly independent**, if equation

$$\sum_{j=0}^{n} \alpha_j v_j = 0$$

has no non-trivial solution for α_j: that is, all α_j's must be zero. For a finite-dimensional vector space $X \neq \{0\}$, a linearly independent subset $B \subset X$ is called a basis of X if span$B = X$. A basis is a minimal spanning set in the sense that removal of one element invalidates span$B$=X. A basis is also a maximal linearly independent set. Addition of any vector destroys its linear independence.

For infinite-dimensional vector spaces, the concept of basis is no longer as simple as its finite-dimensional counterpart. It is not our intention to give the details here. For a careful and exhaustive treatment, the reader is referred to Anderson (2001). We restrict our discussion to Hilbert space only.

We now change the indexing for sequences from the set of natural numbers, $\mathbb{N}$, to the ring of integers, $\mathbb{Z}$. We will continue to refer to them as sequences, but from now on all of them are assumed to be endowed with the bi-infinite notion.

According to Anderson (2001), a basis for an infinite-dimensional Hilbert space X is a minimal total set of X. Particularly, a **Schauder basis** for it is a sequence of vectors $\{e_n\} \subset X$, such that for every $x \in X$, there exists a unique coefficient sequence $\alpha_n \in \mathbb{R}$ satisfying

$$\left\| x - \sum_{i=-n}^{n} \alpha_i e_i \right\| \longrightarrow 0 \qquad \text{as } n \to \infty,$$

and, a **Riesz basis** (or stable, unconditional basis) $\{e_n\}$ for X is a Schauder basis such that the coefficients α_i, for all $x \subset X$, is bounded above and below in the sense that[5]

$$A\|x\|^2 \leq \sum_{i\in\mathbb{Z}} |\alpha_i|^2 \leq B\|x\|^2, \tag{2.4}$$

with constants $B \geq A > 0$.

Riesz bases are equivalent to orthonormal bases (Young, pp. 31), in the sense that there exists an invertible bounded operator between a Riesz basis and an orthonormal basis. From now on, all bases are assumed to be orthonormal, (or at least, Riesz), unless otherwise indicated.

[5]This definition follows Strang and Nguyen (1996). Other definitions may be found in Young (1980) on pp. 31-32, and Anderson (2001).

2.1.6 The expansion property for a Hilbert space

Let X be an infinite-dimensional Hilbert space with an orthonormal basis $\{e_j\}$. According to Anderson (2001), each $x \in X$ can be represented as

$$x = \sum_j \alpha_j e_j, \tag{2.5}$$

with $\{\alpha_j\}$ ℓ^2-summable, i.e., $\sum_j |\alpha_j|^2 < \infty$. This is called the **expansion property** by Anderson (2001), with **expansion coefficients**

$$\alpha_j = \langle x, e_j \rangle . \tag{2.6}$$

If the expansion $\sum_j \alpha_j e_j = 0$, all α_j are then zero, a property shown in Anderson (2001) which is the analog of linear independence in a finite dimensional case.

From Eq. (2.6) we may identify an operator $\mathcal{T}$ from X to the space ℓ^2 (space of sequences that are ℓ^2-summable; see Kreyszig, 1989):

$$\mathcal{T}x \longmapsto \{\langle x, e_j \rangle\}, \qquad \forall\, x \in X. \tag{2.7}$$

This is the **analysis** or **transform** of x with respect to $\{e_j\}$. Eq. (2.5), on the other hand, defines another operator

$$\mathcal{R} : \ell^2 \longrightarrow X; \qquad \mathcal{R}\{\alpha_j\} \longmapsto \sum_j \alpha_j e_j. \tag{2.8}$$

$\mathcal{R}$ is called a **synthesis** or **reconstruction** of $x \in X$ with respect to basis $\{e_j\}$, and expansion coefficients $\{\alpha_j\} \subset \ell^2$. Again here $\{e_j\}$ is orthonormal, and in this case the expansion coefficients $\{\alpha_j\}$ are particularly given a name, **Fourier coefficients**. The collection of all the Fourier coefficients also forms a Hilbert space. It is equal to ℓ^2, and we shall refer to it as a **phase space**.

2.1.7 Parseval relation

Suppose X is a Hilbert space, and $\{e_j\} \subset X$ an orthonormal set. For $x \in X$, let $\{\alpha_j\} = \langle x, e_j \rangle$. We then have the **Bessel inequality** (see Kreyszig, p. 157)

$$\sum_j (\alpha_j)^2 \equiv \sum_j |\langle x, e_j \rangle|^2 \leq \|x\|^2. \tag{2.9}$$

Equality holds for every x if and only if $\{e_j\}$ is total in X. Stating in another way, if $\{e_j\}$ is an orthonormal basis of X, then

$$\sum_j (\alpha_j)^2 = \|x\|^2. \tag{2.10}$$

This is the so-called **Parseval relation** or **Parseval equality**. It reads that the squared Fourier coefficients sum to the corresponding squared norm. In a careful treatment, Anderson (2001) has established the equivalence between the expansion property (2.5), totality, and the Parseval equality for an infinite-dimensional Hilbert space.

For any vector in a Hilbert space, Parseval relation sets up a connection between its representations in the physical and phase spaces. As we will see later, it forms the basis of multiscale energy studies.

2.2 Multiscale window transform

2.2.1 Multi-resolution analysis

By a **multi-resolution analysis** of $L_2(\mathbb{R})$, we mean a sequence of *closed subspaces* $\{V_j\}_{j\in\mathbb{Z}}$, such that (Hernández and Weiss, 1996; Meyer, 1990)[6]

(1) $\ldots V_{-1} \subset V_0 \subset V_1 \subset V_2 \ldots$. (**Nestedness**)

(2) $\mathrm{Cl} \bigcup_{j\in\mathbb{Z}} V_j = L_2(\mathbb{R})$. (**Totality**)

(3) $\bigcap_{j\in\mathbb{Z}} V_j = \{0\}$. (**Emptiness**)

(4) There exists a translation invariant (affine) orthonormal set[7] $\{\phi(t-n),\ n \in \mathbb{Z}\}$ which is total in V_0.

(5) $x(t) \in V_j$ iff $x(2t) \in V_{j+1}$ (**Refinability**).

[6] In Meyer's original definition, the index of V_j runs in the opposite direction, namely, V_j is associated with basis function $\phi(2^{-j}t)$ rather than $\phi(2^j t)$. This notation is also seen in other literature such as Holschneider (1995). We will follow Strang and Nguyen (1997)'s convention (also seen in Louis et al. (1997) and Wojtaszczyk (1997)).

[7] That is to say, an orthonormal set formed by shifting the independent variable by integers.

The translation invariant set in condition (4) forms a space which we will refer to as a **sampling space**. In a formal language, a sampling space is, for a Hilbert space X, a closed subspace $H \subset X$ in which all the functions can be expanded with respect to a translation invariant orthonormal basis. In some literature (e.g., Holschneider, 1995), the "translation invariant orthonormal basis" is relaxed to be "translation invariant Riesz basis".

Conditions (4) and (5) combine to imply that every **multi-resolution approximation** of $L_2(\mathbb{R})$, V_j, $j \in \mathbb{Z}$, is a sampling space. Condition (5) also implies that the subspaces V_j thus-constructed are self-similar. For $j > 0$, it has a compressed orthonormal basis $\{2^{j/2}\phi(2^j t - n), n \in \mathbb{Z}\}$ ($2^{j/2}$ the normalization factor), while if $j < 0$, the basis function is enlarged, with the same form.

Scaling decomposition in a multi-resolution analysis is attained by considering the *orthogonal complements* of V_j in V_{j+1}, W_j, in the sense that $W_j \perp V_j$ and

$$V_j \oplus W_j = V_{j+1}, \qquad \forall\, j \in \mathbb{Z}.$$

If V_j has an orthonormal basis $\{2^{j/2}\phi(2^j t - n), n \in \mathbb{Z}\}$, called a **scaling basis** (or **dilation basis**), then it can be proved W_j is also a sampling space and has an orthonormal basis in the form $\{2^{j/2}\psi(2^j t - n), n \in \mathbb{Z}\}$ called a **wavelet basis** (Holschneider, 1995, p. 205).[8] The whole $L_2(\mathbb{R})$ is now decomposed as

$$\bigoplus_{j \in \mathbb{Z}_+} W_j \oplus V_0 = \bigoplus_{j \in \mathbb{Z}} W_j = L_2(\mathbb{R}), \tag{2.11}$$

where $\mathbb{Z}_+ = \{j \mid j \in \mathbb{Z}, j \geq 0\}$ is the set of nonnegative integers. The relationship between each W_j and V_j is schematized in Fig. 2.1. Note subspace W_j embodies features of only one scale 2^{-j}, or **scale level** j, in contrast to V_j, where all scales below level j are included. For this reason, features on a scale level j (or simply scale j, if no confusion arises) are represented by vectors in W_j. In this decomposition, W_i is orthogonal to W_j for all $i \neq j$, $i, j \in \mathbb{Z}$. We obtain consequently an orthonormal set $\{2^{j/2}\psi(2^j t - n)\} \equiv \{\psi_n^j\}$, $j, n \in \mathbb{Z}$. ($2^{j/2}$ the normalization factor) whose span is dense in $L_2(\mathbb{R})$.

[8] The functions $\phi(t)$ and $\psi(t)$, $t \in \mathbb{R}$ are then called a **scaling** or **dilation function** and a **wavelet function**, respectively. In some literature, such as Meyer (1990), they are also referred to as a *father wavelet* and *mother wavelet*.

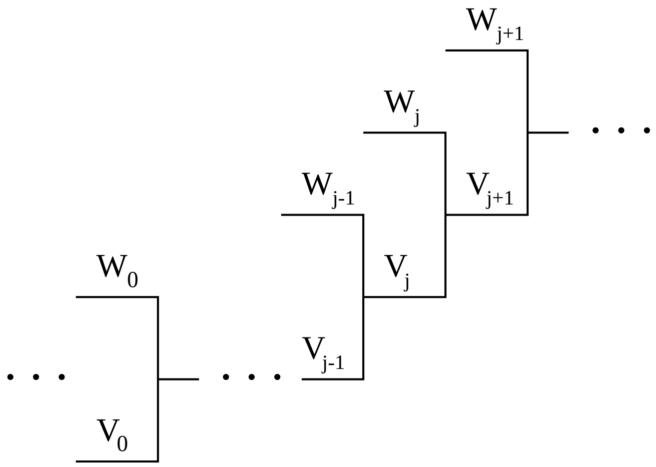

Figure 2.1: A schematic of the multi-resolution analysis, $\{..., V_{j-1}, V_j, V_{j+1}, ...\}$. W_j is the orthogonal complement of V_j in V_{j+1}, and $W_j \oplus V_j = V_{j+1}$, $V_j \perp W_j$.

2.2.2 Multiscale analysis and energy on multiple scales

By multiscale analysis we mean an analysis with respect to some basis with indices corresponding to scales. In the multi-resolution analysis setting, it is just another name for wavelet analysis. Given $x \in X \subset L_2(\mathbb{R})$, let α_n^j be the transform of it with respect to $\{\psi_n^j\}$. As the basis is orthonormal, the Parseval relation states that

$$\sum_j \sum_n \left(\alpha_n^j\right)^2 = \|x\|^2 = \int_{\mathbb{R}} (x(t))^2 \ dt. \tag{2.12}$$

This relation allows a definition of energy on a scale-location mesh. In this example, $E(t) = x^2(t)$ is a generalized energy at some t. In geophysical fluid dynamics, if the vector x is velocity, E is simply the kinetic energy except for a constant multiplier; if the vector is density anomaly ρ, then E is the available potential energy except for a factor $\frac{1}{2}\frac{g^2}{\rho_0 N^2}$ with N the Brunt-Väisälä frequency. The integration, $\int_{\mathbb{R}} E(t)dt$, may have a different interpretation depending on the independent variable t is chosen. If t is a spatial coordinate, then $\int_{\mathbb{R}} E(t)dt$ is simply the total energy over that dimension; if t is time, it can be understood as the energy integrated over the time domain. Either way $\int_{\mathbb{R}} E(t)dt$ boils down to a quadratic quantity associated with a notion of energy.

On the left hand side of Eq. (2.12) is a sum of indexed quantities. They are tagged with the corresponding scales and locations. For every j and n, we may then square the Fourier coefficient to make an energy, i.e., to define $(\alpha_n^j)^2 \equiv E_n^j$, to be the **energy** on scale j and location n, in a generalized sense as explained above.

Note E_n^j depends on n, the location index, which is lacked in the classical representation. This dependence comes from sampling space-generating basis $\{\psi_n^j\}$. It provides the information of locality in the multiscale energy representation.

Besides E_n^j, a quantity that measures the total energy on scale j can also be defined:

$$E^j = \sum_{n \in \mathbb{Z}} E_n^j. \tag{2.13}$$

We will call it the **marginal energy** corresponding to that scale.

2.2.3 Scale windows and energy on scale windows

A faithful representation of the scale windowing phenomenon is a requirement for an adequate MS-EVA formulation. A scale window, in a plain language, is just a subspace of some Hilbert space with only a designated window of scales involved. Its rigorous definition may vary with different notions of scale. Here we define it in the setting of multi-resolution analysis.

Given a multi-resolution approximation $V_{j_2} \subset L_2(\mathbb{R})$ with $j_2 > 0$ some known integer, a **scale window**, or simply a **window**, between scale levels j_a and j_b, with $j_a < j_b \le j_2$ is defined to be a subspace $V^{a\text{-}b} \subset V_{j_2}$ such that

$$V^{a\text{-}b} = W_{j_a} \oplus W_{j_a+1} \oplus ... \oplus W_{j_b}. \tag{2.14}$$

In other words, it is a subspace of V_{j_2} generated by a subfamily of $\{\psi_n^j(t),\ j \le j_2, n \in \mathbb{Z}\}$, $\{\psi_n^j(t),\ j \in J, n \in \mathbb{Z}$, with $J = \{j_a,\ j_a+1,\ ...,\ j_b\}$, where $j_a < j_b \in \mathbb{Z}$ correspond to the upper and lower scale bounds. With this definition, we construct three scale windows,[9] V^L, V^M, and V^S. They are

[9]Arbitrarily many scale windows may be constructed. We build three because we will use only three in the upcoming MS-EVA formulation.

Scale window	Notation	Generating set	
Large scale window	V^L	$\{\psi_n^j\}_{j\in J_0, n\in\mathbb{Z}}$	$J_0 = \{j \in \mathbb{Z} \mid j \leq j_0\}$
Mesoscale window	V^M	$\{\psi_n^j\}_{j\in J_1, n\in\mathbb{Z}}$	$J_1 = \{j \in \mathbb{Z} \mid j_0 < j \leq j_1\}$
Sub-mesoscale window	V^S	$\{\psi_n^j\}_{j\in J_2, n\in\mathbb{Z}}$	$J_2 = \{j \in \mathbb{Z} \mid j_1 < j \leq j_2\}$

Accordingly, V_{j_2} can be decomposed into a direct sum of these scale windows

$$V_{j_2} = V^L \oplus V^M \oplus V^S,$$

with V^L, V^M, and V^S *mutually orthogonal.* For each vector $x \in V_{j_2}$, there is a unique representation

$$x = x^L + x^M + x^S,$$

where x^L, x^M, and x^S are vectors in scale windows V^L, V^M, and V^S, respectively. Let E^j denote the marginal energy for scale level j. The marginal energies of x associated with these windows are, by definition,

$$E^L = \sum_{j\leq j_0} E^j, \tag{2.15}$$

$$E^M = \sum_{j=j_0+1}^{j_1} E^j, \tag{2.16}$$

$$E^S = \sum_{j=j_1+1}^{j_2} E^j. \tag{2.17}$$

The Parseval relation then states that, for any $x \in V_{j_2}$,

$$E^L + E^M + E^S = \|x\|^2, \tag{2.18}$$

i.e., energy is conserved with this decomposition.

The above energy expression requires *a priori* knowledge of the orthonormal basis $\{\psi_n^j\}$. Sometimes, however, this may be unnecessary. An expression of the energy alternatively in physical space could be more convenient. Recalling how V^L, V^M, and V^S are constructed, we have, for any $x \in V_{j_2}$, with $n \in \mathbb{Z}$,

$$\alpha_n^j = \left\langle x, \psi_n^j \right\rangle = \begin{cases} \left\langle x^L, \psi_n^j \right\rangle & j \leq j_0 \\ \left\langle x^M, \psi_n^j \right\rangle & j_0 < j \leq j_1 \\ \left\langle x^S, \psi_n^j \right\rangle & j_1 < j \leq j_2 \end{cases} \tag{2.19}$$

As $\{\psi_n^j,\ j \leq j_0, n \in \mathbb{Z}\}$ furnishes a total orthonormal set for V^L, the Parseval relation implies

$$\sum_{j \leq j_0} \left| \left\langle x^L, e^j \right\rangle \right|^2 = \|x^L\|^2.$$

By (2.19), this is,

$$E^L = \sum_{j \leq j_0} E^j = \|x^L\|^2. \tag{2.20}$$

Likewise,

$$E^M = \|x^M\|^2 \tag{2.21}$$

$$E^S = \|x^S\|^2 \tag{2.22}$$

Equations (2.20), (2.21), and (2.22) are the three parts of energy for $x \in V_{j_2}$ represented in the physical space.

We now examine what is required for a proper scale window decomposition. This is useful when approaches other than multi-resolution analysis are considered; plus, it also serves to verify what we have already obtained. Remember our objective is to have a function decomposed into three parts, each part characteristic of a window of scales. In this context, this is to say that we need to have a Hilbert space V_{j_2} decomposed as

$$V_{j_2} = V^L \oplus V^M \oplus V^S.$$

This decomposition is for the MS-EVA formulation. The first requirement is thus with energy. Energy must be conserved in the sense of (2.18); otherwise interactions in terms of energy exchange between processes defined in different subspaces wouldn't make any sense. This energy conservation requirement should not only apply to the three windows as a whole, but to any two of them also. With the notation adopted above, this is to say, we must have

$$\|x^L\|^2 + \|x^M\|^2 = \|x^L + x^M\|^2 \tag{2.23}$$

$$\|x^M\|^2 + \|x^S\|^2 = \|x^M + x^S\|^2 \tag{2.24}$$

as well as

$$\|x^L\|^2 + \|x^M\|^2 + \|x^S\|^2 = \|x\|^2, \tag{2.25}$$

for any $x^L \in V^L$, $x^M \in V^M$, $x^S \in V^S$, $x \in V_{j_2}$, and $x = x^L + x^M + x^S$. It is easy to prove that (2.23) - (2.25) hold if and only if subspaces V^L, V^M, and V^S are mutually orthogonal. In the forward direction, this is trivially true. Conversely, for all $x^L \in V^L$, $x^M \in V^M$, (2.23) implies

$$\left\langle x^L, x^M \right\rangle = 0,$$

or in other words,

$$V^L \perp V^M.$$

Similarly we also have $V^M \perp V^S$, and $V^L \perp V^S$. Consequently, to have energy conserved is equivalent to demanding mutual orthogonality between the three subspaces.

The energy conservation requirement excludes some extensively used approaches from our consideration for the MS-EVA formulation. Objective analysis (OA) is just such an example.[10] In forming (2.14), mutual orthogonality is guaranteed by the multi-resolution analysis. Energy conservation is not a problem here.

Another requirement for the scale window decomposition concerns the multiscale energy localization (see below for a definition of localization). As will be clear soon, a subspace of some sampling space with a properly localized translation invariant basis contains functions which will yield an energy satisfying the requirement. In our definition with the multi-resolution analysis, V^L, V^M, and V^S are all subspaces of the sampling space V_{j_2}. The remaining problem is then the basis localization, and this is the subject of the next subsection.

2.2.4 Localization

By the definition from the preceding subsection, energy is unfolded on a scale-location plane in the phase space. The information of locality is represented by the location coordinate of that plane. But this is not enough for the MS-EVA purpose. We need introduce the concept of localization.

[10] One only needs to check the non-perpendicularity between an estimator vector and the residual to it. Refer to Appendix A5.2 for an introduction of OA.

A function $f(t)$ is said to have a **polynomial localization** about the origin (or simply **localization**) of order γ if

$$|f(t)| \leq \frac{C}{(1+|t|^2)^{\gamma/2}} \equiv \kappa_\gamma^c(t), \qquad \forall\, t \in \mathbb{R} \tag{2.26}$$

for $\gamma > 0$, where C is some positive constant. Parameter γ describes how fast the function decays with increasing $|t|$. The larger the γ, the more localized the function $f(t)$. We consider $\gamma > 1$ only.

As $|t| \to \infty$, $\kappa_\gamma^c(t)$ approaches to $\frac{C}{|t|^\gamma} \equiv \pi_\gamma^c(t)$. With this $\pi_\gamma^c(t)$, we may introduce a quantity called ε-length to measure the effective support of a family of localized functions. Mathematically, an **ε-length** $L_\varepsilon > 0$ is defined, for sufficiently small positive ε, such that

$$\int_{|t|>L_\varepsilon} \pi_\gamma^c(t)\ dt = \varepsilon. \tag{2.27}$$

The integration on the left hand side converges for $\gamma > 1$, and,

$$L_\varepsilon = \left(\frac{2C}{\gamma - 1}\right)^{\frac{1}{\gamma-1}} \cdot \varepsilon^{-\frac{1}{\gamma-1}}. \tag{2.28}$$

Apparently L_ε depends on C and γ as well as ε; but we will suppress these two dependences if it is clear in the context. If a function $f(t)$ is polynomially localized as (2.26), then

$$\int_{|t|>L_\varepsilon>0} |f(t)|\ dt \leq \int_{|t|>L_\varepsilon} \kappa_\gamma^c(t)\ dt < \int_{|t|>L_\varepsilon} \pi_\gamma^c(t)\ dt = \varepsilon. \tag{2.29}$$

In this case, we say $f(t)$ is **effectively supported** on $[-L_\varepsilon, L_\varepsilon]$ up to ε.

Other types of decay are also possible: exponential localization is just such an example (e.g., Holschneider, 1995). But here the polynomial localization (2.26) is enough, since more often than not we need only study our problems qualitatively. An exponentially localized function may always be made smaller than some κ_γ^c.

With the polynomial localization, it is possible to examine whether and how the energy on a scale-location plane is localized. Recall the MS-EVA problem in this regard concerns the localization of energy for localized events. In other words, a localized event in physical space should have energy locally represented in phase space also. This can be achieved if the analysis bases, i.e., the translation invariant bases $\{\phi_n^j\}$ and $\{\psi_n^j\}$ introduced above, are localized. In fact, the Theorem 11.0.2 of Holschneider (1995) states that, for $|x(t)| \leq \kappa_\gamma^c(t)$, $\gamma > 1$, if the basis function φ (ϕ or ψ) is such that $|\varphi(t)| \leq \kappa_\gamma^c(t)$, then

$$|\alpha_n^j| \leq C'\kappa_\gamma^c\left(\frac{2^{-j}n}{1+2^{-j}}\right) \tag{2.30}$$

for some constant $C' = C'(\gamma, j) < \infty$, where $\alpha_n^j = \langle x, \varphi_n^j \rangle$. By Theorem 11.0.1 of the same source, we then have

$$E_n^j = |\alpha_n^j|^2 \leq C'' \cdot \kappa_\gamma^c\left(\frac{2^{-j}n}{1+2^{-j}}\right), \tag{2.31}$$

where C'' is some other constant independent of n, and $2^{-j}n$ the location in the physical space. E_n^j is therefore polynomially localized around $2^{-j}n$.

Consequently, as long as φ (ϕ or ψ) is localized, the multiscale energy localization is guaranteed. The only problem now is how localized they need to be, and this will be clear later in terms of requirements on γ. From now on, any scaling or wavelet function is tacitly assumed to have some localized structure of a desired order.

2.2.5 Multiscale window transform

Scale window construction

Scale windows are defined in the multi-resolution analysis setting (2.11) with direct sums of W_j (see (2.14)). The advantage of this definition is clear: It allows one to identify with ease any scale windows of interest on a wavelet power spectrum (e.g., Kumar and Foufoula-Georgian, 1997). However, one will encounter two difficulties when attempting to calculate from (2.14) the energy for a window at a specific location.

The first difficulty is resolution. Multi-resolution analysis earns its name from the relation between scale level and resolution, as is easily seen in the form of basis functions $\psi(2^j t - n)$ and $\phi(2^j t - n)$. As shown in Fig. 1.2b, the larger the scale (or the lower the scale level), the lower the resolution. Built with these W_j generated by $\{\psi(2^j t - n), n \in \mathbb{Z}\}$, the subspaces V^L, V^M, and V^S will then have different resolutions, with V^S the highest, and V^L the lowest. When a signal is short,[11] as is usually the case in a real-time forecast, the largest scale is more

[11] Here an extension of the finite signal is needed; but we will postpone the discussion of this to § 2.3.

often than not the duration or basin scale, which yields the coarsest resolution - only 1 (or 2, depending upon how the location index is arranged) energy value on the whole definition domain. This obviously violates the MS-EVA spirit, degenerating to something similar to that in the classical mean-deviation formalism. This is the first problem we have to fix before further actions are taken.

The second difficulty is encountered in building the scale windows. It is also related to resolution. It is stated independently because the problem itself is fundamentally a different one. Recall that a scale window is generated by a cluster of shift-invariant basis functions $2^{j/2}\psi(2^j t - n)$ involving many scale levels. The corresponding energy thus requires a cross-scale summation of the square of Fourier coefficients. The different resolution at different scale level, however, essentially makes this kind of summation impossible. In the example of Fig. 1.2b, energy takes value at only one location for the lowest level, which is at the center of the time domain as shown. Above one level, it takes values at two locations, and one more level higher, four, and so forth. These locations do not coincide in the time domain, and values taken on them hence cannot be summed together. No interpolation should be made since they are now in ℓ_2, rather than $L_2(\mathbb{R})$.

In order to get rid of the above difficulties, the key is to fix the resolution with scale j. However, there is no way to claim a higher resolution for the Fourier coefficients simultaneously in the physical space and phase space. Raising one is at the cost of the other, as stated clearly in Heisenberg's uncertainty principle (see Strang and Nguyen, 1997). What is schematized in Fig. 1.2 reflects an aspect of this problem.

Consequently, an alternative representation of scale windows in the spirit of (2.14), but with the above difficulties overcome, is necessary. Forget for the time being the wavelet transform and look back to the multi-resolution analysis $\{V_j\}$. For each $j \in \mathbb{Z}$, subspace V_j is generated by $\{\phi(2^j t - n), n \in \mathbb{Z}\}$. It encompasses all the scales with level below j, which means V_j itself is a scale window. Particularly, given j_0 the upper bound of scale window V^L, we may take

$$V^L = V_{j_0}. \tag{2.32}$$

Its basis in the form $\{2^{j_0/2}\phi(2^{j_0}t - n), n \in \mathbb{Z}\}$ guarantees that it is a sampling space. In doing this, we completely avoid taking sums across scales to obtain the energy for a scale window at a specific location. Besides, the so-obtained window has a basis with the highest physical space resolution to which its scales correspond (2^{-j_0} here).

The construction of V^M and V^S follows the same track as that of V^L, except for some minor modifications. Suppose that j_1 and j_2 are the upper limits[12] of scale levels for V^M and V^S, respectively. Consider the orthogonal complement of V_{j_0} in $V_{j_1} \subset V_{j_2} \subset L_2(\mathbb{R})$, written as $V_{j_1} \ominus V_{j_0}$. This subspace encompasses only features with scale level between $j_0 < j \leq j_1$, and is orthogonal to the large-scale window V^L constructed previously. Now the problem concerns with localization, but this is not in question since it is a subspace of V_{j_1}, a sampling space with an orthonormal and localized translation invariant basis $\{2^{j_1/2}\phi(2^{j_1}t - n), n \in \mathbb{Z}\}$. It is then reasonable to choose

$$V^M = V_{j_1} \ominus V_{j_0}. \tag{2.33}$$

Likewise,

$$V^S = V_{j_2} \ominus V_{j_1}. \tag{2.34}$$

Again, it is a subspace of a sampling space (V_{j_2}), and it is orthogonal to both V^L and V^M, meeting the conservation constraint, as well as the localization requirement if the scaling basis is localized to some desired order.

The story is not complete as yet. The three scale windows thus-formed still bear bases with different physical space resolution, of which V^L is no doubt the worst. We need to have the lower resolutions increased.

Notice that all the scale windows are subspaces of V_{j_2}. Vectors in them can definitely be represented uniquely with respect to the orthonormal basis $\{2^{j_2/2}\phi(2^{j_2}t - n), n \in \mathbb{Z}\}$. The Fourier coefficients (and hence the energy) obtained from the transform with this basis will have the highest resolution admissible in the physical space (2^{-j_2} here). This applies to all the windows generated above, and the resolution is not a problem any more.

[12]The upper bound of scale level j_2 for the sub-mesoscale window V^S corresponds to the smallest scale that the series has (here 2^{-j_2}). Theoretically it should be $+\infty$, but in practice we may assume there exists a finest resolution. We will justify this in §2.4.

Multiscale window transform

With the above preparation, we are now able to introduce a new analysis tool: the **multiscale window transform/analysis**, or merely **scale window transform/analysis**.

Given a multi-resolution approximation V_{j_2}, and a function $p(t) \in V_{j_2}$, $t \in \mathbb{R}$. Suppose V_{j_2} is spanned by an orthonormal scaling basis

$$\left\{\phi_n^{j_2}(t)\right\}_{n\in\mathbb{Z}} \quad = \quad \left\{2^{j_2/2}\phi(2^{j_2}t - n),\ n \in \mathbb{Z}\right\},$$

where $2^{j_2/2}$ is the normalized multiplier. The three scale windows, V^L, V^M, and V^S, are defined as in §2.2.3 such that they correspond to intervals of scale levels $(-\infty, j_0]$, $(j_0, j_1]$, $(j_1, j_2]$, respectively. All these windows are subspaces of V_{j_2}.

Write the scaling transform (transform with the scaling basis) of p as

$$\widehat{p}_n^j = \int_{\mathbb{R}} p(t)\ \phi_n^j(t)\ dt. \tag{2.35}$$

The caret is from now on used for transforms only, with different subscripts and/or superscripts signifying different types of transforms. In this definition, j symbolizes scale while n is location. Likewise, a wavelet transform with $\{\psi_n^j\}$ is written

$$\widetilde{p}_n^j = \int_{\mathbb{R}} p(t)\ \psi_n^j(t)\ dt \tag{2.36}$$

for later use.

With the caret notation we define three *operators*: large-scale, meso-scale, and sub-mesoscale window syntheses. The **large-scale window synthesis** (or merely **large-scale synthesis**) denoted in the superscript form $[]^{\sim 0}$: $V_{j_2} \to V^L$, is such that

$$p^{\sim 0}(t) = \sum_{n\in\mathbb{Z}} \widehat{p}_n^{j_0}\ \phi_n^{j_0}(t). \tag{2.37}$$

It keeps only the large-scale feature in $p(t)$. Similarly, the **meso-scale window synthesis** (or **meso-scale synthesis**) $[]^{\sim 1}$ is an operator from V_{j_2} to V^M:

$$p^{\sim 1}(t) = \sum_{n\in\mathbb{Z}} \widehat{p}_n^{j_1} \phi_n^{j_1}(t) - p^{\sim 0}(t), \tag{2.38}$$

and the **sub-mesoscale window synthesis** (or **sub-mesoscale synthesis**) $[]^{\sim 2}$ from V_{j_2} to V^S performs an operation such that

$$p^{\sim 2}(t) = p(t) - \sum_{n\in\mathbb{Z}} \widehat{p}_n^{j_1} \phi_n^{j_1}(t). \tag{2.39}$$

Scale window transforms are introduced with the aid of these syntheses. A **large-scale window transform/analysis** is an operator

$$\mathcal{T}_{0,n} : V_{j_2} \longrightarrow \ell_2; \qquad p \longmapsto \left\{ \int_{\mathbb{R}} p^{\sim 0} \phi_n^{j_2} dt, \ n \in \mathbb{Z} \right\}. \tag{2.40}$$

Similarly the **meso-scale window transform/analysis** $\mathcal{T}_{1,n}$ and **sub-mesoscale window transform/analysis** $\mathcal{T}_{2,n}$ can be defined, with $p^{\sim 0}$ in (2.40) replaced by $p^{\sim 1}$ and $p^{\sim 2}$, respectively. For mnemonic reasons, we also use carets as in (2.35) to denote these window transforms

$$\widehat{p}_n^{\sim w} = \int_{\mathbb{R}} p^{\sim w}(t) \phi_n^{j_2}(t) \ dt, \qquad w = 0, 1, 2. \tag{2.41}$$

except for the superscripts suggestive of their corresponding scale windows.

Note in (2.41), each transform involves information from a window of scales. Although the basis $\{\phi_n^{j_2}\}$ is employed, it is quite different from that in (2.35). One doesn't have a basis to accomplish a scale window analysis through a transform in the usual sense. Nevertheless, we may still refer them simply as **large-scale transform** (analysis), **meso-scale transform** (analysis), and **sub-mesoscale transform** (analysis), with the word *window* omitted, if no confusion arises. With $\widehat{p}_n^{\sim 0}$, $\widehat{p}_n^{\sim 1}$, and $\widehat{p}_n^{\sim 2}$, Eqs. (2.37), (2.38), and (2.39) can be re-stated as:

$$p^{\sim w}(t) = \sum_{n \in \mathbb{Z}} \widehat{p}_n^{\sim w} \phi_n^{j_2}(t), \qquad w = 0, 1, 2 \tag{2.42}$$

which forms a transform pair together with (2.41).

2.2.6 Properties of the multiscale window transform

In this subsection, we explore some properties of the multiscale window transform defined above. These properties are important to the later MS-EVA formulation.

Theorem 2.2.1 *Given two functions $p, q \in V_{j_2}$, and two constants $c_1, c_2 \in \mathbb{R}$, we have*

$$\widehat{(c_1 p + c_2 q)}_n^{\sim w} = c_1 \ \widehat{p}_n^{\sim w} + c_2 \ \widehat{q}_n^{\sim w}, \qquad \text{(linearity)} \tag{2.43}$$

$$(p^{\sim w})^{\sim v} = (p^{\sim v})^{\sim w} = \delta_{vw} p^{\sim w} \tag{2.44}$$

$$\widehat{(p^{\sim v})}_n^{\sim w} = \delta_{vw} \ \widehat{p}_n^{\sim w}, \tag{2.45}$$

$$\sum_{n \in \mathbb{Z}} \widehat{p}_n^{\sim v} \cdot \widehat{q}_n^{\sim w} = \delta_{vw} \int_{\mathbb{R}} p^{\sim w}(t) \cdot q^{\sim w}(t) \ dt \tag{2.46}$$

for each $n \in \mathbb{Z}$ and $v, w = 0, 1, 2$.

Properties (2.43)-(2.45) can be proved with (2.41) and (2.42) and the orthogonality between windows 0, 1, and 2. We give a little more discussion of (2.46). Notice on the left hand side, it is a sum of the products over all the location indices. This corresponds to the reversal process of the scale window sampling with a shift-invariant basis. In accordance to the concept of marginal energy introduced before, (2.46) may be referred to as a property regarding the **marginalization** of sampled quantities. When $p = q$ and $v = w$, this marginalization is similar to the Parseval relation we have seen before.

The proof of (2.46) is straightforward, with the aid of Eqs. (2.42), (2.41), and the orthonormality of scaling function $\phi_n^{j_2}(t)$. In fact,

$$\begin{aligned}
\sum_{n\in\mathbb{Z}} \widehat{p}_n^{\sim v} \cdot \widehat{q}_n^{\sim w} &= \sum_{n_1,n_2\in\mathbb{Z}} \widehat{p}_{n_1}^{\sim v} \cdot \widehat{q}_{n_2}^{\sim w}\ \delta_{n_1,n_2} \\
&= \sum_{n_1,n_2\in\mathbb{Z}} \widehat{p}_{n_1}^{\sim v} \cdot \widehat{q}_{n_2}^{\sim w} \int_{\mathbb{R}} \phi_{n_1}^{j_2}(t)\ \phi_{n_2}^{j_2}(t)\ dt \\
&= \int_{\mathbb{R}} \left[\sum_{n_1\in\mathbb{Z}} \widehat{p}_{n_1}^{\sim v}\ \phi_{n_1}^{j_2} \cdot \sum_{n_2\in\mathbb{Z}} \widehat{q}_{n_2}^{\sim w}\ \phi_{n_2}^{j_2} \right] dt \\
&= \int_{\mathbb{R}} p^{\sim v} \cdot q^{\sim w}\ dt.
\end{aligned}$$

Recall that subspaces V^L, V^M, and V^S are mutually orthogonal. That is to say,

$$\langle p^{\sim v}, q^{\sim w} \rangle = \int_{\mathbb{R}} p^{\sim v}(t) \cdot q^{\sim w}(t)\ dt = 0, \qquad \forall\ v \neq w,\ v, w = 0, 1, 2,$$

which immediately leads to (2.46).

2.3 Boundary extension and finite domain multiscale analysis

So far all the functions considered are defined over $\mathbb{R}$. In practice, however, function evaluations are usually available on some finite domain only. We must have them extended to $\mathbb{R}$ before multiscale analyses in the above sense begin. In this section, we first describe some available extension schemes and then modify the previously developed technique to handle the analysis according to these schemes.

2.3.1 Extensions

For continuous functions $x_1(t)$ and $x_2(t)$,

$$x_1 : \quad \mathcal{D}(x_1) \subseteq \mathbb{R} \longrightarrow \mathcal{R}(x_1) \subseteq \mathbb{R},$$
$$x_2 : \quad \mathcal{D}(x_2) \subseteq \mathbb{R} \longrightarrow \mathcal{R}(x_2) \subseteq \mathbb{R},$$

if $\mathcal{D}(x_1) \subseteq \mathcal{D}(x_2)$ and $x_1(t) = x_2(t)$, $\forall\ t \in \mathcal{D}(x_1)$, x_2 is called an **extension** of x_1, and x_1 the **restriction** of x_2 on $\mathcal{D}(x_1)$. Without loss of generality, let $\mathcal{D}(x_1)$ be $[0, 1)$. (A translation followed by a rescaling will make any bounded half-open interval be so if it is not.) We will see how a reasonable extension can be achieved in this context.

There exist many extensions of x_1. Constraints are needed to select one for our purpose. First, as a common practice in wavelet analysis, we impose continuity on the extended signals, which requires the extension $x_2(t)$ be continuous at the boundaries of $\mathcal{D}(x_1)$. The second, and the most important one in the light of MS-EVA, is preservation of energy. That is to say, the extension must yield the same total energy over $\mathcal{D}(x_1)$ in physical space and phase space alike, up to an error within some tolerance, if there is any. The phase space energy representation relies on the function space structure and will be clear soon.

The problem can now be cast as follows: Given an $x_1(t) \in L_2[0, 1)$, find an extension of it to $\mathbb{R}$, called $x_2(t)$, such that

(1) x_2 is continuous at $t = 0, 1$;

(2) x_2 is consistent in energy in the sense that the energy of x_2 over $\mathcal{D}(x_1)$ calculated from the physical space is equal to the energy obtained from the phase space over the same domain.

In constraint (2) we require that the total energy be preserved with the extended signal. No energy leakage outside the definition domain is allowed with the transform or synthesis. A specific formulation of this constraint requires a knowledge of the structure of the space, and we will postpone it to §2.3.5.

In wavelet analysis, people usually consider three types of extension. The first is zero-padding, which pads zeros on $\mathbb{R}$ outside $\mathcal{D}(x_1)$. The second is periodic extension, with the signal of concern wrapped around at the boundaries of $\mathcal{D}(x_1)$ (here 0 and 1):

$$x_2(t+\ell) = x_1(t), \qquad \forall\, t \in [0,1), \ell \in \mathbb{Z}. \tag{2.47}$$

In the third extension, $x_1(t)$ is extended such that $x_1(t)$ is symmetric about the boundary points. It is called symmetric extension, or extension by reflection accordingly:

$$x_2(t+\ell) = \begin{cases} x_1(t) & \ell \text{ even, } \ell \in \mathbb{Z},\ t \in \mathcal{D}(x_1), \\ x_1(1-t) & \ell \text{ odd, } \ell \in \mathbb{Z},\ t \in \mathcal{D}(x_1). \end{cases} \tag{2.48}$$

Among the three extensions, zero-padding, if adopted, will validate all the previous efforts in constructing the multiscale analysis. However, this simple scheme seldom has continuity at the boundaries of $\mathcal{D}(x_1)$, and is therefore excluded from consideration. Periodical extension may not be able to meet the continuity requirement, either. But for periodic functions, it will produce an exact extension. The third type, symmetric extension, gives an $x_2(t)$ which is continuous at the boundaries, though the first derivative is, in general, discontinuous. We discuss hereafter only periodic and symmetric extensions.

To approach the two extensions in a unified way, notice that with the aforementioned periodic extension functions are actually periodized with period one. This definition can be easily generalized to the case with any periodicities. For $\varrho \in \mathbb{R}$, a ϱ-periodic extension (or an extension by periodization with period ϱ) of a $[0,\varrho)$-defined function $x_1(t)$ is another function x_2 on $\mathbb{R}$ such that

$$x_2(t+\varrho\ell) = x_1(t), \qquad \forall\, t \in [0,\varrho),\ \ell \in \mathbb{Z}. \tag{2.49}$$

With this definition, the previously introduced periodization is actually a 1-periodic extension. For the symmetric extension, it can be achieved through two steps: an extension from [0, 1) to [0, 2) by reflection, followed by an extension by periodization with period two. Since both schemes are essentially the same, we need only study the case of periodization with some generic period ϱ. For simplicity, suppose ϱ is a power of 2 (say, $\varrho = 2^\lambda$).[13] If a periodization scheme is referred to but with no period specified, ϱ is always assumed to be one.

[13] This assumption is actually not necessary. We have it here in order for Hernández and Weiss's results (1996) to apply in our proofs of the following theorems. Anyway, we consider only two cases: $\varrho = 1$ and 2, either with an integer power ($\lambda = 0, 1$).

2.3.2 Multi-resolution analysis of $L_2(\mathbb{T}_\varrho)$

We have made extensions for the finite-lengthed signals. Now we need a space for these newly formed functions. Specifically, a Hilbert space X is needed which contains all the ϱ-periodic functions on $\mathbb{R}$ such that

$$\int_0^\varrho |x(t)|^2\ dt < \infty, \qquad \forall\ x \in X.$$

Clearly this space is just the $L_2(\mathbb{T}_\varrho)$ we have introduced before (p. 30). Our problem is hence to be studied in $L_2(\mathbb{T}_\varrho)$ instead of $L_2(\mathbb{R})$ from now on.

The change in function space requires a change in the entire system we have developed before. Specifically, we need to build a sequence of closed subspaces of $L_2(\mathbb{T}_\varrho)$, $\{V_{\varrho,j}\}_{j\geq 0}$, in analogy to the multi-resolution analysis of $L_2(\mathbb{R})$, such that

(1) $V_{\varrho,0} \subset V_{\varrho,1} \subset ....$ (**Nestedness**)

(2) $\cup_{j=0}^{\infty} V_{\varrho,j}$ is dense in $L_2(\mathbb{T}_\varrho)$. (**Totality**)

(3) There exists an affine orthonormal set which is total in $V_{\varrho,j}$.

(4) $x(t) \in V_{\varrho,j}$ iff $x(2t) \in V_{\varrho,j+1}$. (**Refinability**)

and redefine the multiscale window and multiscale window transform in accordance to the new space structures. For convenience, we will assign this sequence a name, **multi-resolution analysis of** $L_2(\mathbb{T}_\varrho)$, which, with the space $L_2(\mathbb{T}_\varrho)$ clearly specified, is easy to be distinguished from what we have defined before on page 34 with $L_2(\mathbb{R})$. The task of this subsection is to construct such an analysis, and the procedure of construction generally follows that of Hernández and Weiss (1996).

Recall that $\{\phi_n^j(t)\}_{n\in\mathbb{Z}}$ is an orthonormal basis of $V_j \subset L_2(\mathbb{R})$. Assuming that ϕ has a polynomial localization of order $\gamma > 1$, we can form a new sequence

$$\{\phi_n^{\varrho,j}(t)\}_{n\in\mathcal{N}_\varrho^j} = \{\sum_{\ell\in\mathbb{Z}} \phi_n^j(t+\varrho\ell), \quad n = 0,1,2,...,2^j\varrho - 1\}, \qquad \forall\ t \in [0,\varrho), \tag{2.50}$$

where $\mathcal{N}_\varrho^j = \{0,1,2,...,2^j\varrho - 1\}$. (With $\gamma > 1$ the convergence of the infinite sum is not an issue. Refer to the Lemma A2.5.1 on p. 109) Clearly $\{\phi_n^{\varrho,j}(t)\}_{n\in\mathcal{N}_\varrho^j} \subset L_2(\mathbb{T}_\varrho)$, and moreover, it has the following property:

Theorem 2.3.1 *$\{\phi_n^{\varrho,j}(t)\}_{n\in\mathcal{N}_\varrho^j}$ furnishes an orthonormal system in $L_2(\mathbb{T}_\varrho)$.*

Proof: See Appendix A2.1

Theorem 2.3.1 allows us to build a subspace of $L_2(\mathbb{T}_\varrho)$:

$$V_{\varrho,j} = \text{span}\{\phi_n^{\varrho,j}\}_{n\in\mathcal{N}_\varrho^j} \subset L_2(\mathbb{T}_\varrho) \tag{2.51}$$

which has an orthonormal basis $\{\phi_n^{\varrho,j}(t)\}_{\mathcal{N}_\varrho^j}$ and is of finite dimensionality $2^j\varrho$ (hence closed). This process can be repeated again and again for any $j \in \mathbb{Z}$, $j \geq 0$ and hence a sequence of finite-dimensional subspaces of $L_2(\mathbb{T}_\varrho)$ is obtained. We wish this sequence to be the multi-resolution analysis we want, and the following two theorems make the wish come true.

Theorem 2.3.2 *For the subspace $V_{\varrho,j}$ of $L_2(\mathbb{T}_\varrho)$ formed by (2.51), the following inclusion holds:*

$$V_{\varrho,j} \subset V_{\varrho,j+1}, \qquad \forall\ j = 0, 1, 2, ...$$

Proof: See Appendix A2.2

Theorem 2.3.3 *$\bigcup_{j=0}^{\infty} V_{\varrho,j}$ is dense in $L_2(\mathbb{T}_\varrho)$.*

Proof: See Appendix A2.3

We have thence found a multi-resolution analysis of $L_2(\mathbb{T}_\varrho)$, $\{V_{\varrho,0}, V_{\varrho,1}, V_{\varrho,2}, ...\}$, which are spanned respectively by the periodized bases $\{\phi_n^{\varrho,0}\}$, $\{\phi_n^{\varrho,1}\}$, $\{\phi_n^{\varrho,2}\}$, and so forth. Likewise, another periodized basis

$$\{\psi_n^{\varrho,j}(t)\}_{n\in\mathcal{N}_\varrho^j} = \{\sum_{\ell\in\mathbb{Z}} \psi_n^j(t+\varrho\ell), \quad n = 0, 1, 2..., 2^j\varrho - 1\}, \qquad \forall\ t \in [0, \varrho) \tag{2.52}$$

and correspondingly a subspace of $L_2(\mathbb{T}_\varrho)$

$$W_{\varrho,j} = \text{span}\{\psi_n^{\varrho,j}\}_{n\in\mathcal{N}_\varrho^j} \tag{2.53}$$

may also be formed. Clearly $\dim W_{\varrho,j} = 2^j\varrho$. Following a procedure exactly the same as that for $\phi_n^{\varrho,j}$ and $V_{\varrho,j}$, it is easy to show that

- $\{\psi_n^{\varrho,j}\}_{n\in\mathcal{N}_\varrho^j}$ is an orthonormal basis of $W_{\varrho,j}$,

- $W_{\varrho,j} \subset V_{\varrho,j+1}$,

and, as in the $L_2(\mathbb{R})$ case,

Theorem 2.3.4 *$W_{\varrho,j} \perp V_{\varrho,j}$ and $V_{\varrho,j} \oplus W_{\varrho,j} = V_{\varrho,j+1}$.*

Proof: The perpendicularity between $W_{\varrho,j}$ and $V_{\varrho,j}$ can be proved by following the same steps as the proof of Theorem 2.3.1, together with the orthogonality between W_j and V_j. Besides, we know

$$\mathrm{dim}W_{\varrho,j} + \mathrm{dim}V_{\varrho,j} = 2^j\,\varrho + 2^j\,\varrho = \mathrm{dim}V_{\varrho,j+1},$$

so $V_{\varrho,j} \oplus W_{\varrho,j} = V_{\varrho,j+1}$.

Parallel to $V_j, W_j \subset L_2(\mathbb{R})$, the two subspaces of $L_2(\mathbb{T}_\varrho)$, $V_{\varrho,j}$ and $W_{\varrho,j}$, also classify the function collections in the universe according to scales. $V_{\varrho,j}$ encompasses features with all the scales from level j down to 0 (2^{-j} to 2^0), while $W_{\varrho,j}$ contains only that with scale 2^{-j}. For convenience, $V_{\varrho,j}$ will be referred to as the multi-resolution approximation of $L_2(\mathbb{T}_\varrho)$ up to level j.

2.3.3 Multiscale window and multiscale window transform in $L_2(\mathbb{T}_\varrho)$

With the multi-resolution analysis of $L_2(\mathbb{T}_\varrho)$, we now proceed to modify the previously defined scale windows to the new space. Suppose the large-scale, meso-scale, and sub-mesoscale windows are delimited by scales 2^{-j_0}, 2^{-j_1}, 2^{-j_2} (argument pre-scaled by the signal duration), with integer levels $j_0 < j_1 < j_2$, then the three windows can be constructed as follows:

Scale window	Notation	Definition
Large-scale window	V_ϱ^L	$V_{\varrho,j_0} \quad (j_0 \geq 0)$
Meso-scale window	V_ϱ^M	$V_{\varrho,j_1} \ominus V_{\varrho,j_0} \quad (j_1 > j_0)$
Sub-mesoscale window	V_ϱ^S	$V_{\varrho,j_2} \ominus V_{\varrho,j_1} \quad (j_2 > j_1)$

By Theorem 2.3.4, V_ϱ^L, V_ϱ^M, and V_ϱ^S are mutually orthogonal. Note here we tacitly assume that the smallest scale is cut off at some finite level j_2. That is to say, all functions are assumed to be in V_{ϱ,j_2}, a proper subspace of $L_2(\mathbb{T}_\varrho)$. This remains to be justified but the justification will be put off to the next section. We merely take it for granted at this moment.

The modified multiscale transform now can be introduced. Before proceeding, we disclaim that all the related symbols and terms used before will be kept for their respective usages. The ambiguity of notation and terminology will not cause confusion, as from now on only extended signals will be considered. As before, we begin with defining a scaling-like transform (we still call it a scaling transform)

$$\widehat{p}_n^j = \int_0^\varrho p(t)\ \phi_n^{\varrho,j}(t)\ dt \tag{2.54}$$

for $p \in V_{\varrho,j}$. From this transform three reconstructions of p, called large-scale, meso-scale, and sub-mesoscale synthesis, are defined, respectively, as

$$\begin{aligned} p^{\sim 0}(t) &= \sum_{n\in\mathcal{N}_\varrho^{j_2}} \widehat{p}_n^{j_0}\ \phi_n^{\varrho,j_0}(t), & (2.55)\\ p^{\sim 1}(t) &= \sum_{n\in\mathcal{N}_\varrho^{j_2}} \widehat{p}_n^{j_1}\ \phi_n^{\varrho,j_1}(t) - p^{\sim 0}(t), & (2.56)\\ p^{\sim 2}(t) &= p(t) - \sum_{n\in\mathcal{N}_\varrho^{j_2}} \widehat{p}_n^{j_1}\ \phi_n^{\varrho,j_1}(t), & (2.57)\end{aligned}$$

for all $t \in [0,\varrho)$. And, accordingly, the multiscale window transform can be introduced:

$$\widehat{p}_n^{\sim w} = \int_0^\varrho p^{\sim w}(t)\ \phi_n^{\varrho,j_2}(t)\ dt, \tag{2.58}$$

where $w = 0, 1, 2$ signifies the three scale windows. With (2.58), the three equations, (2.55)-(2.57), may be condensed to a single one

$$p^{\sim w}(t) = \sum_{n\in\mathcal{N}_\varrho^{j_2}} \widehat{p}_n^{\sim w} \phi_n^{\varrho,j_2}(t), \qquad w = 0, 1, 2. \tag{2.59}$$

Eqs. (2.58) and (2.59) form the transform pair of multiscale window analysis in $L_2(\mathbb{T}_\varrho)$. Note in both equations, n runs over $\mathcal{N}_\varrho^{j_2} = \{0, 1, 2, ..., 2^{j_2}\varrho - 1\}$. In showing the final result, however, we need only the first 2^{j_2} values (i.e., those in $\mathcal{N}_1^{j_2}$), which corresponds to the domain $[0, 1)$ in the physical space.

2.3.4 Localization and self-similarity with the periodized basis

Recall that localness and orthonormality are two crucial issues in choosing a basis for the MS-EVA study. For $\{\phi_n^{\varrho,j}\}_{n\in\mathcal{N}_\varrho^j}$, the orthonormality is guaranteed by Theorem 2.3.1. The localness, however, could be problematic, as a basis after periodization is by no means as local as it used to be. We will now examine this problem.

For a periodized basis $\{\phi_n^{\varrho,j}\}$ of $V_{\varrho,j}$, consider the element $\phi_n^{\varrho,j}(t)$, $n \in \mathcal{N}_\varrho^j$, $t \in [0,1)$. Involved in it is an infinite series sum:

$$\ldots + \phi(2^j t - n - 2^j \varrho) + \phi(2^j t - n) + \phi(2^j t - n + 2^j \varrho) + \ldots .$$

We know $\phi(t)$, $t \in \mathbb{R}$ is localized and has a localization of an order $\gamma > 1$ by the assumption invoked in the beginning of this section. $\phi(2^j t - n)$ is hence localized around $2^{-j}n$ in the physical space. The sum of the terms other than $\phi(2^j t - n)$, denoted by OTHER hereafter, corrupts the localization of $\phi_n^{\varrho,j}(t)$ on $t \in [0,1)$ around $2^{-j}n$. The degree of this corruption depends on the scale level j as well as the localization order γ of ϕ. For a fixed γ, the higher the scale level (the smaller the scale), the weaker the influence of OTHER; for a fixed j, a weak influence of OTHER is possible with a highly localized scaling function ϕ. This is to say, although $\phi_n^{\varrho,j}(t)$ is theoretically not localized on $t \in [0,1)$, in practice it may look like so if the scale level is higher than a certain threshhold given a localization order. This may be best illustrated with the aid of an example. As shown in Fig. 2.2, at a high scale level (the bottom panel), $\phi_n^{\varrho,j}(t)$ ($\varrho = 1$ here) is essentially the same as that of $\phi_n^j(t)$ (not shown), for almost all $n \in \mathcal{N}_1^j$ except those near 0 and $2^j - 1$ where the boundaries are hit. With the decreasing of j (increasing in scale), the effect of periodization encroaches more and more upon the interior until all the domain feels the boundary modification on a scale comparable to one (top panel). Therefore, if we consider only those j's greater than some critical value, and focus only on the interior region, i.e., a region not within the ε-lengths to the two ends of the signal, the localization is practically not a problem. In this example, j should be greater than or equal to 3 for a satisfactorily localized $\phi_n^{\varrho,j}$. When a symmetric extension is used, i.e., $\varrho = 2$, a smaller j could have the same effect for the same ϕ. In that case, by our experiment, $j = 2$ can generate a $\phi_n^{2,j}$ as localized as what $j = 4$ can do for $\phi_n^{1,j}$ (figure not shown).

Self-similarity or scale invariance is another feature we care about with the periodized basis. Again, the periodization distorts this elegance only on a scale comparable to 1. (See Fig. 2.2). If the signal is long enough, or if we focus on features with scales smaller enough (at least eight times smaller than the duration scale for the above example), the self-similarity won't make much an issue in practice, either.

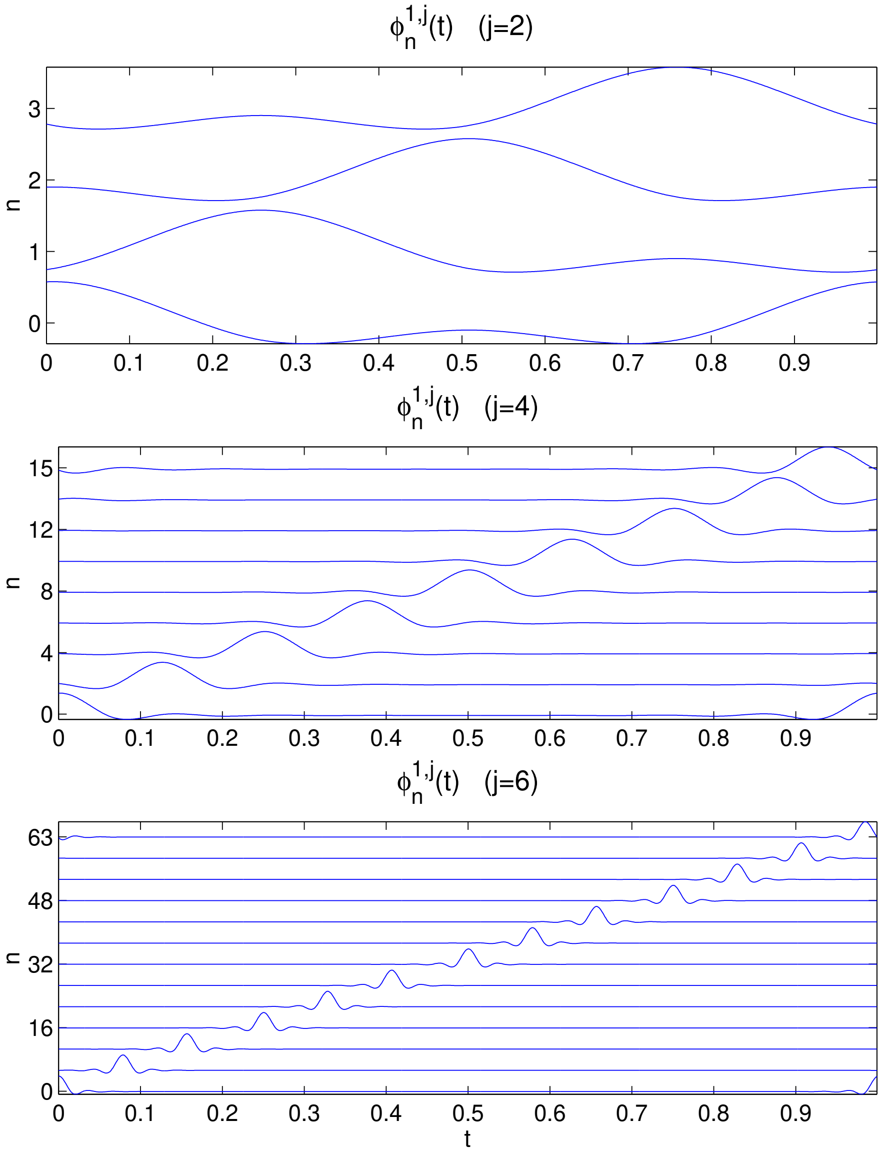

Figure 2.2: An example showing the localization of periodized bases with scale levels $j = 2$ (top), $j = 4$ (middle), and $j = 6$ (bottom), respectively. The scaling function ϕ is shown in Fig. 2.6 and its construction will be elaborated later.

2.3.5 The energy consistency requirement revisited

With the function space established, we are now able to examine the issue of energy leakage raised in the beginning of this section. We want to see if the above two extensions satisfy this no-leak requirement, or if not, what additional condition should be imposed on the building blocks we have developed thus far.

Let x_2 be an extended function which lies in $V_{\varrho,j}$ for some $j > 0$. Let

$$E^j_{\text{phys}} \equiv \int_0^1 |x_2(t)|^2 \; dt, \tag{2.60}$$

$$E^j_{\text{phase}} \equiv \sum_{n=1}^{N-1} (\alpha_n)^2 + \frac{1}{2}\left[(\alpha_0)^2 + (\alpha_N)^2\right], \qquad N = 2^j, \tag{2.61}$$

where $\alpha_n = \langle x_2, \phi_n^{\varrho,j} \rangle$, then the energy consistency requirement reads

$$E^j_{\text{phys}} = E^j_{\text{phase}}. \tag{2.62}$$

Here the energy in the phase space over $[0, 1)$, E^j_{phase}, is defined to be (2.61), rather than $\sum_{n=0}^{N-1} (\alpha_n)^2$, to well balance the two boundaries, $n = 0$ and $n = N$. Our task in this subsection is to show that Eq. (2.62) holds for $\varrho = 1$ and 2. For later convenience, we explore a more general property than just energy.

Theorem 2.3.5 *For any functions $p, q \in V_{\varrho,j}$ with period $\varrho = 1$, we have the following equality:*

$$\int_0^1 p(t)q(t) \; dt = \sum_{n=1}^{N-1} \alpha_n\beta_n + \frac{1}{2}(\alpha_0\beta_0 + \alpha_N\beta_N)$$

where $N = 2^j$, and $\alpha_n = \langle p, \phi_n^{1,j} \rangle$, $\beta_n = \langle q, \phi_n^{1,j} \rangle$, for $n = 0, 1, ..., N$

Proof:

As a periodized basis with $\varrho = 1$, $\phi_N^{1,j} = \phi_0^{1,j}$, so

$$\alpha_N = \alpha_0, \qquad \beta_N = \beta_0, \qquad \frac{1}{2}(\alpha_0\beta_0 + \alpha_N\beta_N) = \alpha_0\beta_0$$

and what we need to prove then becomes

$$\int_0^1 p(t) \; q(t) \; dt = \sum_{n=0}^{N-1} \alpha_n\beta_n.$$

But this is just the **generalized Parseval relation** (see Anderson, 2001), the theorem is therefore established.

Theorem 2.3.6 *For any two symmetrically extended function p and q, if they belong to $V_{2,j}$, then*

$$\int_0^1 p(t)q(t)\ dt = \sum_{n=1}^{N-1} \alpha_n \beta_n + \frac{1}{2}(\alpha_0\beta_0 + \alpha_N\beta_N)$$

provided that $\phi(t)$ is symmetric about zero. Here N and α_n, β_n have the same meaning as they do in Theorem 2.3.5, except that $\{\phi_n^{2,j}\}_{n\in\mathcal{N}_2^j}$ is now a periodized basis with period two.

Proof: See Appendix A2.4

Let $p = q$, Theorem 2.3.5 implies that the extension by periodization preserves energy unconditionally. For the symmetric extension case, there will also be no energy leak if the scaling function ϕ used for the basis construction is symmetric about the origin.[14] So, if a symmetric ϕ is constructed, either of the two extension schemes meets the energy preservation requirement.

2.3.6 Properties of the finite domain multiscale window transform

We have shown in §2.2.6 some properties of the multiscale window transform in $V_{j_2} \subset L_2(\mathbb{R})$. We shall see that in $V_{\varrho,j_2} \subset L_2(\mathbb{T}_\varrho)$, similar properties also hold. The following identities are in a form same as those of §2.2.6:

Theorem 2.3.7 *Given two functions $p, q \in V_{\varrho,j_2}$, with either periodic or symmetric extension adopted, and two constants $c_1, c_2 \in \mathbb{R}$, we have*

$$\widehat{(c_1 p + c_2 q)}_n^{\sim w} = c_1\ \widehat{p}_n^{\sim w} + c_2\ \widehat{q}_n^{\sim w}, \tag{2.63}$$

$$(p^{\sim w})^{\sim v} = (p^{\sim v})^{\sim w} = \delta_{vw} p^{\sim w} \tag{2.64}$$

$$\widehat{(p^{\sim v})}_n^{\sim w} = \delta_{vw}\ \widehat{p}_n^{\sim w}, \tag{2.65}$$

for each $n \in \mathcal{N}_\varrho^{j_2}$ and $v, w = 0, 1, 2$.

The proofs of these identities are straightforward. They follow directly from the definitions (2.58) and (2.59), and the mutual orthogonality between the three distinct windows.

[14]Wavelet analysis doesn't have this property as the wavelet basis function $\psi(t)$ is not symmetric about $t = 0$. Energetic studies with wavelet transform thus may encounter theoretical problems if a non-periodic extension is used.

Apart from (2.63)-(2.65), finite domain multiscale window analysis also has idiosyncratic properties of its own. The following two theorems are important in the MS-EVA development.

Theorem 2.3.8 *Suppose $p(t)$, $t \in [0,1)$, is a function in V_{ϱ,j_2} with period $\varrho = 1$. Suppose further that the scaling function for V_{ϱ,j_2}, ϕ, has a polynomial localization of order $\gamma > 1$. We have, for $j_0 = 0$,*

$$p^{\sim 0}(t) = \widehat{p}_0^{j_0} = \overline{p(t)}, \qquad \text{for } t \in [0,1), \tag{2.66}$$

where the overline signifies an averaging over the duration, i.e. $\overline{p(t)} = \int_0^1 p(t)\ dt$.

Proof

To prove this theorem, notice if the scaling function ϕ is of polynomial localization of order $\gamma > 1$, then

$$\phi_n^{\varrho,0}(t) = \sum_{\ell \in \mathbb{Z}} \phi(t + \ell - n) = 1, \tag{2.67}$$

for all $t \in [0,1)$, a direct consequence of the fact that $\sum_{l \in \mathbb{Z}} \phi(t+l) = 1$ when $\gamma > 1$. (e.g., Hernández and Weiss, 1996, p. 222; but the condition they impose is weaker.) By property (2.65),

$$\widehat{(p^{\sim 0})}_n^0 = \widehat{(p^{\sim 0})}_n^{\sim 0} = \widehat{p}_n^{\sim 0} = \widehat{p}_n^{j_0} = \widehat{p}_n^0.$$

So

$$\begin{aligned}
p^{\sim 0}(t) &= \sum_{n=0}^{2^0-1} \widehat{(p^{\sim 0})}_n^0 \cdot \phi_n^{\varrho,0}(t) \\
&= \widehat{p}_0^0 \\
&= \int_0^1 p(t) \phi_0^{\varrho,0}(t)\ dt \\
&= \int_0^1 p(t)\ dt \\
&= \overline{p(t)}.
\end{aligned}$$

Q.E.D.

Theorem 2.3.8 implies that, for a periodically extended signal, if the large scale window has an upper level bound $\underline{j_0 = 0}$, then its synthesis on this window is simply the average over the duration. This relation between the large-scale synthesis and duration average, as we will see later, is very useful in connecting the MS-EVA to the classical energetic formulations.

Another useful theorem regards something we shall refer to as *finite domain marginalization*, in parallel to the marginalization property in $V_{j_2} \subset L_2(\mathbb{R})$.

Theorem 2.3.9 (marginalization) *Suppose $p(t), q(t) \in L_2(\mathbb{T}_\varrho)$ are either periodically or symmetrically extended ($\varrho = 1$ or 2 depending on the extension scheme). Suppose further that they lie in the subspace V_{ϱ,j_2} of $L_2(\mathbb{T}_\varrho)$, which is formed with a scaling function $\phi(t)$ symmetric about the origin, then with the notation $N \equiv 2^{j_2}$, the following equality holds:*

$$\sum_{n=1}^{N-1} \widehat{p}_n^{\sim w} \widehat{q}_n^{\sim w} + \frac{1}{2}\left[\widehat{p}_0^{\sim w} \widehat{q}_0^{\sim w} + \widehat{p}_N^{\sim w} \widehat{q}_N^{\sim w}\right] = \int_0^1 p^{\sim w}(t) q^{\sim w}(t)\ dt, \tag{2.68}$$

for all scale windows $w = 0, 1, 2$.

Proof

For any $p, q \in V_{\varrho,j_2}$, we can find the multiscale synthesis $p^{\sim w}, q^{\sim w} \in V_{j_2}$ for windows $w = 0, 1, 2$. Let $N = 2^{j_2}$, substitute respectively $p^{\sim w}, q^{\sim w}$ for the p, q in the equalities of Theorems 2.3.5 and 2.3.6, and use definition (2.58) which may be alternatively stated as

$$\widehat{(p^{\sim w})}_n^{j_2} = \widehat{p}_n^{\sim w},$$

the result then follows immediately.

Theorem 2.3.9 is actually a re-statement of the marginalization property (2.46) on the finite domain $[0, 1)$. For this reason, we call the left hand side of (2.68) the **marginalization of** $\widehat{p}_n^{\sim w} \widehat{q}_n^{\sim w}$ **over** $[0, 1)$ (or merely marginalization of $\widehat{p}_n^{\sim w} \widehat{q}_n^{\sim w}$ if the domain is clear in the context), and denote it as

$$\sum_{n=1}^{N-1} \widehat{p}_n^{\sim w} \widehat{q}_n^{\sim w} + \frac{1}{2}\left[\widehat{p}_0^{\sim w} \widehat{q}_0^{\sim w} + \widehat{p}_N^{\sim w} \widehat{q}_N^{\sim w}\right] \equiv \mathcal{M}_n\left(\widehat{p}_n^{\sim w} \widehat{q}_n^{\sim w}\right). \tag{2.69}$$

Using the overline notation for a duration average such that $\bar{p} = \int_0^1 p\ dt$ for any integrable function $p = p(t)$ over [0,1), the marginalization property, Eq. (2.68), may be succinctly written as

$$\boxed{\mathcal{M}_n\left(\widehat{p}_n^{\sim w} \widehat{q}_n^{\sim w}\right) = \overline{p^{\sim w} q^{\sim w}},} \tag{2.70}$$

for the three windows $w = 0, 1, 2$.

2.3.7 Multiscale energy representation

The multiscale window transform equipped with the marginalization property (2.70) allows a simple representation of the energy for any scale window in concern. For $p(t) \in V_{\varrho,j_2}$, $N = 2^{j_2}$, let $E_n^{W*} \equiv (\widehat{p}_n^{\sim w})^2$ ($w = 0, 1, 2$, and correspondingly $W = L, M, S$). By (2.70),

$$\mathcal{M}_n E_n^{W*} \;=\; \int_0^1 [p^{\sim w}(t)]^2 \;dt, \tag{2.71}$$

which is the energy of window W (up to some constant factor) integrated over $[0, 1)$. E_n^{W*} thus can be viewed as the energy of window W summed over a small interval of length $\Delta t = 2^{-j_2}$ around location $t = 2^{-j_2}n$. An energy variable for window W at $2^{-j_2}n$ consistent with the fields at that location is therefore a locally averaged quantity

$$\boxed{E_n^W = \frac{1}{\Delta t} E_n^{W*} = 2^{j_2} \cdot (\widehat{p}_n^{\sim w})^2,} \tag{2.72}$$

for $w = 0, 1, 2$, and $W = L, M, S$ correspondingly.

2.4 Treatment of discrete signals

So far all functions have been assumed to be in some finite level multi-resolution approximation V_{ϱ,j_2} of $L_2(\mathbb{T}_\varrho)$. In this section we shall justify this assumption.

In practice, time signals generated with a numerical model are not continuous. They are discretized as well as of finite length. For convenience, suppose the signals in question are output at $N = 2^{j_2}$ equi-distant locations, for some positive integer j_2. Taking for example some $x(t)$, t being scaled by the signal duration, this is equivalent to saying that we have a realization $x_n = x(t_n)$, with $t_n = n/N$, and $n = 0, 1, 2, ..., N-1$. Since we have no idea about features with scale less than $1/N$, what we need to justify is that the interpolation of these N x_n's with some basis lies in some $V_{\varrho,j}$ (and hence in any multi-resolution of $L_2(\mathbb{T}_\varrho)$ with level higher than j). Choose the interpolation basis to be $\{\phi_n^{\varrho,j}\}_{n\in\mathcal{N}_\varrho^j}$. Then we need to show that, *under a specific extension*,

$$x_n = \sum_{m\in\mathcal{N}_\varrho^j} \alpha_m \phi_m^{\varrho,j}(t_n), \qquad n = 0, 1, ..., N-1 \tag{2.73}$$

has a unique solution for coefficients α_m. To fulfill the mission, we distinguish the following two cases:

- Case with a periodic extension;
- Case with a symmetric extension.

2.4.1 Periodic extension

A signal $x(t)$ obtained by periodic extension means that it lies in some subspace $V_{\varrho,j} \subset L_2(\mathbb{T}_\varrho)$ with period $\varrho = 1$. This subspace has a dimensionality 2^j. Recall we have $N = 2^{j_2}$ observations for $x(t)$. They will furnish N equations for N unknown α's if j is chosen to be j_2:

$$\sum_{m=0}^{N-1} \alpha_m \phi_m^{\varrho,j_2}(t_n) = x(t_n), \qquad t_n = \frac{n}{N}, \quad n = 0, 1, ..., N-1 \tag{2.74}$$

Written in a matrix form, this is

$$\underline{\underline{\mathbf{H}}}\ \underline{\alpha} = \underline{\mathbf{x}}, \tag{2.75}$$

with the entries of $\underline{\underline{\mathbf{H}}}$ formed by summing up the scaling basis function:

$$H_{nm} = \phi_m^{\varrho,j_2}(t_n) = \sum_{l \in \mathbb{Z}} \phi_{m-lN}^{j_2}(t_n). \tag{2.76}$$

If $\underline{\underline{\mathbf{H}}}$ is invertible , then $\alpha = \underline{\underline{\mathbf{H}}}^{-1}\underline{\mathbf{x}}$, and this x thus interpolated lies surely in V_{ϱ,j_2}. Therefore, the question whether $x \in V_{\varrho,j_2}$ is justifiable is transformed into a problem regarding the invertibility of $\underline{\underline{\mathbf{H}}}$. The following theorem gives this question an answer.

Theorem 2.4.1 *The matrix $\underline{\underline{\mathbf{H}}}$ formed above is of full rank, if the scaling function $\phi(t)$ is maximized at $t = 0$ and*

$$|\phi(n)| \le \frac{\phi(0)}{(1+n^2)^{\gamma/2}}, \qquad \forall\ n \in \mathbb{Z} \tag{2.77}$$

for some $\gamma > 1$, which satisfies the inequality

$$\frac{2}{2^{\gamma/2}} + \frac{2}{2^{\gamma-1}-1} + \left(1 + \frac{4}{1-2^{1-\gamma}}\right)\frac{1}{N^{\gamma-1}} < 1 \tag{2.78}$$

where $N = 2^{j_2}$.

Proof: See Appendix A2.5

Note (2.77) is much weaker than the corresponding localization requirement for ϕ. It involves only integer points on the definition domain. For many scaling functions, as the one we will build soon, $\phi(n)$ is almost negligible in comparison to $\phi(0)$. The parameter γ is thence usually very large, albeit the localization order of ϕ is actually much smaller. In this case, inequality (2.78) will hold provided that N is larger than a certain value. For example, when $\gamma = 3.6$, a signal with a length $N > 8$ will always have (2.78) satisfied.

2.4.2 Symmetric extension

For the symmetric extension, one may use the same trick as above to justify $x(t) \in V_{\varrho,j_2} \subset L_2(\mathbb{T}_2)$. But it is easier to accomplish the task by availing ourselves of the proved Theorem 2.4.1.

Extend the series $x_n = x(\frac{n}{N})$, $n = 0, 1, ..., N-1$ symmetrically by a half to form a new series x'_n such that

$$x'\left(\frac{n}{N}\right) = x'_n = \begin{cases} x_n, & n = 0, 1, ..., N-1 \\ x_N, & n = N \\ x_{2N-n}. & n = N+1, ..., 2N-1 \end{cases} \tag{2.79}$$

Here $x'_N = x_N$ is not provided by data. It could be filled with any number that is reasonable, e.g., $x'_N = \frac{1}{2}(x'_{N-1} + x'_{N+1}) = x_{N-1}$. This extra point emerges in the symmetric extension for discrete signals. For a continuous signal, it is not necessary.

Periodically extrapolating the newly formed signal x'_n, $n = 0, 1, ..., 2N-1$, fulfills a symmetric extension of x_n, $n = 0, 1, ..., N-1$. The justification problem is now to show that the interpolated function $x'(t), t \in [0, 2)$ with periodic extension lies in V_{ϱ,j_2}. To do this, re-scale the argument of x' to construct a new function

$$y(t) = x'(2t), \qquad t \in [0, 1), \tag{2.80}$$

and its corresponding series

$$y_n = y\left(\frac{n}{2N}\right) = x'_n, \qquad n = 0, 1, ..., 2N-1. \tag{2.81}$$

It is at this point that Theorem 2.4.1 applies. Note the N in (2.78) now should be replaced with $2N = 2^{j_2+1}$. This replacement further weakens the already weak condition (2.78) for a family of scaling function, under which $y(t)$, $t \in [0, 1)$ is a vector in V_{ϱ,j_2+1} by the above theorem. This is to say, when (2.78) is satisfied, there exists a unique series α_n such that

$$
\begin{aligned}
x'(t) = y(t/2) &= \sum_{m=0}^{2N-1} \alpha_m \phi_m^{\varrho,j_2+1}(\frac{t}{2}) \qquad t \in [0,2) \\
&= \sum_{m=0}^{2N-1} \sqrt{2}\alpha_m \phi_m^{\varrho,j_2}(t). \qquad (2.82)
\end{aligned}
$$

In this case, $x'(t), t \in [0, 2)$ is a function in V_{ϱ,j_2} with period 2, and so is $x(t), t \in [0, 1)$. This completes the justification under a symmetric extension.

2.5 Construction of the scaling basis

At this stage, the ground work has been finished and our problem left is to build a scaling basis $\{\phi(t-k)\}_{k\in\mathbb{Z}}$. This scaling basis is not just an arbitrary one. It must meet the following requirements according to the above sections:

- $\{\phi(t-k)\}_{k\in\mathbb{Z}}$ orthonormal;
- $\phi(t)$ symmetric about the origin and maximized at zero;
- $\phi(t)$ polynomially localized up to order $\gamma > 1$.

We begin this section with an introduction of the dilation equation, which must be satisfied by any scaling function. Dilation equation brings together a scaling function and filter bank, which is particularly important in scaling analysis and basis construction. The second subsection applies this equation to show that a scaling basis, if orthonormalized, is also scaling. With this result, we start the construction with a function, the cubic spline function, which is known to satisfy the dilation equation, and orthonormalized it to achieve our goal (§2.5.3).

Notation remarks: (1) Generally Strang and Nguyen (1997)'s conventions are adopted throughout. Capital Latin or Greek letters with a hat represent functions in the frequency domain (Fourier transform), while lowercase Latin or Greek letters without a hat stand for time domain representation. (2) While throughout this thesis, all scaling bases are assumed to be orthonormal, in this section orthonormal and non-orthonormal bases do coexist. In case any confusion arises, notation ϕ without any superscript or subscript is always reserved for the orthonormal class, while bases in a more generic sense are distinguished with some proper subscripts (e.g., ϕ_3 for a cubic spline).

2.5.1 Dilation equation and filter bank

The construction of scaling function ϕ begins with an equation called the *dilation equation*. It is a direct result of the multi-resolution analysis (MRA).

We know, by the nestedness of MRA, that $V_0 \subset V_1$. The scaling function $\phi \in V_0$ consequently can be represented with the basis of V_1, $\{2^{1/2}\phi(2t-k)\}_{k\in\mathbb{Z}}$:

$$\phi(t) = \sqrt{2}\sum_{k\in\mathbb{Z}} h_0(k)\phi(2t-k), \qquad \forall\ t \in \mathbb{R}. \tag{2.83}$$

This is the so-called **dilation equation**, with $h_0(k)$ being the expansion coefficients. Likewise, $W_0 \subset V_1$ implies

$$\psi(t) = \sqrt{2}\sum_{k\in\mathbb{Z}} h_1(k)\phi(2t-k), \qquad \forall\ t \in \mathbb{R}, \tag{2.84}$$

an equality referred to as a **wavelet equation**. Note in Eqs. (2.83) and (2.84), the premultiplied factor is $\sqrt{2}$, instead of 2 as in Strang and Nguyen (1997). This difference arises from the different function spaces which have been chosen ($L_2(\mathbb{R})$ here).

More generally, for any j, the nestedness of MRA implies that $V_j \subset V_{j+1}$. So there exists a sequence $\{\alpha_k^j\}_{k\in\mathbb{Z}} \in \ell_2$, such that

$$\phi_0^j(t) = \sum_{k\in\mathbb{Z}} \alpha_k^j \phi_k^{j+1}(t), \tag{2.85}$$

or

$$2^{j/2}\phi(2^j t) = \sum_{k\in\mathbb{Z}} \alpha_k^j \cdot 2^{(j+1)/2}\phi(2^{j+1}t-k),$$

which becomes, with t replaced by $2^{-j}t'$,

$$\phi(t') = \sqrt{2}\sum_{k\in\mathbb{Z}} \alpha_k^j \phi(2t'-k).$$

This is just the dilation equation (2.83), with

$$\alpha_k^j = h_0(k) \tag{2.86}$$

independent of the scale level j. So, if sampled on t, (2.85) can be viewed as a signal output from a convolution of the input with h_0 which does not depend on time. This time-invariant operator, h_0*: $V_{j+1} \longrightarrow V_j$, is called a **low-pass filter**. Likewise,

$$\begin{aligned} W_j \subset V_{j+1} \quad &\Longrightarrow \quad \psi_0^j(t) = \sum_{k\in\mathbb{Z}} \beta_k^j \phi_k^{j+1}(t) \\ &\Longrightarrow \quad \beta_k^j = h_1(k) \qquad \text{independent of } j. \end{aligned}$$

And $h_1* : V_{j+1} \longrightarrow W_j$ is called a **high-pass filter**. Filters may be represented either by the *impulse response* or by *frequency response*. $\{h_0(k),\ k \in \mathbb{Z}\}$ forms the **impulse response** for the low-pass filter, while

$$H_0(\omega) = \sum_{k\in\mathbb{Z}} h_0(k) e^{-ik\omega}, \quad \omega \in \mathbb{R},$$

gives the **frequency response**. Similarly, we have two responses, $h_1(k)$, and $H_1(\omega)$, for the high-pass filter. For convenience, we will refer to a filter simply by its impulse or frequency response.

h_0 and h_1 form a set of filters, called a **filter bank**.[15] Filter banks are very important in building basis functions and facilitating function analysis, as we will see later.

Filters can be found from a given scaling function and/or wavelet function. For the low-pass filter, the dilation equation (2.83) implies

$$\begin{aligned} h_0(k) &= \int_{-\infty}^{\infty} \phi(t) \cdot \sqrt{2}\phi(2t-k)\ dt \\ &= \frac{1}{\sqrt{2}} \int_{-\infty}^{\infty} \phi(\frac{t}{2}) \phi(t-k)\ dt \end{aligned} \tag{2.87}$$

by orthonormality of $\{\phi_k^1(t)\}_{k\in\mathbb{Z}}$. The frequency response is also easy to obtain. In fact, the dilation and wavelet equations,

$$\begin{aligned} \phi(t) &= 2^{1/2} \sum_{k\in\mathbb{Z}} h_0(k)\phi(2t-k), \\ \psi(t) &= 2^{1/2} \sum_{k\in\mathbb{Z}} h_1(k)\phi(2t-k), \end{aligned}$$

[15] To be precise, it is an **analysis bank** here.

have a simpler representation in the frequency domain:

$$\hat{\phi}(\omega) = \frac{1}{\sqrt{2}} H_0(\frac{\omega}{2}) \hat{\phi}(\frac{\omega}{2}), \tag{2.88}$$

$$\hat{\psi}(\omega) = \frac{1}{\sqrt{2}} H_1(\frac{\omega}{2}) \hat{\phi}(\frac{\omega}{2}). \tag{2.89}$$

Eq. (2.88) yields

$$H_0(\frac{\omega}{2}) = \sqrt{2}\, \frac{\hat{\phi}(\omega)}{\hat{\phi}(\frac{\omega}{2})}. \tag{2.90}$$

In a similar way, the high-pass filter response may also be obtained. In real problems, however, there is seldom knowledge of ψ in advance. The resulting formulas for H_1 and h_1 in these forms thus are not of practical use. A common exercise to get them is from their low-pass counterparts. We will see an example in the next subsection.

There is a very interesting property for the filter bank obtained from an orthonormal scaling and wavelet basis: the **double shift orthonormality**. By the dilation equation and orthonormality of the scaling bases formed by ϕ,

$$\begin{aligned}
\delta_{m,0} &= \int_{\mathbb{R}} \phi(t)\phi(t-m)\ dt \\
&= \int_{\mathbb{R}} 2 \sum_{k\in\mathbb{Z}} h_0(k)\phi(2t-k) \cdot \sum_{l\in\mathbb{Z}} h_0(l)\phi(2t-2m-l)\ dt \\
&= \sum_k \sum_l h_0(k)h_0(l) \int_{\mathbb{R}} \phi_k^1(t)\ \phi_{2m+l}^1(t)\ dt \\
&= \sum_k \sum_l h_0(k)h_0(l) \cdot \delta_{k,2m+l}.
\end{aligned}$$

That is to say,

$$\sum_{k\in\mathbb{Z}} h_0(k)\ h_0(k-2m) = \delta_{m,0}, \qquad \forall\ m \in \mathbb{Z}. \tag{2.91}$$

Similarly we have, for h_1 and h_0,

$$\sum_{k\in\mathbb{Z}} h_0(k)\ h_1(k-2m) = 0, \qquad \forall\ m \in \mathbb{Z}, \tag{2.92}$$

by the perpendicularity between V_1 and W_1. Eqs. (2.91) and (2.92) describe the double-shift orthonormality for a filter bank.

2.5.2 Orthonormalization of the scaling basis

As mentioned before, one of the central issues in this study is the orthonormality of the analysis basis, on which the Parseval relation relies. Until now, we have assumed the scaling basis $\{\phi_n^j\}$ to be orthonormal. But what if we have in hand one which is non-orthonormal? In this subsection, we give a way to deal with this situation.

Suppose, for the time being, we have a translation invariant basis $\{\phi_*(t-k)\}_{k\in\mathbb{Z}}$ which is non-orthonormal. We want to find from it an orthonormal one $\{\phi(t-k)\}_{k\in\mathbb{Z}}$. In the time domain, the orthonormality of $\{\phi(t-k)\}_{k\in\mathbb{Z}}$ reads that

$$a(\phi, k) \equiv \int_{-\infty}^{\infty} \phi(t)\phi(t-k)dt = \delta(k),$$

which, if transformed to the frequency domain, becomes (Strang and Nguyen, 1997)

$$A(\hat{\phi}, \omega) = \sum_{n=-\infty}^{\infty} |\hat{\phi}(\omega + 2\pi n)|^2 = 1.$$

This condition is also sufficient, a detailed proof can be seen in Strang and Nguyen (1997, p. 205). We thus have

$$\{\phi(t-k)\}_{k\in\mathbb{Z}} \text{ orthonormal} \qquad \text{iff} \qquad A(\hat{\phi}, \omega) = 1. \tag{2.93}$$

If, for some dilation function ϕ_*, $A(\hat{\phi}_*, \omega) > 0$, condition (2.93) implies that we may orthonormalize $B_* = \{\phi_*(t-k)\}_{k\in\mathbb{Z}}$ in the frequency domain by dividing it by $\sqrt{A(\hat{\phi}_*, \omega)}$. But the positivity of $A(\hat{\phi}_*, \omega)$ is guaranteed by the Riesz basis property (cf. (2.4)) by Theorem 6.13 of Strang and Nguyen (1997); the orthonormalization of B_* is therefore easily attained.

Our next task is to show that the orthonormalized basis

$$B_{orth} \equiv \{\phi(t-k)\}_{k\in\mathbb{Z}}, \qquad \phi(t) = \mathcal{F}^{-1}\left(\frac{\hat{\phi}_*(\omega)}{\sqrt{A(\hat{\phi}_*, \omega)}}\right), \tag{2.94}$$

with $\mathcal{F}^{-1}$ being an inverse Fourier transform, is a scaling basis for V_0. To be specific, we need to show

- (i) $\phi(\omega)$ satisfies the dilation equation (2.88);
- (ii) B_{orth} is a basis of V_0.

Notice that here the shift orthonormality of B_{orth} is automatically satisfied. The following proof generally follows Truchetet (1998).[16]

Recall that $B_* = \{\phi_*(t-k)\}_{k\in\mathbb{Z}}$ is a scaling basis of V_0. ϕ_* must satisfy an equation in the form of (2.88) (The derivation of (2.83) and (2.88) don't require the orthonormality of $\{\phi_*(t-k)\}_{k\in\mathbb{Z}}$):

$$\hat{\phi}_*(\omega) = \frac{1}{\sqrt{2}} H_{*0}(\frac{\omega}{2})\hat{\phi}_*(\frac{\omega}{2}). \tag{2.95}$$

So

$$\hat{\phi}(\omega) = \frac{\hat{\phi}_*(\omega)}{\sqrt{A(\hat{\phi}_*,\omega)}} = \frac{1}{\sqrt{2}}\left(H_{*0}(\frac{\omega}{2})\cdot\frac{\sqrt{A(\hat{\phi}_*,\frac{\omega}{2})}}{\sqrt{A(\hat{\phi}_*,\omega)}}\right)\hat{\phi}(\frac{\omega}{2}),$$

which apparently satisfies the dilation equation with a low-pass filter

$$H_0(\frac{\omega}{2}) = H_{*0}(\frac{\omega}{2})\cdot\frac{\sqrt{A(\hat{\phi}_*,\frac{\omega}{2})}}{\sqrt{A(\hat{\phi}_*,\omega)}}. \tag{2.96}$$

This completes the first part of the proof.

The second part of proof requires one to show that the sequence $B_{orth} = \{\phi(t-k)\}_{k\in\mathbb{Z}}$ furnishes a basis for V_0. Observe that $A_{RT}(\omega) \equiv \sqrt{A(\hat{\phi}_*,\omega)}$ is 2π-periodic. It can be represented with the Fourier basis as

$$A_{RT}(\omega) = \sum_{k\in\mathbb{Z}} a_{RT}(k)e^{-i\omega k} \tag{2.97}$$

Hence

$$\begin{aligned}\hat{\phi}_*(\omega) &= \hat{\phi}(\omega)\cdot A_{RT}(\omega) \\ &= \left(\sum_k a_{RT}(k)e^{-i\omega k}\right)\cdot\hat{\phi}(\omega),\end{aligned} \tag{2.98}$$

which, when transformed to the time domain, becomes

$$\begin{aligned}\phi_*(t) &= \left(\sum_k a_{RT}(k)\cdot\mathcal{F}^{-1}\left(e^{-i\omega k}\hat{\phi}_{orth}(\omega)\right)\right) \\ &= \sum_k a_{RT}(k)\cdot\phi(t-k).\end{aligned} \tag{2.99}$$

[16] The English translation of the relevant part of this was generously provided by Dr. F. Nicolier of the University of Burgundy, France.

Translating t by any integer yields a similar equation. So all $f(t) \in \text{span}B_*$ can be written as a linear combination of the elements in another family $B_{orth} = \{\phi(t-k)\}_{k\in\mathbb{Z}}$. In other words, B_{orth} is a basis of the space which is the closure of the span of B_*, namely, V_0. Q.E.D.

In the above proofs we appeal to the low-pass filter bank many times to bring together the dilation functions at different scales. Actually it is also possible to find this filter bank in terms of the pre-orthonormalized ϕ_*. From Eq. (2.96),

$$H_0(\omega) = H_{*0}(\omega)\frac{\sqrt{A(\hat{\phi}_*, \omega)}}{\sqrt{A(\hat{\phi}_*, 2\omega)}}. \tag{2.100}$$

But from (2.95),

$$H_{*0}(\omega) = \frac{\sqrt{2}\hat{\phi}_*(2\omega)}{\hat{\phi}_*(\omega)}. \tag{2.101}$$

A relation bridging H_0 and $\hat{\phi}_*$ is then established:

$$H_0(\omega) = \sqrt{2}\frac{\hat{\phi}_*(2\omega)}{\hat{\phi}_*(\omega)} \cdot \frac{\sqrt{A(\hat{\phi}_*, \omega)}}{\sqrt{A(\hat{\phi}_*, 2\omega)}}. \tag{2.102}$$

Note here $H_0(0) \neq 1$, different from that in Strang and Nguyen (1997), as we are considering problems in space $L_2(\mathbb{R})$ rather than $L_\infty(\mathbb{R})$.

To summarize, a scaling basis of V_0, if not orthonormal, may be orthonormalized with the technique of (2.94). The resulting basis spans subspace V_0, and the corresponding scaling function satisfies the dilation equation. It is, consequently, also a scaling basis.

2.5.3 B-spline scaling and wavelet functions

With the above preparation, we are now ready to build our scaling and wavelet bases, the B-spline bases. The problem is to find the dilation and wavelet functions $\phi(t)$ and $\psi(t)$, for $t \in \mathbb{R}$.

B-splines and scaling functions

It is well known that splines provide bases for certain piecewise polynomials (e.g., Strang and Nguyen, 1997). They can be constructed to any degree. For the widely used cubic spline, it consists of polynomials of the third degree on unit intervals within $[-2, 2]$ with derivatives continuous up to the second order. Splines are also referred to as *B-splines*, a name with the origin not quite clear.[17] As will been seen soon, B-splines are symmetric scaling bases.

Splines can also be defined with time-domain box functions. $n+1$ box functions convolute to form an n-th degree B-spline. Particularly, a cubic B-spline is

$$\phi_3(t) = B(t) * B(t) * B(t) * B(t) \tag{2.103}$$

with the box function $B(t)$ defined as

$$B(t) = \phi_0(t) = \begin{cases} 1 & |t| < \frac{1}{2} \\ 0 & \text{otherwise.} \end{cases}$$

Fig. 2.3 shows the four splines of different degrees thus formed. Later on we will see that

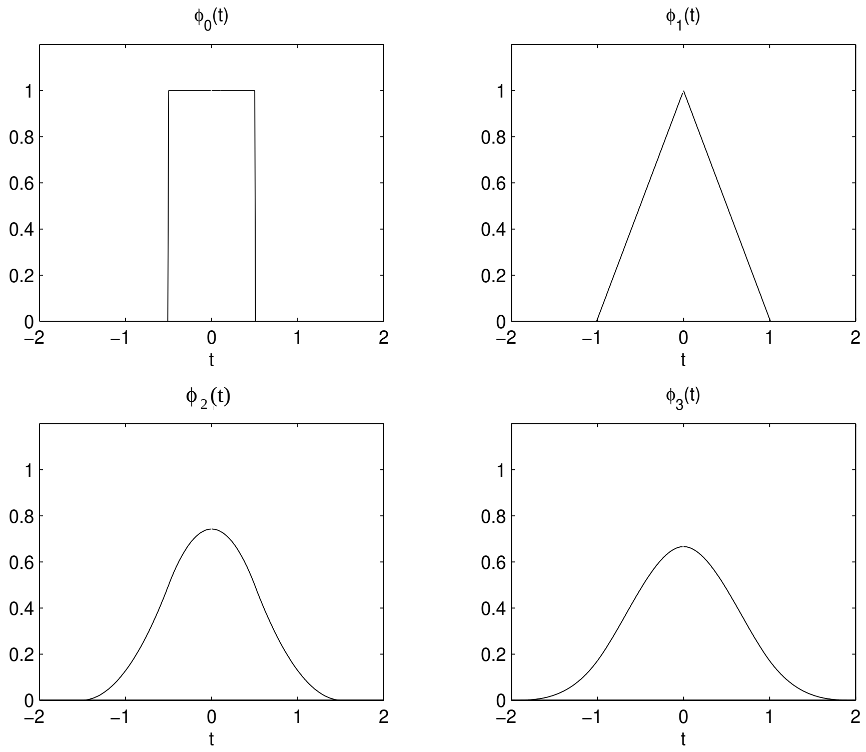

Figure 2.3: B-splines of zeroth, first, second, and third degrees

definition (2.103) allows an easy algebraic manipulation of ϕ_3 in the frequency domain.

[17]The letter "B" might signify "bell" or "basic".

The family of splines $\{\phi_n(t), n \geq 0\}$ has an appealing property, as established in Strang and Nguyen (1997, pp. 252-253). In their book, the abscissa in the box function diagram is shifted rightward by a half. As a result all the convolutions lie on the right hand side of the ordinate. In that coordinate frame, by the cascading algorithm (p. 234 therein), the average filter $(\frac{1}{2}, \frac{1}{2})$ leads to $\phi_0(t)$ (the scaling function for the Haar), $(\frac{1}{2}, \frac{1}{2}) * (\frac{1}{2}, \frac{1}{2})$ to the Mexican hat, and so on. In general, the convolution of $n+1$ average filters $(\frac{1}{2}, \frac{1}{2})$ iterates to $\phi_n(t)$. Therefore, $\phi_n(t)$, particularly $\phi_3(t)$, satisfies the dilation equation of multi-resolution analysis (since it comes from the recursive iteration of low-pass filters). It is thus a scaling or dilation function and its translation invariant family forms a scaling basis (see Strang and Nguyen, 1997, for details).

The orthonormalized scaling function and low-pass filter

The cubic B-spline has a good compact support. The corresponding filter is accordingly an FIR (finite impulse response). It is also symmetric, if the data points are evenly spaced (we don't consider irregularly spaced series in this study). It is , however, not orthonormal on translating. This can easily be seen from Fig. 2.3. In fact, Daubechies (1992) has established that no orthogonal scaling basis with finite support can ever be found in the B-spline family except for the Haar system. In other words, if we insist on orthogonality, the compactness has to be abandoned; the two properties are not consistent within this family. For this reason, the B-spline bases constructed are usually biorthonormal instead of orthonormal. In our problem, however, orthonormality is of paramount importance, while compactness is not that essential. A basis which is "effectively supported" on some finite domain is enough to meet the MS-EVA needs (the localization requirement, see below). Our objective is, in spite of its non-compactness, to find from $\phi_3(t)$ an orthonormal scaling basis $\{\phi(t-k)\}_{k \in \mathbb{Z}}$. As established in the previous subsection, ϕ may be simply be taken to be the orthonormalized $\phi_3(t)$. The following shows how the orthonormalization is achieved.

Taking Fourier transform of Eq. (2.103), one obtains

$$\hat{\phi}_3(\omega) = \mathrm{sinc}^4(\frac{\omega}{2}) = \left(\frac{\sin \omega/2}{\omega/2}\right)^4 \tag{2.104}$$

where ω is frequency in radians. The orthonormalization is gained by dividing $\hat{\phi}_3(\omega)$ by

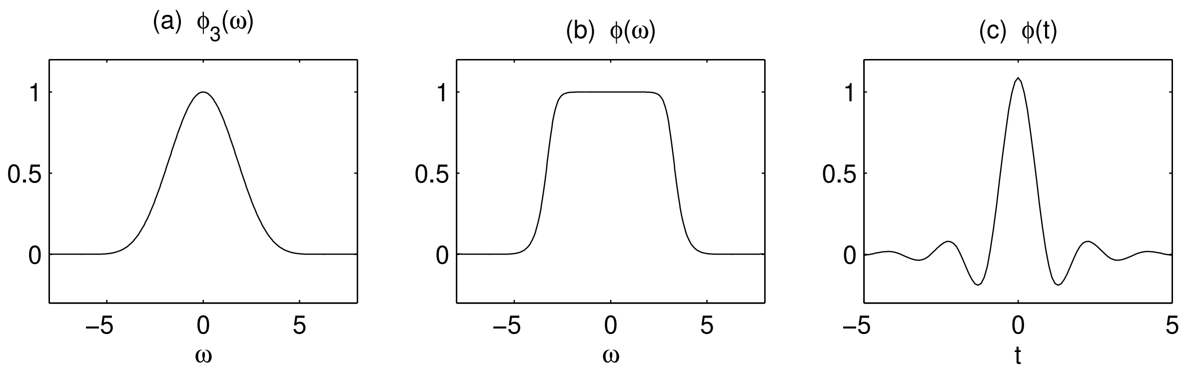

Figure 2.4: The cubic B-spline (in frequency domain ω in radians) (a), and its orthonormalization (in both frequency domain (b) and time domain (c)).

$\sqrt{A(\hat{\phi}_3, \omega)}$:

$$\hat{\phi}(\omega) = \frac{\hat{\phi}_3(\omega)}{\sqrt{A(\hat{\phi}_3, \omega)}}, \tag{2.105}$$

where

$$A(\hat{\phi}_3, \omega) = \sum_{n=-\infty}^{\infty} |\hat{\phi}_3(\omega + 2\pi n)|^2. \tag{2.106}$$

An inverse Fourier transform of $\hat{\phi}(\omega)$ (Use the package from Press et al., 1993) yields the desired $\phi(t)$, which is plotted in Fig. 2.4.

The $\phi(t)$ obtained is symmetric about and maximized at $t = 0$. By translating it furnishes an orthonormal basis of V_0. The localization of $\phi(t)$ has also been studied. It decays exponentially as $|t|$ increases (Holschneider, 1995, p. 193), a property which guarantees the polynomial localization which has been investigated before. It is not our intention here to go into details of the localization problem. We just want to repeat a fact that the ϕ thus-obtained meets the localization requirement: $|\phi(t)| \leq \kappa_\gamma^c(t)$ with $\gamma > 1$. As an example, Fig. 2.5 shows that $|\phi(t)| < \kappa_2^c(t)$, where $c = \phi(0) = 1.0889$, and κ_2^c is of order $\gamma = 2$. Here the localization order is twice the threshold value $\gamma = 1$, implying a fast convergence rate for the series summations involved in periodizing the bases. Besides, the ϕ obtained satisfies (figure not shown)

$$|\phi(n)| < \frac{\phi(0)}{(1+n^2)^{\gamma^*/2}}, \quad \forall\, n \in \mathbb{Z},$$

with γ^* being a number greater than 3.6. Signals with length N larger than 8 will then meet the inequality constraint (2.78) of Theorem 2.4.1 (cf. p. 62).

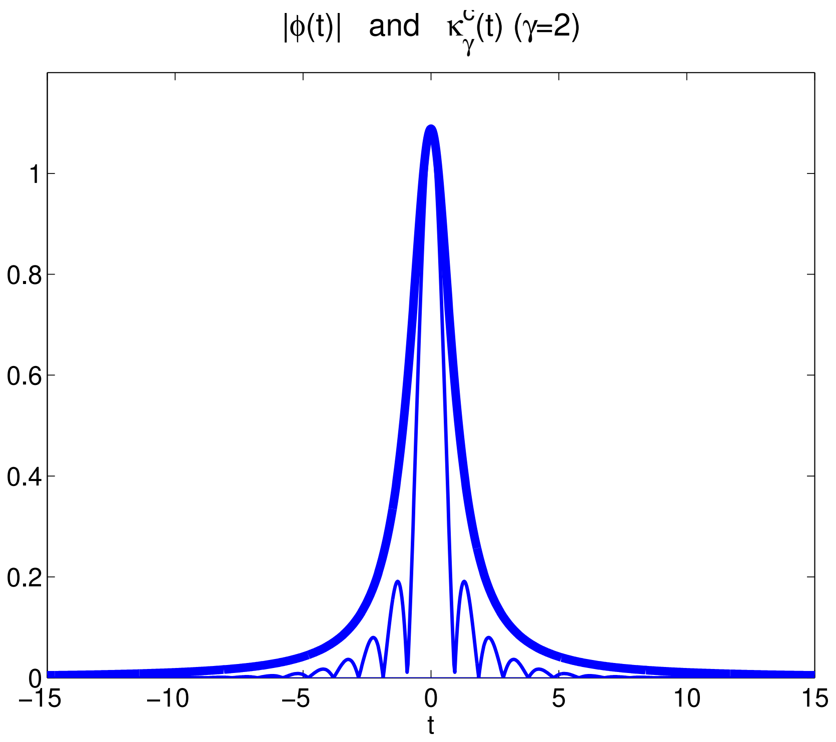

Figure 2.5: The absolute cubic B-spline scaling function $|\phi(t)|$ versus $\kappa_\gamma^c(t)$ (thick) with γ=2, C=$\phi(0) = 1.09$.

Associated with $\phi(t)$ is the low-pass filter. It is now easy to obtain, either by Eq. (2.87),

$$h_0(k) = \frac{1}{\sqrt{2}} \int_{-\infty}^{\infty} \phi(\frac{t}{2}) \phi(t-k) dt, \tag{2.107}$$

or by (2.88),

$$\begin{aligned} H_0(\omega) &= \sum_k h_0(k) e^{-ik\omega} \\ &= \frac{\sqrt{2}\ \hat{\phi}(2\omega)}{\hat{\phi}(\omega)}. \end{aligned} \tag{2.108}$$

It is also easy to verify that the obtained sequence $\{h_0(k)\}_{k\in\mathbb{Z}}$ is double shift orthonormal, as expected from the orthonormality of $\{\phi(t-k)\}_{k\in\mathbb{Z}}$.

The wavelet function and high-pass filter

A wavelet basis is needed in computing the localized power spectrum for scale window identification. Wavelet bases are formed by dilating and translating wavelet functions $\psi(t)$, while

$\psi(t)$ results from the high-pass filter $h_1(k)$ by Eq. (2.84), given the scaling function $\phi(t)$. There are many techniques available for the construction of h_1, and alternating flip is just one such (Strang and Nguyen, 1997, Eq. (2.17)). It states[18]

$$H_1(\omega) = e^{-i\omega}\overline{H_0(\omega + \pi)}. \tag{2.109}$$

Or, in the time domain

$$\begin{aligned} h_1(k) &= \int_{\mathbb{R}} H_1(\omega) e^{ik\omega} d\omega \\ &= \int e^{i\omega(k-1)} \overline{\sum_m h_0(m) e^{-im(\omega+\pi)}} d\omega \\ &= \sum_m (-1)^m h_0(m) \int e^{i\omega(m-1+k)} d\omega \\ &= \sum_m (-1)^m h_0(m) \delta(m-1+k) \\ &= (-1)^{1-k} h_0(1-k) \\ &= (-1)^{k+1} h_0(1-k). \end{aligned} \tag{2.110}$$

Notice this formula[19], differs from the usual ones like Strang and Nguyen (1997), where the power of (-1) is k,

$$h_1(k) = (-1)^k h_0(1-k). \tag{2.111}$$

It is easy to verify that both of these formulas give the desired double-shift orthonormality (2.92). That is to say, what matters here is the alternating feature of the minus sign, rather than its location of appearance. Both (2.111) and (2.110) may be used. But keep in mind, the time-domain counterpart of (2.109) is (2.110).

In addition to alternating flip, other approaches are also available to the building of $h_1(k)$ or $H_1(\omega)$. The QMF (Quadrature Mirror Filters, or alternating sign), for example, is just one of these approaches.

With the high-pass filter available, the wavelet function is now easily to be obtained. By (2.84):

$$\hat{\psi}(\omega) = \frac{1}{\sqrt{2}} H_1(\frac{\omega}{2}) \cdot \hat{\phi}(\frac{\omega}{2}). \tag{2.112}$$

[18] In this subsection, overline denotes the complex conjugate.

[19] Other forms may also be possible. Actually one may choose, for any *odd* N, the alternating flip to be

$$h_1(k) = (-1)^k h_0(N-k).$$

Again, the $(-1)^k$ can be replaced by $(-1)^{k+1}$.

Inversion of this equation gives the time domain representation of $\psi(t)$, the wavelet function we are seeking for. Figure (2.6) plots the ϕ and ψ obtained, and the corresponding filters h_0 and h_1.

2.5.4 Wrap-up for this section

We have obtained, from the cubic B-spline function $\phi_3(t)$, an orthonormal scaling basis $\{\phi(t-k)\}_{k\in\mathbb{Z}}$ (and its corresponding wavelet basis) adequate for the multiscale window transform. B-splines form a class of scaling bases, also Riesz bases, with compact supports. They are, however, not orthonormal in the space where they are considered. By the results of §2.5.2, these Riesz bases can be conveniently orthonormalized, and the resulting orthonormal outputs retain their respective dilation properties. The construction procedure of $\phi(t)$ is thereby simply the orthonormalization of $\{\phi(t-k)\}_{k\in\mathbb{Z}}$. The $\phi(t)$ thus-formed is symmetric about and maximized at zero, and has a polynomial localization of order $\gamma > 2$. For later use, its corresponding filter bank and wavelet basis are also acquired.

2.6 Computation of the multiscale window transform and synthesis

Given a $p \in V_{\varrho,j_2}$, the key to its multiscale window analysis is finding the scaling transform

$$\widehat{p}_n^j = \int_0^{\varrho} p(t)\ \phi_n^{\varrho,j}(t)\ dt,$$

and, if necessary, the wavelet transform

$$\widetilde{p}_n^j = \int_0^{\varrho} p(t)\ \psi_n^{\varrho,j}(t)\ dt$$

for the three window bounds $j = j_0, j_1, j_2$. Once they are obtained, all the window transforms and syntheses then can be easily computed by their respective definitions.

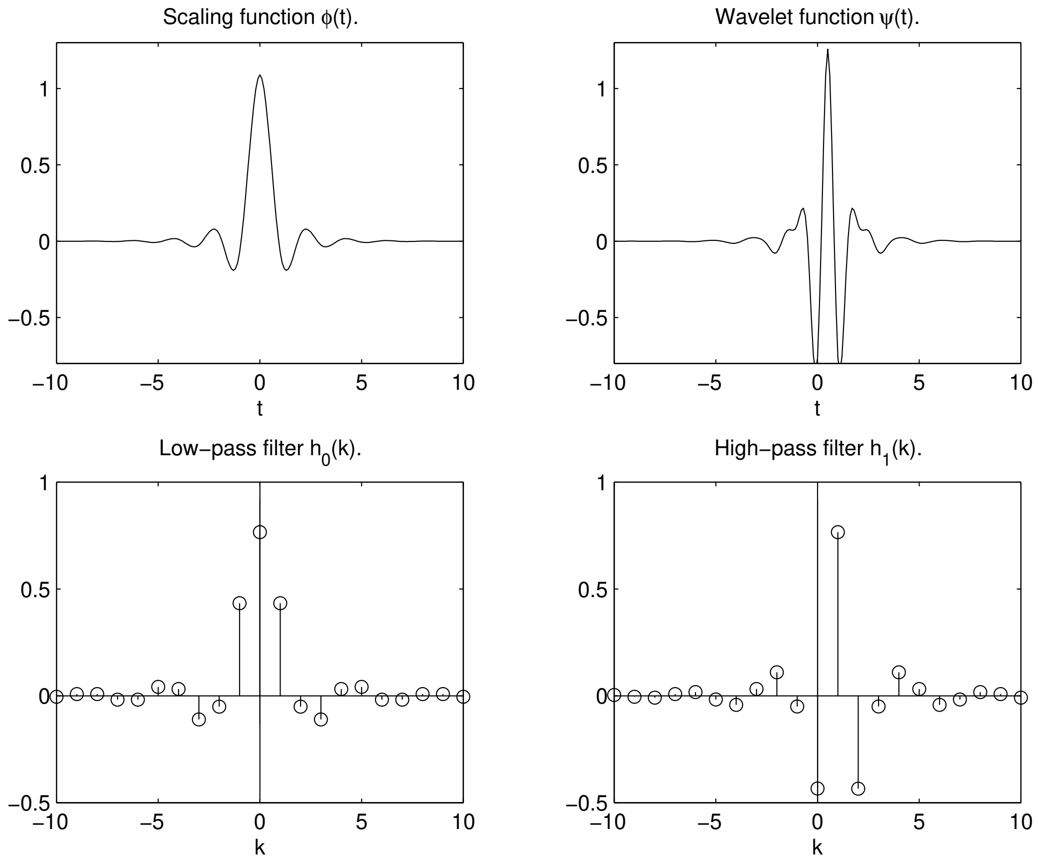

Figure 2.6: The orthonormalized cubic B-spline scaling and wavelet functions and their corresponding filter bank.

2.6.1 Fast scaling transform below level j_2

Suppose that we already know $\widehat{p}_n^{j+1}$ for some scale level j. The projection of $p(t)$ on $V_{\varrho,j+1}$, the subspace generated by

$$\{\phi_n^{\varrho,j+1}\}_{n\in\mathcal{N}_\varrho^j} = \left\{2^{(j+1)/2}\sum_{\ell\in\mathbb{Z}}\phi(2^{j+1}(t+\varrho\ell)-n)\right\}_{n\in\mathcal{N}_\varrho^{j+1}}$$

is represented as

$$p_{j+1}(t) = \sum_{n\in\mathcal{N}_\varrho^{j+1}} \widehat{p}_n^{j+1}\phi_n^{\varrho,j+1}(t). \tag{2.113}$$

By multi-resolution analysis, this can be decomposed into a sum of two terms, i.e.,

$$\sum_{m\in\mathcal{N}_\varrho^{j+1}} \widehat{p}_m^{j+1}\phi_m^{\varrho,j+1}(t) = \sum_{n\in\mathcal{N}_\varrho^j}\widehat{p}_n^j\phi_n^{\varrho,j}(t) + \sum_{n\in\mathcal{N}_\varrho^j}\widetilde{p}_n^j\psi_n^{\varrho,j}(t). \tag{2.114}$$

Using the orthogonality of $V_{\varrho,j}$ and $W_{\varrho,j}$, we may compute from this equation the scaling transform of p at location n and level j:

$$\begin{aligned}
\widehat{p}_n^j &= \int_0^\varrho \sum_{m\in\mathcal{N}_\varrho^{j+1}} \widehat{p}_m^{j+1}\phi_m^{\varrho,j+1}(t)\phi_n^{\varrho,j}(t)\ dt \\
&= \sum_{m\in\mathcal{N}_\varrho^{j+1}} \widehat{p}_m^{j+1}\int_0^\varrho \phi_m^{\varrho,j+1}(t)\ \phi_n^{\varrho,j}(t)\ dt \\
&= \sum_{m\in\mathcal{N}_\varrho^{j+1}} \widehat{p}_m^{j+1}\sum_{\ell_1\in\mathbb{Z}}\sum_{\ell_2\in\mathbb{Z}}\int_0^\varrho \phi_m^{j+1}(t+\varrho\ell_1)\ \phi_n^j(t+\varrho\ell_2)\ dt \\
&= \sum_{m\in\mathcal{N}_\varrho^{j+1}} \widehat{p}_m^{j+1}\sum_{\Delta\ell\in\mathbb{Z}}\sum_{\ell_2\in\mathbb{Z}}\int_{\varrho\ell_2}^{\varrho(\ell_2+1)} \phi_m^{j+1}(t'+\varrho\Delta\ell)\ \phi_n^j(t')\ dt' \\
&\qquad\qquad (t' = t+\varrho\ell_2,\ \Delta\ell = \ell_1-\ell_2) \\
&= \sum_{m\in\mathcal{N}_\varrho^{j+1}} \widehat{p}_m^{j+1}\sum_{\Delta\ell\in\mathbb{Z}}\int_{\mathbb{R}} \phi_m^{j+1}(t'+\varrho\Delta\ell)\ \phi_n^j(t')\ dt' \\
&= \sum_{m\in\mathcal{N}_\varrho^{j+1}} \widehat{p}_m^{j+1}\sum_{\Delta\ell\in\mathbb{Z}}\int_{\mathbb{R}} \sqrt{2}\phi(2t''+2n-m+2^{j+1}\varrho\Delta\ell)\ \phi_n^j(t'')\ dt'' \\
&\qquad\qquad (t'' = 2^jt'-n) \\
&= \sum_{m\in\mathcal{N}_\varrho^{j+1}} \widehat{p}_m^{j+1}\sum_{\Delta\ell\in\mathbb{Z}}\int_{\mathbb{R}} \phi^1_{m-2n-2^{j+1}\varrho\Delta\ell}(t'')\ \left[\sum_k h_0(k)\phi_k^1(t'')\right]\ dt'' \\
&\qquad\qquad \text{(dilation equation)} \\
&= \sum_{m\in\mathcal{N}_\varrho^{j+1}} \widehat{p}_m^{j+1}\sum_{\Delta\ell\in\mathbb{Z}}\sum_{k\in\mathbb{Z}} h_0(k)\int_{\mathbb{R}} \phi^1_{m-2n-2^{j+1}\varrho\Delta\ell}(t'')\phi_k^1(t'')\ dt''
\end{aligned}$$

$$= \sum_{m\in\mathcal{N}_\varrho^{j+1}} \widehat{p}_m^{j+1} \sum_{\Delta\ell\in\mathbb{Z}} \sum_{k\in\mathbb{Z}} h_0(k)\delta(k-(m-2n-2^{j+1}\varrho\Delta\ell)),$$

i.e.,

$$\begin{aligned} \widehat{p}_n^j &= \sum_{m\in\mathcal{N}_\varrho^{j+1}} \widehat{p}_m^{j+1} \sum_{\ell\in\mathbb{Z}} h_0(m-2n-2^{j+1}\varrho\ell) \\ &\equiv \sum_{m\in\mathcal{N}_\varrho^{j+1}} \widehat{p}_m^{j+1} h_0^{\varrho,j+1}(m-2n). \end{aligned} \tag{2.115}$$

The convergence of

$$h_0^{\varrho,j+1}(m-2n) = \sum_{\ell\in\mathbb{Z}} h_0(m-2n-2^{j+1}\varrho\ell) \tag{2.116}$$

is not in question. This can be seen from the equation (2.87), which we re-write here for easy reference

$$h_0(k) = \frac{1}{\sqrt{2}} \int_{\mathbb{R}} \phi(\frac{t}{2})\ \phi(t-k)\ dt. \tag{2.117}$$

From it h_0 may be viewed as the Fourier coefficient of a γ-order localized function $\phi(t/2)$ with respect to a γ-order localized basis $\{\phi(t-k)\}$, and hence it itself is also γ-order localized by the Theorem 11.0.2 of Holschneider (1995). As $\gamma > 1$, by the way in which ϕ is built, the infinite sum of (2.116) converges (cf., Lemma A2.5.1). Likewise, the wavelet transform

$$\begin{aligned} \widetilde{p}_n^j &= \sum_{m\in\mathcal{N}_\varrho^{j+1}} \widetilde{p}_m^{j+1} \sum_{\ell\in\mathbb{Z}} h_1(m-2n-2^{j+1}\varrho\ell) \\ &= \sum_{m\in\mathcal{N}_\varrho^{j+1}} \widetilde{p}_m^{j+1}\ h_1^{\varrho,j+1}(m-2n). \end{aligned} \tag{2.118}$$

Again, the convergence of the infinite sum is not an issue provided that ψ is polynomially localized up to an order $\gamma > 1$. Therefore, once $\widehat{p}_n^{j_2}$, the scaling transform at the highest level, is obtained, all the transforms below level j_2 can be computed efficiently with (2.115) and (2.118).

2.6.2 Scaling transform at the scale level j_2

The remaining issue is now is to find $\widehat{p}_n^{j_2}$. The basic idea has been already mentioned before in many places in §2.4. These pieces are re-organized in this subsection to make the story a whole. Following the same procedure, we distinguish two cases: extension by periodization and extension by reflection.

Extension by periodization

Recall measurements of a function $p(t)$ are available only at $2^{j_2} = N$ points: $t_n = \frac{n}{N}$, $n = 0, ..., N-1$. It has been justified in §2.4 that the $p(t)$ reconstructed from these measurements by interpolation lies in V_{ϱ,j_2}. When a periodic extension is used ($\varrho = 1$), this is to say,

$$p(t_m) = \sum_{n=0}^{N-1} \hat{p}_n^{j_2} \phi_n^{1,j_2}(t_m), \qquad m = 0, ..., N-1, \tag{2.119}$$

or, in a matrix form,

$$\underline{\underline{\mathbf{H}}}\, \underline{\hat{\mathbf{p}}}^{j_2} = \underline{\mathbf{p}}, \tag{2.120}$$

where $\underline{\mathbf{p}} = [p(t_0), p(t_1), ..., p(t_{N-1})]^T$, $\underline{\hat{\mathbf{p}}}^{j_2} = \left[\hat{p}_0^{j_2}, \hat{p}_1^{j_2}, ..., \hat{p}_{N-1}^{j_2}\right]^T$, and

$$\underline{\underline{\mathbf{H}}} = \begin{bmatrix} H_{0,0} & H_{0,1} & ... & H_{0,N-1} \\ H_{1,0} & H_{1,1} & ... & H_{1,N-1} \\ \vdots & \vdots & \vdots & \vdots \\ H_{N-1,0} & H_{N-1,1} & ... & H_{N-1,N-1} \end{bmatrix}, \tag{2.121}$$

with the entries

$$H_{mn} = \phi_n^{1,j_2}(t_m) = \sum_{l \in \mathbb{Z}} \phi_{n-Nl}^{j_2}(t_m), \qquad m, n = 0, ...N-1. \tag{2.122}$$

with the ϕ obtained, this $\underline{\underline{\mathbf{H}}}$ is invertible. So

$$\underline{\hat{\mathbf{p}}}^{j_2} = \underline{\underline{\mathbf{H}}}^{-1} \underline{\mathbf{p}}, \tag{2.123}$$

which gives the scaling transform of p at level j_2.

Notice that $\underline{\underline{\mathbf{H}}}$ is determined completely by the structure of the chosen space. It doesn't depend on the signal at all. Therefore the inverse of it may be obtained once and for all. Notice further that $H_{m,n+N} = H_{m,n}$, and $H_{m,n} = H_{m+\alpha,n+\alpha}$, for all integers α, m, and n that make the indices meaningful, which is due to the periodic property of $\phi_n^{1,j_2}(t)$. This simple fact implies that $\underline{\underline{\mathbf{H}}}$ is a circulant matrix, and hence $\underline{\underline{\mathbf{H}}}^{-1}$ is also circulant (e.g., Davis, 1979). This $\underline{\underline{\mathbf{H}}}^{-1}$ therefore behaves like a cyclic filter (except for a constant multiplier), pre-treating signals before they enters the fast analysis bank. For the orthonormal scaling basis built before, this "pre-filter" with $j_2 = 10$ is shown in Fig. 2.7. From the plot, it is rather weak (side lobes negligible compared to the value at zero), and will exert effects only on grid-size features.

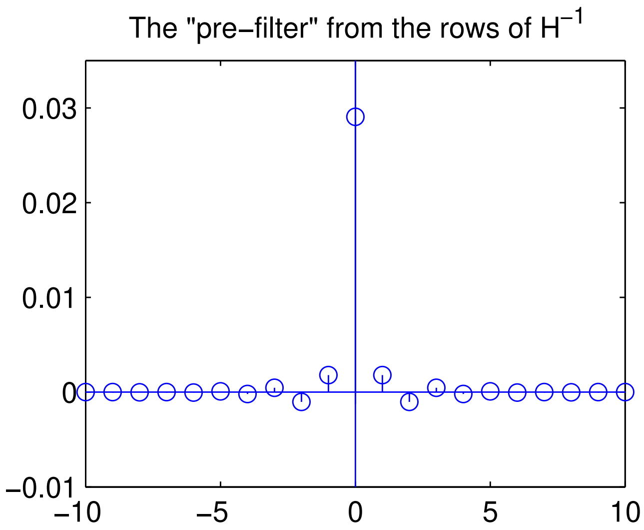

Figure 2.7: The "pre-filter" formed from the rows of $\underline{\underline{\mathbf{H}}}^{-1}$ in computing the scaling transforms at the highest scale level allowed by signal. $j_2 = 10$ in this example.

Extension by reflection

The idea to calculate $\widehat{p}_n^{j_2}$ with an extension by reflection has also been mentioned in §2.4. We first extend the discrete signal $p(t_n)$ from $[0, 1)$ to $[0, 2)$ as we did for x_n in (2.79), then starting from this extended series, extrapolate it periodically to the whole $\mathbb{R}$ (period=2 now). The computational problem with a symmetric extension now boils down to calculating the transform of $p(t)$ reconstructed from this extended signal. Let $q(t) = p(2t)$. $q(t)$ is then in V_{ϱ,j_2+1} and is formed by periodic extension from domain $[0, 1)$. By the result of the preceding subsection, we can obtain $\widehat{q}_n^{j_2+1}$ by solving equation

$$\sum_{n=0}^{2N-1} \widehat{q}_n^{j_2+1} \sum_{l\in\mathbb{Z}} \phi_{n-2Nl}^{j_2+1}(\frac{m}{2N}) = q(\frac{m}{2N}). \qquad (2.124)$$
$$m = 0, 1, ..., 2N - 1.$$

If N satisfies condition (2.78), so does $2N$. This equation must have a unique solution for $\widehat{q}_n^{j_2+1}$. By what we have shown in (2.82), $\widehat{p}_n^{j_2}$ is simply related to $\widehat{q}_n^{j_2+1}$ as:

$$\widehat{p}_n^{j_2} = \sqrt{2}\ \widehat{q}_n^{j_2+1}, \qquad n = 0, ..., 2N - 1. \qquad (2.125)$$

Consequently, the scaling transform of $p(t)$ at level j_2 can be conveniently computed, either with a periodic extension or a symmetric extension. Note $\widehat{p}_n^{j_2}$ is not p_n itself (up to a constant factor). The fast scaling/wavelet transform therefore doesn't apply directly to p_n. In other words, signals must be "pre-filtered" before entering the filter bank. (Strang and Nguyen, 1997, p. 232). In this study, the filter for the pre-filtering is from the rows of the circulant matrix $\underline{\underline{\mathbf{H}}}^{-1}$.

Once $\widehat{p}_n^{j_2}$ is obtained, all the scaling transforms below level j_2 can be obtained easily with the aid of Eqs. (2.115) and (2.118). These transforms, when substituted in (2.58) and (2.59), immediately yield the desired multiscale window analysis and synthesis.

2.6.3 Testing examples

To validate the multiscale window analysis developed in this chapter, two finite-lengthed signals

$$f_1(t) = e^{-\alpha(t-\frac{1}{2})^2} + ae^{-\beta(t-\frac{1}{2})^2}\sin\omega(t-\frac{1}{2}), \tag{2.126}$$

$$f_2(t) = t-\frac{1}{2} + ae^{-\beta(t-\frac{1}{2})^2}\sin\omega(t-\frac{1}{2}), \tag{2.127}$$

$t \in [0,1)$, which are highly localized in time and distinctly windowed on scales, are used to test whether their multiscale features can be correctly reconstructed. With the following parameters

$$\alpha = 25, \quad \beta = 1\times 10^4, \quad \omega = 150\pi, \quad a = 0.5,$$

these signals are plotted in Fig. 2.8 (top row). According to their distributions, we choose for $f_1(t)$ and $f_2(t)$ respectively a periodic extension and a symmetric extension, and perform and a two-scale window decomposition (windows 0 and 1) for the extended signals. In doing so, the upper level bound of the large-scale window (*i.e.*, window 0) is set to be $j_0 = 5$.

The results of the test are shown in Fig. 2.8, at the middle and bottom rows. In either case, the two-scale localized features have been well reproduced in the synthesis channel (bottom), with only some tiny error which is almost indiscernible by naked eye near the boundaries in the right-bottom subplot. The capability of our multiscale window transform is thus tested, satisfactorily with highly localized signals.

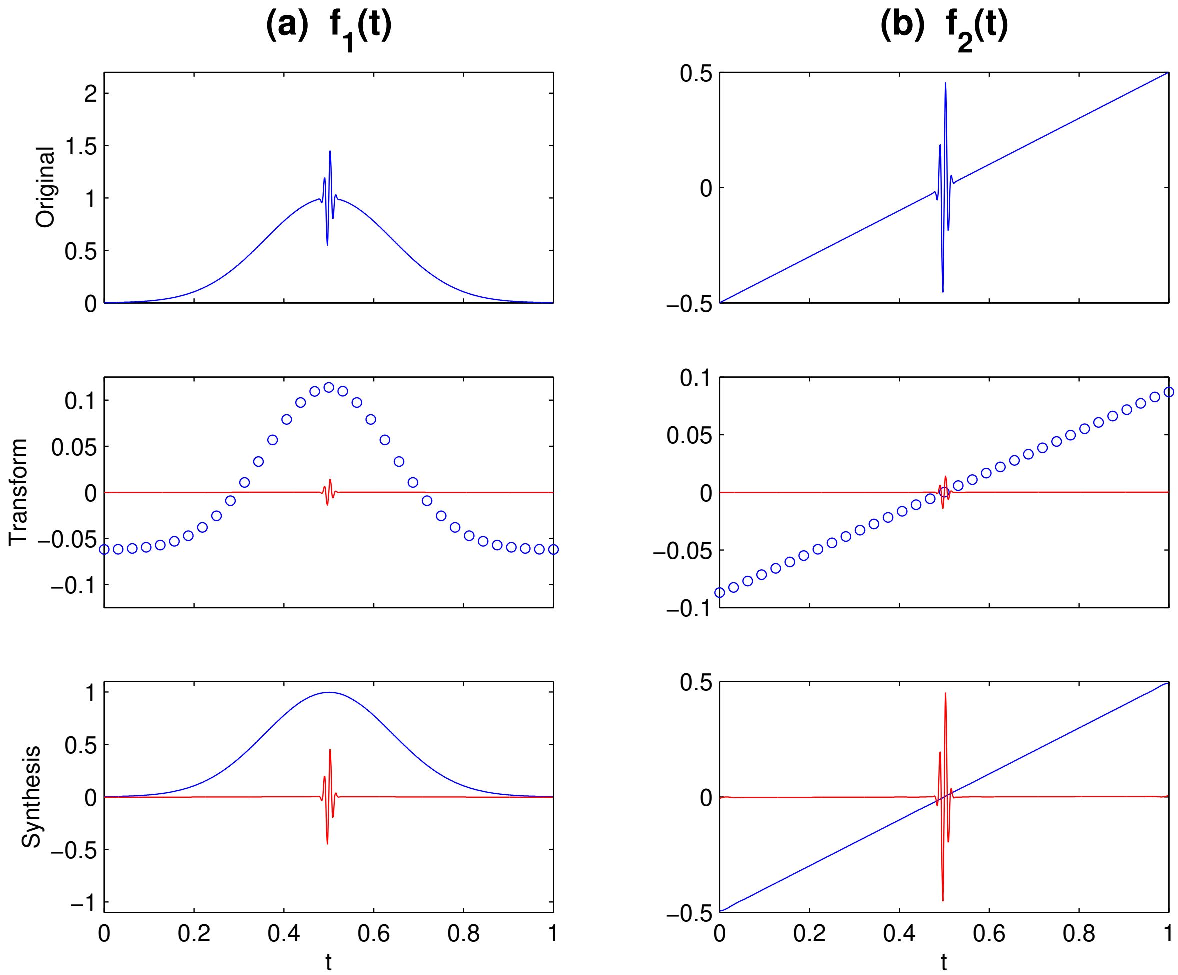

Figure 2.8: Testing examples of multiscale window analysis with functions $f_1(t)$ (column a), and $f_2(t)$ (column b). Sitting on the top are the original signals of f_1 and f_2, and the transforms and syntheses are shown at the middle and bottom rows, respectively. The blue is for scale window 0, and the red for window 1. Note the transforms are discretized in value. The window-1 distributions (the red "lines" at the middle row) look continuous because there are too many data points involved. The abscissa of the middle-row plots have been converted into locations in the physical space. In performing the analysis, $j_0 = 5$ is chosen and extensions by periodization and by reflection are applied to f_1 and f_2, respectively.

2.7 Two-dimensional multiscale window transform

As mentioned in Chapter 1, and will be more clear after Chapter 4, we need a multi-dimensional or multivariate large-scale window transform to handle the scale decomposition in the space direction. In this section, we illustrate how the one-dimensional (1-D) multiscale window transform developed above can be extended to a multivariate case (particularly, a 2-D case). In order to make the correspondence clear, the 1-D notations, such as those for filter responses, scaling functions, scale window transforms/syntheses, etc., will be used henceforth in this section without change. The dimensionalities of the field over which they are defined can be distinguished by the arguments they take, a practice resembling to the *function overloading* in the C++ computer language.

2.7.1 Dilation matrix and multi-resolution analysis

The extension of the 1-D transform begins with a replacement of the dilation factor "2" in the 1-D scaling function with an $n \times n$ matrix $\underline{\underline{\mathbf{M}}}$. This $\underline{\underline{\mathbf{M}}}$ is called a **dilation or sampling matrix**. It must meet the following requirements, as summarized in Kovacevic and Vetterli (1992):

(i) Each entry $m_{ij} \in \mathbb{Z}, \forall\ i, j = 1, 2, ..., n$;

(ii) Every eigenvalue $|\lambda_i| > 1, \forall\ i = 1, 2, ..., n$.

The corresponding dilation equation, parallel to the 1-D case, is

$$\phi(\underline{\mathbf{x}}) = \sqrt{N} \sum_{\underline{\mathbf{k}} \in \mathbb{Z}^n} h_0(\underline{\mathbf{k}}) \phi(\underline{\underline{\mathbf{M}}}\ \underline{\mathbf{x}} - \underline{\mathbf{k}}), \qquad \underline{\mathbf{x}} \in \mathbb{R}^n \tag{2.128}$$

with $N = |\det \underline{\underline{\mathbf{M}}}|$, the absolute value of the determinant of $\underline{\underline{\mathbf{M}}}$. Here $\underline{\mathbf{x}} \in \mathbb{R}^n$ is the coordinate, and $\underline{\mathbf{k}} \in \mathbb{Z}^n$ an integer vector. As in the 1-D case, $h_0(\underline{\mathbf{k}})$ represents a low-pass filter. The coefficient $\sqrt{N}$ appears so that the sequence $\{\sqrt{N}\phi(\underline{\underline{\mathbf{M}}}\ \underline{\mathbf{x}} - \underline{\mathbf{k}}),\ \underline{\mathbf{k}} \in \mathbb{Z}^n\}$ is orthonormal in the function space $L_2(\mathbb{R}^n)$. N itself is referred as the **dilation factor** because it measures the overall dilation effect. When $n = 1$, the only reasonable choice for $\underline{\underline{\mathbf{M}}}$ is 2. The dilation factor is then also 2. Compared to this 1-D case, multi-dimensional problems have more freedom in choosing $\underline{\underline{\mathbf{M}}}$, and therefore more dilation factors are possible. For instance, when $n = 2$, the dilation matrix could be

$$\underline{\underline{\mathbf{M}}}_1 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}, \qquad \underline{\underline{\mathbf{M}}}_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \qquad \underline{\underline{\mathbf{M}}}_3 = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}. \tag{2.129}$$

For the first case, $N = 4$, for the rest, $N = 2$. Dilation matrices like $\underline{\underline{\mathbf{M}}}_2$ or $\underline{\underline{\mathbf{M}}}_3$ are said to be **nonseparable**, since they cannot be linearly reduced to a diagonal form by real transformation. The opposite of the nonseparable dilation matrices are **separable**. $\underline{\underline{\mathbf{M}}}_1$ is just such an example.

A dilation matrix $\underline{\underline{\mathbf{M}}}$ determines a sampling scheme. Given an $\underline{\underline{\mathbf{M}}}$ with $|\det \underline{\underline{\mathbf{M}}}| = N$, since

$$\underline{\underline{\mathbf{M}}}(\mathbb{Z}^n) \equiv \{\underline{\underline{\mathbf{M}}}\ \underline{\mathbf{k}},\ \underline{\mathbf{k}} \in \mathbb{Z}^n\}$$

is a subspace of $\mathbb{Z}^n$, we can form

$$\mathbb{Z}^n / \underline{\underline{\mathbf{M}}}(\mathbb{Z}^n),$$

the quotient space of $\mathbb{Z}^n$ by $\underline{\underline{\mathbf{M}}}(\mathbb{Z}^n)$. It has been shown that this space contains only N elements-cosets (see Wojtaszczyk, 1997), which constitute close to the organ a *principal parallelepiped* on the sampling lattice generated by the $\underline{\underline{\mathbf{M}}}$ (strictly, by the basis of its column vectors). This parallelepiped is referred to be a **Voronoi cell** (cf., Kovacevic and Vetterli, 1992). Fig. 2.9 shows the Voronoi cells and cosets with respect to $\underline{\underline{\mathbf{M}}}_1(\mathbb{Z}^2)$ and $\underline{\underline{\mathbf{M}}}_2(\mathbb{Z}^2)$, respectively.

The question is, among the many choices of dilation or sampling matrices, which one should be picked? Before answering, we first look at the definition of a multi-resolution analysis over $\mathbb{R}^n$.

Definition 2.7.1 Given an $n \times n$ dilation matrix $\underline{\underline{\mathbf{M}}}$, a multi-resolution analysis with respect to it is a sequence of closed subspaces $\{V_j, j \in \mathbb{Z}\}$ of $L_2(\mathbb{R}^n)$ such that

(i) $V_i \subset V_j, \quad \forall\, i < j, \quad i, j \in \mathbb{Z}$; (**Nestedness**)

(ii) $\bigcup_{j \in \mathbb{Z}} V_j$ is dense in $L_2(\mathbb{R}^n)$; (**Totality**)

(iii) $\bigcap_{j \in \mathbb{Z}} V_j = \{0\}$; (**Emptiness**)

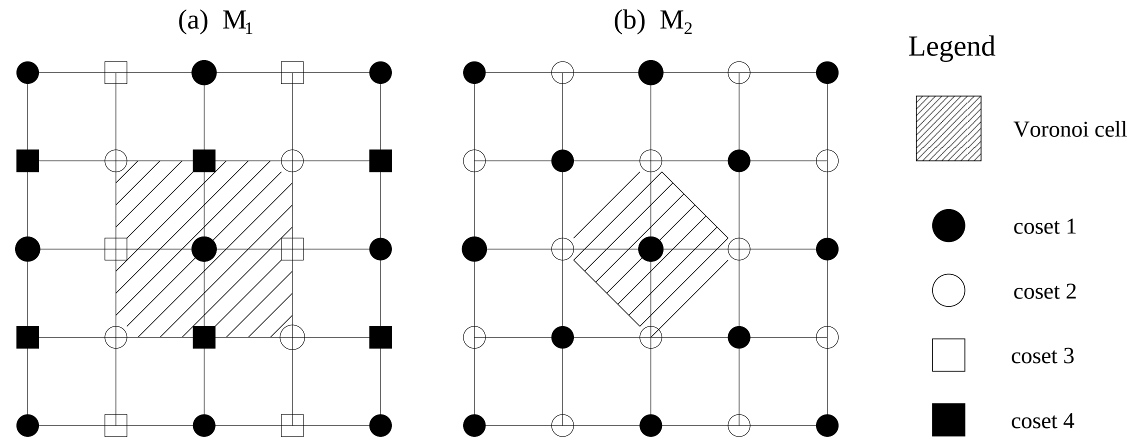

Figure 2.9: Voronoi cells and cosets with respect to $\underline{\underline{\mathbf{M}}}_1(\mathbb{Z}^2)$ (left) and $\underline{\underline{\mathbf{M}}}_2(\mathbb{Z}^2)$ (right).

(iv) V_0 is spanned by some orthonormal shift (or translation) invariant basis $\{\phi(\underline{\mathbf{x}} - \underline{\mathbf{k}}),\ \underline{\mathbf{k}} \in \mathbb{Z}^n\}$;

(v) $f(\underline{\mathbf{x}}) \in V_j$ iff $f(\underline{\underline{\mathbf{M}}}\underline{\mathbf{x}}) \in V_{j+1}$. **(Refinability)**

This definition corresponds exactly to its 1-D version given in §2.2.1. Denote

$$\phi^j_{\underline{\mathbf{k}}}(\underline{\mathbf{x}}) \equiv N^{j/2}\phi(\underline{\underline{\mathbf{M}}}^j\ \underline{\mathbf{x}} - \underline{\mathbf{k}}). \tag{2.130}$$

It has been established (see, for example, Meyer, 1992; Daubechies, 1993; Wojtaszczyk, 1997) that the sequence $\{V_j,\ j{\in}\mathbb{Z}\}$, where

$$V_j = \text{span}\{\phi^j_{\underline{\mathbf{k}}}(\underline{\mathbf{x}}),\ \underline{\mathbf{k}} \in \mathbb{Z}^n\},$$

does form a multi-resolution analysis, provided that

- $\phi(\underline{\mathbf{x}}) = N^{1/2}\sum_{\underline{\mathbf{k}}\in\mathbb{Z}^n} h_0(\underline{\mathbf{k}})\phi(\underline{\underline{\mathbf{M}}}\underline{\mathbf{x}} - \underline{\mathbf{k}}))$ converges in $L_2(\mathbb{R}^n)$;
- The Fourier transform of $\phi(\underline{\mathbf{x}})$, $\hat{\phi}(\underline{\mathbf{0}})$, is not zero, and $\hat{h}_0(\underline{\omega})$ is continuous at $\underline{\omega} = \underline{\mathbf{0}}$.

(Wojtaszczyk's theorem also requires orthonormality of $\{\phi(\underline{\mathbf{x}} - \underline{\mathbf{k}}\}$, but this is not an issue here as it is included in the above definition of $\phi^0_{\underline{\mathbf{k}}}(\underline{\mathbf{x}})$.)

According to the multi-resolution analysis, V_{j+1}, the **multi-resolution approximation** of $L_2(\mathbb{R}^n)$ to level $j{+}1$, can be written as

$$V_{j+1} = V_j \oplus V_j^{\perp}.$$

Meyer (1992) shows that, the orthogonal complementary $V_j^{\perp}$ can be further analyzed into a direct sum of $N-1$ mutually orthogonal subspaces ($N = |\det \underline{\underline{\mathbf{M}}}|$):

$$V_j^{\perp} = \bigoplus_{m=1}^{N-1} W_{jm}, \tag{2.131}$$

which, together with V_j, correspond(s) to the cosets with respect to $\underline{\underline{\mathbf{M}}}(\mathbb{Z}^n)$. So, when $N = 2$, only one subspace is admitted complement to V_j at that scale level. This is similar to the 1-D problem. But if $N = 4$, as is for $\underline{\underline{\mathbf{M}}} = \underline{\underline{\mathbf{M}}}_1$, the number of subspaces grows to 3, a multi-resolution analysis we have never met before in Chapter 2.

As explained by Meyer (1992), the above result may be understood intuitively as follows: When V_j gets dilated to V_{j+1}, the size of the measure is incremented by a factor of N. In other words, the space has been "enlarged" by N times. Therefore, V_{j+1} is the direct sum of V_j and $N-1$ other subspaces. A related fact is also seen in Fig. 2.9, where the Voronoi cell with respect to $\underline{\underline{\mathbf{M}}}_1(\mathbb{Z}^2)$ is twice as large as that with respect to $\underline{\underline{\mathbf{M}}}_2(\mathbb{Z}^2)$ or $\underline{\underline{\mathbf{M}}}_3(\mathbb{Z}^2)$.

Our previous questions about the choice of dilation matrix $\underline{\underline{\mathbf{M}}}$ can now be answered. For the analysis of V_{j+1}, when $N > 2$, the features represented in subspaces W_{jm} (m=1,...,N-1) are not just the details in the usual sense. They may involve averages (approximations) in one dimension but details in another, as that of the tensor product formulation (corresponding to $\underline{\underline{\mathbf{M}}}_1$). In other words, not all averages at level j are included in V_j. Some of them may also be dispersed in other subspaces at the same level. This is obviously not what we want for the local average. The dilation matrix $\underline{\underline{\mathbf{M}}}$ thus must be such that $N = |\det \underline{\underline{\mathbf{M}}}| = 2$. As only two-dimensional cases are considered, we choose the $\underline{\underline{\mathbf{M}}}_2$ of Eq. (2.129) henceforth.

2.7.2 The low-pass filter by McClellan's method

To begin constructing a scaling function, $\phi(\underline{\mathbf{x}})$, we need first find a 2-D low-pass filter, $\{h_0(\underline{\mathbf{k}})\}_{\underline{\mathbf{k}} \in \mathbb{Z}^2}$. There are many ways to approach the problem, Here we introduce a widely used one, **McClellan's method**.

McClellan's method is elegantly simple. It involves the extension of 1-D low-pass filter impulse functions. For a symmetric 1-D filter $h_0(x_1)$, its Fourier domain representation involves only cosines, which are definitely symmetric about the x_2-axis on a two-dimensional plane represented by coordinate pair (x_1, x_2). Mathematically, it is

$$\begin{aligned} H_0(\omega) &= \sum_{k \in Z} h_0(k) e^{-ik\omega} \\ &= h_0(0) + 2 \sum_{k=1}^{\infty} h_0(k) \cdot \cos(k\omega). \end{aligned} \tag{2.132}$$

Using the Tchebycheff polynomials $\{T_k, k = 0, 1, 2, ...\}$,[20] the cosine terms can be written as

$$\cos(k\omega) = T_k(\cos\omega). \tag{2.133}$$

It is at this point that McClellan applied his transform. Replacing $\cos\omega$ by a 2-D function $F(\omega_1, \omega_2)$, the filter is made into a 2-D one:

$$H_0(\omega_1, \omega_2) = h_0(0) + 2 \sum_{k=1}^{\infty} h_0(k) \cdot T_k[F(\omega_1, \omega_2)], \tag{2.134}$$

where $F(\omega_1, \omega_2)$ is required to be real, and its magnitude (absolute value) no larger than 1 in accordance with the replaced $\cos\omega$ (McClellan, 1973; Strang and Nguyen, 1997; Shapiro, 1994; Nicolier et al., 1998).

Equation (2.134) opens a vast variety of possibilities in building the 2-D filters. Shapiro (1994) proposed for $F(\omega_1,\ \omega_2)$ a form

$$F(\omega_1, \omega_2) = \frac{1}{2}(\cos\omega_1 + \cos\omega_2) + \frac{a}{2}(\cos\omega_1 - \cos\omega_2) + \frac{b}{8}(\sin 2\omega_1 \sin\omega_2 + \sin\omega_1 \sin 2\omega_2), \tag{2.135}$$

with parameters a and b varying (on condition that $|F| \leq 1$) to adapt to any specific problems. In the case of perfect isotropicity, $a = b = 0$, so only the first term is retained:

$$F(\omega_1, \omega_2) = \frac{1}{2}(\cos\omega_1 + \cos\omega_2). \tag{2.136}$$

In this appendix, we consider this simplest form only.

Using the 1-D $h_0(n)$ derived before from the cubic spline scaling function, the 2-D $h_0(\underline{\mathbf{k}})$ is easily obtained by inverse transforming (2.134). The result is shown in Fig. 2.10.

[20] By definition a Tchebycheff polynomial of the nth degree is

$$T_n(x) = \cos[n \cos^{-1}(x)],$$

for $|x| \leq 1$. It is usually calculated from the recursion relations (cf., Press et al., 1993):

$$\begin{cases} T_0(x) = 1 \\ T_1(x) = x \\ T_n(x) = 2xT_{n-1}(x) - T_{n-2}(x). \end{cases}$$

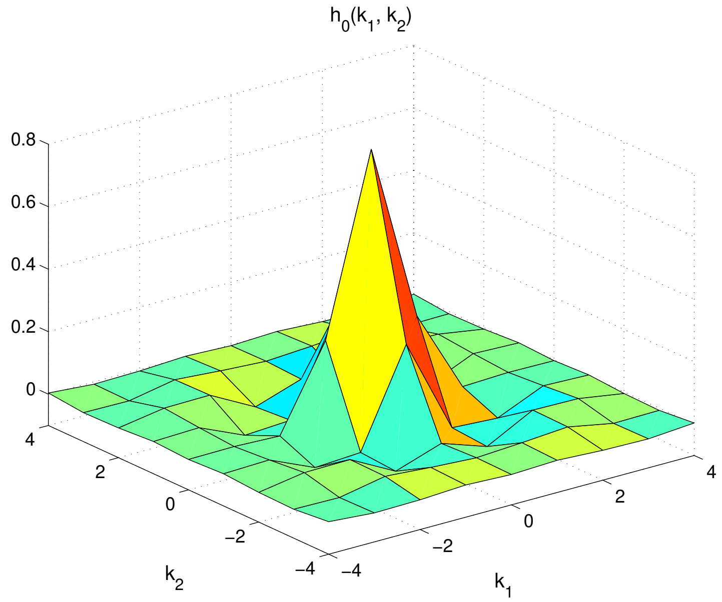

Figure 2.10: Impulse response of the 2-D low-pass filter $h_0(\underline{\mathbf{k}})$. Note $h_0(\underline{\mathbf{k}})$ is defined on the lattice $(k_1, k_2) \in \mathbb{Z}^2$. But for the sake of clarity, its discrete points are joined together with straight lines.

2.7.3 Scaling function from the low-pass filter with the cascade algorithm

The obtained 2-D low-pass filter $h_0(\underline{\mathbf{k}})$ is now adopted to build the scaling function $\phi(\underline{\mathbf{x}})$, with an algorithm called the **cascade algorithm** (Strang and Nguyen, 1997). The cascade algorithm relies on the convergence of an iteration, which is beyond the scope of this section. We just re-examine an example which has been used with success (see Nicolier et al., 1998, 1999, and references therein) and hence avoid invoking that issue. Let us assume there exists such a function, ϕ, such that

$$\phi(\underline{\mathbf{x}}) = N^{1/2} \sum_{\underline{\mathbf{k}} \in \mathbb{Z}^2} h_0(\underline{\mathbf{k}}) \phi(\underline{\underline{\mathbf{M}}}\, \underline{\mathbf{x}} - \underline{\mathbf{k}}), \tag{2.137}$$

where the factor $N^{1/2}$ is for normalization. The scheme for the cascade algorithm is henceforth

- (i) Start from some $\phi^0(\underline{\mathbf{x}})$,
- (ii) $\phi^{n+1}(\underline{\mathbf{x}}) = N^{1/2} \sum_{\underline{\mathbf{k}} \in \mathbb{Z}^2} h_0(\underline{\mathbf{k}}) \phi^n(\underline{\underline{\mathbf{M}}}\underline{\mathbf{x}} - \underline{\mathbf{k}}), \qquad n \geq 0,$
- (iii) Repeat (ii) until $\phi^n(\underline{\mathbf{x}})$ converges. The converged limit is the desired $\phi(\underline{\mathbf{x}})$.

Theoretically the starting function $\phi^0(\underline{\mathbf{x}})$ could be any element in $L_2(\mathbb{R}^2)$. In reality a careful choice of $\phi^0(\underline{\mathbf{x}})$ can greatly speed up the convergence. Since $\phi(\underline{\mathbf{x}})$ is symmetrical about the origin, we choose a box function

$$B(x_1, x_2) = \begin{cases} 1 & |x_1| \leq \frac{1}{2}, \quad |x_2| \leq \frac{1}{2} \\ 0 & \text{otherwise} \end{cases}$$

for the initialization.

The outcome of the iteration relies crucially on the dilation matrix. We have chosen

$$\underline{\underline{\mathbf{M}}} = \underline{\underline{\mathbf{M}}}_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix},$$

whose two column vectors, $(1,1)^T$ and $(1,-1)^T$, generate a quincunx lattice as shown in Fig. 2.9. The Voronoi cell on that lattice has an area only half as large as the one formed by tensor product ($\underline{\underline{\mathbf{M}}}_1$), and accordingly the normalization coefficient is $\sqrt{2}$ compared to the 2 in the latter case.

Figure 2.11 is the scaling function $\phi(\underline{\mathbf{x}})$ from the dilation matrix $\underline{\underline{\mathbf{M}}}_2$ and the low-pass filter $h_0(\underline{\mathbf{k}})$. The convergence is not slow. Twenty iterations are enough to have the relative error reduced to below $\frac{1}{100}$. We perform beyond this step 30 iterations more in case accuracy presents an issue.

2.7.4 Orthonormalization of the scaling basis

We have found the dilation function $\phi(\underline{\mathbf{x}})$, and hence a basis $\{\phi(\underline{\mathbf{x}} - \underline{\mathbf{k}}\}_{\underline{\mathbf{k}} \in \mathbb{Z}^2}$ for the subspace V_0. In what follows, we shall orthonormalize this basis. Once this is done, every dilation basis that derives therefore,

$$\{\phi^j_{\underline{\mathbf{k}}}\}_{\underline{\mathbf{k}} \in \mathbb{Z}^2} = \{N^{j/2} \phi(\underline{\underline{\mathbf{M}}}^j \underline{\mathbf{x}} - \underline{\mathbf{k}}), \ \underline{\mathbf{k}} \in \mathbb{Z}^2\},$$

will form a total orthonormal system in its corresponding subspace, V_j.

The orthonormalization of $\{\phi(\underline{\mathbf{x}} - \underline{\mathbf{k}})\}_{\underline{\mathbf{k}} \in \mathbb{Z}^2}$ is similar to what we have done for the 1-D case. It is based on the following fact, which is a generalization of the Theorem 6.10 of Strang and Nguyen (1997):

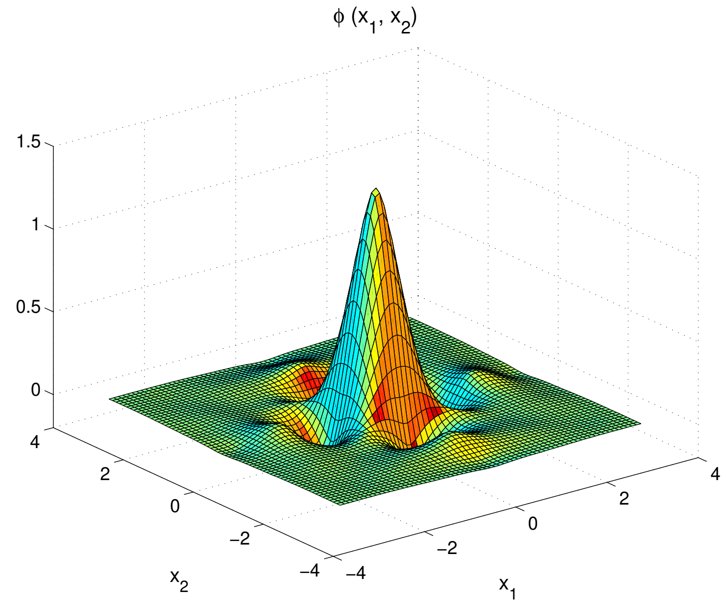

Figure 2.11: The 2-D pre-orthonormalized scaling function.

Theorem 2.7.1 *$\{\phi(\underline{\mathbf{x}}-\underline{\mathbf{k}})\}_{\underline{\mathbf{k}}\in\mathbb{Z}^2}$ is an orthonormal system of V_0 if and only if*

$$A(\hat{\phi},\underline{\omega}) = \sum_{\underline{\mathbf{n}}\in\mathbb{Z}^2} |\hat{\phi}(\underline{\omega}+2\pi\underline{\mathbf{n}})|^2 = 1,$$

where $\hat{\phi}(\underline{\omega})$ is the 2-D Fourier transform of $\phi(\underline{\mathbf{x}})$.

Proof

The proof is the 2-D version of that of Theorems 6.10 and 6.13 of Strang and Nguyen (1997), with the 1-D Fourier transforms therein replaced by their 2-D counterparts.

The theorem implies, given a Riesz basis $B_* = \{\phi_*(\underline{\mathbf{x}}-\underline{\mathbf{k}})\}_{\underline{\mathbf{k}}\in\mathbb{Z}^2}$ of V_0, we may form from it an orthonormal sequence

$$B_{orth} = \{\phi(\underline{\mathbf{x}}-\underline{\mathbf{k}}),\ \underline{\mathbf{k}}\in\mathbb{Z}^2\}, \tag{2.138}$$

where

$$\phi(\underline{\mathbf{x}}) = \mathcal{F}^{-1}\left(\frac{\hat{\phi}_*(\underline{\omega})}{\sqrt{A(\hat{\phi}_*,\underline{\omega})}}\right), \tag{2.139}$$

$$A(\hat{\phi}_*,\underline{\omega}) = \sum_{\underline{\mathbf{n}}\in\mathbb{Z}^2} |\hat{\phi}_*(\underline{\omega}+2\pi\underline{\mathbf{n}})|^2, \tag{2.140}$$

and $\mathcal{F}^{-1}$ the 2-D inverse Fourier transform. The remaining issue is to show that B_{orth} is a scaling basis of V_0. This can be done by following exactly what we did for the 1-D case (pp. 67-69), with appropriate modifications pertaining to the function arguments, and with the dilation matrix $\underline{\underline{\mathbf{M}}}$ instead of the dilation factor 2. Fig. 2.12a shows the basis function orthonormalized from that of Fig. 2.11.

Once an orthonormal $\{\phi(\underline{\mathbf{x}} - \underline{\mathbf{k}})\}_{\underline{\mathbf{k}} \in \mathbb{Z}^2}$ is found, the corresponding low-pass filter can be easily obtained from the dilation equation. Multiply (2.128) by $N^{1/2}\phi(\underline{\underline{\mathbf{M}}}\,\underline{\mathbf{x}} - \underline{\mathbf{k}})$ and integrate with respect to $\underline{\mathbf{x}}$ over $\mathbb{R}^2$. Using the orthonormality, we get

$$h_0(\underline{\mathbf{k}}) = \iint_{\mathbb{R}^2} N^{1/2}\ \phi(\underline{\underline{\mathbf{M}}}\ \underline{\mathbf{x}} - \underline{\mathbf{k}})\ \phi(\underline{\mathbf{x}})\ d\underline{\mathbf{x}}. \tag{2.141}$$

Formula (2.141) states how a filter can be determined from the scaling function. But in practice, we rarely exploit it to calculate the impulse response. It will be more convenient to achieve the goal in the frequency domain. Look at the dilation equation (2.128) again. It becomes, with a Fourier transform taken on either side,

$$\hat{\phi}(\underline{\omega}) = \frac{1}{\sqrt{N}} H_0\left((\underline{\underline{\mathbf{M}}}^{-1})^T \underline{\omega}\right)\ \hat{\phi}\left((\underline{\underline{\mathbf{M}}}^{-1})^T \underline{\omega}\right). \tag{2.142}$$

If the argument ω is replaced by $\left((\underline{\underline{\mathbf{M}}}^{-1})^T \underline{\omega}\right)$, we obtain from it a formula for $H_0(\omega)$

$$H_0(\underline{\omega}) = \sqrt{N} \frac{\hat{\phi}\left(\underline{\underline{\mathbf{M}}}^T \underline{\omega}\right)}{\hat{\phi}(\underline{\omega})}. \tag{2.143}$$

This is the frequency response of the low-pass filter, and an inverse Fourier transform gives the corresponding impulse response $h_0(\underline{\mathbf{k}})$. In Fig. 2.12b we plot such an $h_0(\underline{\mathbf{k}})$ derived from the $\phi(\underline{\mathbf{x}})$ shown in Fig. 2.12a.

2.7.5 Boundary extensions and the 2-D multiscale window transform and synthesis

As the 1-D case, practically, we only have functions defined on a finite domain. They have to be extended to $\mathbb{R}^2$ before the analysis or synthesis applies. Again, we consider for this purpose only two schemes: the extension by periodization (or periodic extension) and extension by reflection (or symmetric extension).

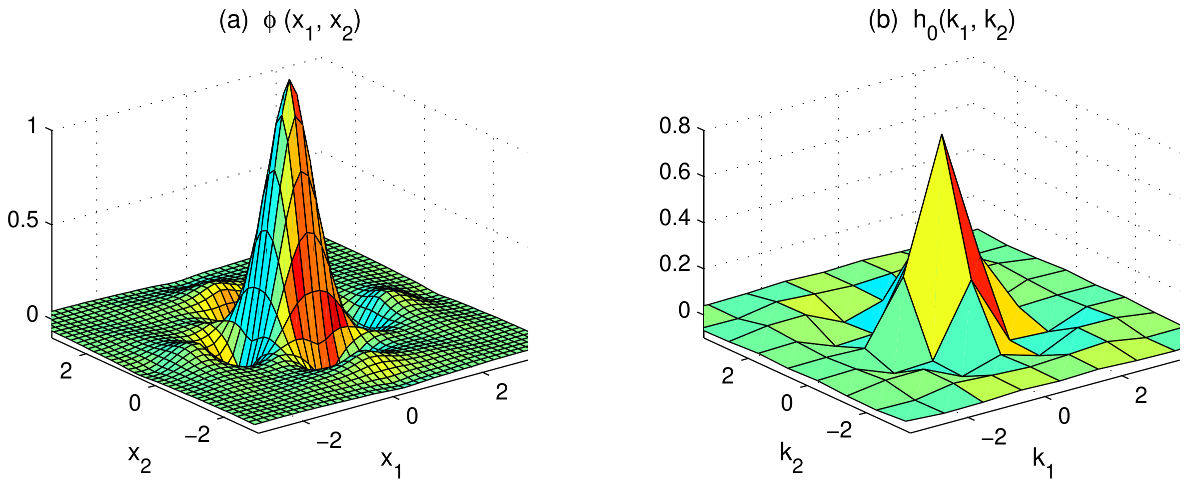

Figure 2.12: The orthonormalized scaling function (a) and its corresponding low-pass filter (b).

Suppose that we have a function $f = f(\underline{\mathbf{x}}) = f(x_1, x_2)$, $\underline{\mathbf{x}} = (x_1, x_2)^T$, defined on $\mathcal{D}$: $[0, 1) \times [0, 1)$. By a periodic extension of f (with periodicity one by default) from $\mathcal{D}$ to $\mathbb{R}^2$ we mean a new function f' over $\mathbb{R}^2$ such that

$$f'(x_1 + l_1, x_2 + l_2) = f'(x_1, x_2), \quad \underline{\mathbf{l}} = (l_1, l_2)^T \in \mathbb{Z}^2 \tag{2.144}$$

$$f'(x_1, x_2) = f(x_1, x_2) \qquad \text{if } \underline{\mathbf{x}} \in \mathcal{D}. \tag{2.145}$$

Likewise, a symmetric extension of f is an f' over $\mathbb{R}^2$ such that

$$f'(x_1, x_2) = f(x_1, x_2), \qquad \forall\ \underline{\mathbf{x}} \in \mathcal{D}, \tag{2.146}$$

but

$$f'(x_1 + l_1, x_2 + l_2) = f'(\mathrm{Odd}(l_1) + (-1)^{l_1} x_1, \mathrm{Odd}(l_2) + (-1)^{l_2} x_2), \tag{2.147}$$

where $(l_1, l_2) \in \mathbb{Z}^2$ and

$$\mathrm{Odd}(l) = \begin{cases} 1, & l \in \mathbb{Z},\ l \text{ odd} \\ 0. & l \in \mathbb{Z},\ l \text{ even} \end{cases} \tag{2.148}$$

In more general cases, a function $f(\underline{\mathbf{x}})$ defined on $[0, \varrho) \times [0, \varrho) \equiv \mathcal{D}_\varrho$, for any integer $\varrho > 0$, can be extended in a similar way. Particularly, a periodic extension of this f with periodicity

ϱ from $\mathcal{D}_\varrho$ to $\mathbb{R}^2$ is a function f' such that

$$f'(\underline{\mathbf{x}} + \varrho\underline{\mathbf{l}}) = f'(\underline{\mathbf{x}}), \qquad \underline{\mathbf{l}} \in \mathbb{Z}^2; \tag{2.149}$$

$$f'(\underline{\mathbf{x}}) = f(\underline{\mathbf{x}}), \qquad \underline{\mathbf{x}} \in \mathcal{D}_\varrho. \tag{2.150}$$

With this definition, the symmetric extension can be understood as an extension first by reflection from $\mathcal{D}$ to $[0,2)\times[0,2)$ (the symmetric extension restricted on $[0,2)\times[0,2)$), followed by a periodic extension with periodicity two. The above two schemes, therefore, may be treated in a unified approach by considering a periodic extension with an integer periodicity ϱ. For simplicity, suppose $\varrho = 2^\lambda$ with λ being some non-negative integer (0 or 1).

Functions through periodization with periodicity ϱ are identified with those defined on a 2-torus $\mathbb{T}^2_\varrho = \mathbb{T}_\varrho \times \mathbb{T}_\varrho$. As in the 1-D case, the problem is now studied in $L_2(\mathbb{T}^2_\varrho)$ instead of $L_2(\mathbb{R}^2)$. We need to construct a sequence of closed subspaces in $L_2(\mathbb{T}\varrho^2)$ in analogy with the multi-resolution analysis of $L_2(\mathbb{R}^2)$. Note it is not our intention to build such a sequence in a generic sense. We just want to fulfill our task with the spline scaling basis we have just obtained, in order to avoid invoking any unnecessary theoretical issues.

Recall V_j is generated by

$$\{\phi^j_{\underline{\mathbf{k}}}(\underline{\mathbf{x}})\}_{\underline{\mathbf{k}}\in\mathbb{Z}^2} = \{N^{j/2}\phi(\underline{\underline{\mathbf{M}}}^j\underline{\mathbf{x}} - \underline{\mathbf{k}}),\ \underline{\mathbf{k}} \in \mathbb{Z}^2\}.$$

From this $\phi^j_{\underline{\mathbf{k}}}(\underline{\mathbf{x}})$ we construct a new function

$$\phi^{\varrho,j}_{\underline{\mathbf{k}}}(\underline{\mathbf{x}}) \equiv \sum_{\underline{\mathbf{l}}\in\mathbb{Z}^2} \phi^j_{\underline{\mathbf{k}}}(\underline{\mathbf{x}} + \varrho\underline{\mathbf{l}}), \quad \forall\ \underline{\mathbf{k}} \in \mathbb{Z}^2. \tag{2.151}$$

With the ϕ obtained before, the convergence of this infinite sum is very fast and hence is not an issue. Obviously $\phi^{\varrho,j}_{\underline{\mathbf{k}}}(\underline{\mathbf{x}}) \in L_2(\mathbb{T}^2_\varrho)$ and, denoting by $\mathcal{V}_j$ the Voronoi cell with respect to $\underline{\underline{\mathbf{M}}}^j$, we have for it the following property:

Theorem 2.7.2 *$\left\{\phi^{\varrho,j}_{\underline{\mathbf{k}}}(\underline{\mathbf{x}})\right\}_{\underline{\mathbf{k}}\in\mathcal{V}_{j+2\lambda}}$, for $\underline{\mathbf{x}} \in \mathcal{D}(\underline{\mathbf{x}}) = [0,\varrho)\times[0,\varrho)$ and $\varrho = 2^\lambda$, furnishes an orthonormal system in $L_2(\mathbb{T}_{\varrho^2})$.*

Proof

For any $\underline{\mathbf{k}}, \underline{\mathbf{k}}' \in \mathcal{V}_{j+2\lambda}$,

$$\iint_{\mathcal{D}(\underline{\mathbf{x}})} \phi^{\varrho,j}_{\underline{\mathbf{k}}}(\underline{\mathbf{x}})\ \phi^{\varrho,j}_{\underline{\mathbf{k}}'}(\underline{\mathbf{x}})\ d\underline{\mathbf{x}}$$

$$
\begin{aligned}
&= \iint_{\mathcal{D}(\underline{\mathbf{x}})} \sum_{\underline{\ell}\in\mathbb{Z}^2} \phi^j_{\underline{\mathbf{k}}}(\underline{\mathbf{x}}+\varrho\underline{\ell}) \sum_{\underline{\ell}'\in\mathbb{Z}^2} \phi^j_{\underline{\mathbf{k}}}(\underline{\mathbf{x}}+\varrho\underline{\ell}')\ d\underline{\mathbf{x}} \\
&= \sum_{\Delta\underline{\ell}'\in\mathbb{Z}^2} \iint_{\mathbb{R}^2} \phi^j_{\underline{\mathbf{k}}}(\underline{\mathbf{y}})\ \phi^j_{\underline{\mathbf{k}}'}(\underline{\mathbf{y}}+\varrho\Delta\underline{\ell}')\ d\underline{\mathbf{y}} \qquad\qquad (\underline{\mathbf{y}}=\underline{\mathbf{x}}+\varrho\underline{\ell},\quad \Delta\underline{\ell}=\underline{\ell}'-\underline{\ell}) \\
&= \iint_{\mathbb{R}^2} \phi^j_{\underline{\mathbf{k}}}(\underline{\mathbf{y}})\ \phi^j_{\underline{\mathbf{k}}'}(\underline{\mathbf{k}}')\ d\underline{\mathbf{y}} + \sum_{\Delta\underline{\ell}'\neq\underline{\mathbf{0}}} \iint_{\mathbb{R}^2} \phi^j_{\underline{\mathbf{k}}}(\underline{\mathbf{y}})\ \phi^j_{\underline{\mathbf{k}}'-\varrho\underline{\underline{\mathbf{M}}}^j\Delta\underline{\ell}'}(\underline{\mathbf{y}})\ d\underline{\mathbf{y}}.
\end{aligned}
$$

If $\Delta\underline{\ell}'\neq\underline{\mathbf{0}}$, as $\underline{\mathbf{k}}'\in\mathcal{V}_{j+2\lambda}$, $\varrho\underline{\underline{\mathbf{M}}}^j\Delta\ell'-\underline{\mathbf{k}}'$ is never $\underline{\mathbf{0}}$. By the orthonormality of $\left\{\phi^j_{\underline{\mathbf{k}}}(\underline{\mathbf{x}})\right\}_{\mathbb{Z}^2}$ the second term hence vanishes. For the same reason, the first term is equal to $\delta(\underline{\mathbf{k}}-\underline{\mathbf{k}}')$. (Here we use $\delta(\underline{\ell})$ to signify $\delta(\ell_1)\delta(\ell_2)$ for any vector $\underline{\ell}=(\ell_1,\ell_2)^T$.) So

$$
\iint_{\mathcal{D}(\underline{\mathbf{x}})} \phi^{\varrho,j}_{\underline{\mathbf{k}}}(\underline{\mathbf{x}})\ \phi^{\varrho,j}_{\underline{\mathbf{k}}'}(\underline{\mathbf{x}})\ d\underline{\mathbf{x}} = \delta(\underline{\mathbf{k}}-\underline{\mathbf{k}}') \tag{2.152}
$$

for all $\underline{\mathbf{k}},\underline{\mathbf{k}}'\in\mathcal{V}_{j+\lambda}$, and the theorem is therefore proved.

Now let

$$
V_{\varrho,j} = \operatorname{span}\left\{\phi^{\varrho,j}_{\underline{\mathbf{k}}},\ \underline{\mathbf{k}}\in\mathcal{V}_{j+2\lambda}\right\}. \tag{2.153}
$$

By the above theorem, this space has an orthonormal basis $\left\{\phi^{\varrho,j}_{\underline{\mathbf{k}}}\right\}_{\underline{\mathbf{k}}\in\mathcal{V}_{j+2\lambda}}$, and hence a finite dimensionality dim $V_{\varrho,j}=2^{j+2\lambda}$. In addition, the following inclusion holds:

Theorem 2.7.3 $V_{\varrho,j}\subset V_{\varrho,j+1}$, *for all* $j=0,1,2,....$

Proof

To begin the proof, first observe

(1) With the low-pass filter h_0 we have built, $\sum_{\underline{\mathbf{k}}\in\mathbb{Z}^2}|h_0(\underline{\mathbf{k}})|<\infty$ and actually converges very fast. (We don't intend to deal with more general cases.)

(2) Since $V_j\subset V_{j+1}$, we have

$$
\phi^j_{\underline{\mathbf{k}}}(\underline{\mathbf{x}}) = \sum_{\underline{\mathbf{n}}\in\mathbb{Z}^2} \alpha_{\underline{\mathbf{k}},\underline{\mathbf{n}}}\phi^{j+1}_{\underline{\mathbf{n}}}(\underline{\mathbf{x}})
$$

where

$$
\alpha_{\underline{\mathbf{k}},\underline{\mathbf{n}}} = \iint_{\mathbb{R}^2} \phi^{j+1}_{\underline{\mathbf{n}}}(\underline{\mathbf{x}})\ \phi^j_{\underline{\mathbf{k}}}(\underline{\mathbf{x}})\ d\underline{\mathbf{x}}
$$

$$
\begin{aligned}
&= N^{j+1/2} \iint_{\mathbb{R}^2} \phi(\underline{\underline{\mathbf{M}}}^{j+1}\underline{\mathbf{x}} - \underline{\mathbf{n}})\ \phi(\underline{\underline{\mathbf{M}}}^{j}\underline{\mathbf{x}} - \underline{\mathbf{k}})\ d\underline{\mathbf{x}} \\
&= N^{1/2} \iint_{\mathbb{R}^2} \phi(\underline{\underline{\mathbf{M}}}\ \underline{\mathbf{x}}' - (\underline{\mathbf{n}} - \underline{\underline{\mathbf{M}}}\ \underline{\mathbf{k}}))\ \phi(\underline{\mathbf{x}}')\ d\underline{\mathbf{x}}' \qquad (\underline{\mathbf{x}}' = \underline{\underline{\mathbf{M}}}^{j}\underline{\mathbf{x}} - \underline{\mathbf{k}}) \\
&= h_0(\underline{\mathbf{n}} - \underline{\underline{\mathbf{M}}}\ \underline{\mathbf{k}}). \qquad \text{(by Eq. (2.141))}
\end{aligned}
$$

From these two facts, we have, for $\underline{\mathbf{x}} \in \mathcal{D}(\underline{\mathbf{x}})$,

$$
\begin{aligned}
\phi_{\underline{\mathbf{k}}}^{\varrho,j}(\underline{\mathbf{x}}) &= \sum_{\underline{\ell} \in \mathbb{Z}^2} \phi_{\underline{\mathbf{k}}}^{j}(\underline{\mathbf{x}} + \varrho\underline{\ell}) \\
&= \sum_{\underline{\ell} \in \mathbb{Z}^2} \sum_{\underline{\mathbf{n}} \in \mathbb{Z}^2} h_0(\underline{\mathbf{n}}\ \underline{\underline{\mathbf{M}}}\ \underline{\mathbf{k}})\ \phi_{\underline{\mathbf{n}}}^{j+1}(\underline{\mathbf{x}} + \varrho\underline{\ell}) \\
&= \sum \underline{\mathbf{n}} \in \mathbb{Z}^2 h_0(\underline{\mathbf{n}} - \underline{\underline{\mathbf{M}}}\ \underline{\mathbf{k}})\ \phi_{\underline{\mathbf{n}}}^{\varrho,j+1}(\underline{\mathbf{x}}) \\
&= \sum_{\underline{\mathbf{s}} \in \mathcal{V}_{j+1+2\lambda}} \left[\sum_{\underline{\mathbf{t}} \in \mathbb{Z}^2} h_0(\varrho\underline{\underline{\mathbf{M}}}^{j+1}\underline{\mathbf{t}} + \underline{\mathbf{s}} - \underline{\underline{\mathbf{M}}}\ \underline{\mathbf{k}}) \right] \phi_{\underline{\mathbf{s}}}^{\varrho,j+1}(\underline{\mathbf{x}}).
\end{aligned}
$$

The sum in the square bracket converges because of fact (1), and hence $V_{\varrho,j} \subseteq V_{\varrho,j+1}$. But we know dim $V_{\varrho,j} <$ dim $V_{\varrho,j+1}$, so $V_{\varrho,j} \subset V_{\varrho,j+1}$. Q.E.D.

Theorem 2.7.3 could lead to a *multi-resolution analysis* of $L_2(\mathbb{T}_\varrho^2)$, in analogy with its 1-D counterpart. But we don't need this analysis here. With the 2-D scaling function ϕ built before, it is justifiable that all functions of our interest should be contained in V_{ϱ,j_2}, for some positive integer j_2. In fact, as in the 1-D case, what we know about the functions in question are only the values on a discretized mesh. The whole picture of these functions can only be obtained approximately through some measure such as interpolation. For function $f(\underline{\mathbf{x}})$ defined on a square mesh grid with 2^{j_2} points in either dimension, we first enlarge it by $\varrho = 2^\lambda$ times in either dimension, depending on which scheme of extension is chosen, then transform the definition domain such the data points lie in the Voronoi cell $\mathcal{V}_{j_2+2\lambda}$. For simplicity, we still use f to denote the extended function and write the independent variable after transformation in its old form tagged with location indices. The justification is then translated into whether the algebraic equation set for α

$$
\sum_{\underline{\mathbf{k}} \in \mathcal{V}_{j_2+2\lambda}} \phi_{\underline{\mathbf{k}}}^{\varrho,j_2}(\underline{\mathbf{x}}_m)\ \alpha_{\underline{\mathbf{k}}} = f(\underline{\mathbf{x}}_m), \qquad \underline{\underline{\mathbf{M}}}^{j_2+2\lambda}\underline{\mathbf{x}}_m \in \mathcal{V}_{j_2+2\lambda} \tag{2.154}
$$

has a unique solution. With the ϕ constructed in the last subsection (again, we consider only this particular case to avoid more theoretical issues), the diagonal elements of the matrix thus formed dominates the sum of the off-diagonal entries by orders and hence the answer is positive.

With the above preparation, we next decompose the space where our problem is studied, V_{j_2}, into a direct sum of three scale windows:

- Large-scale window V_ϱ^L: V_{ϱ,j_0} $(j_0 \geq 0)$,
- Meso-scale window V_ϱ^M: $V_{\varrho,j_1} \ominus V_{\varrho,j_0}$ $(j_1 > j_0)$,
- Sub-mesoscale window V_ϱ^S: $V_{\varrho,j_2} \ominus V_{\varrho,j_1}$ $(j_2 > j_1)$,

where the value of $\varrho = 2^\lambda$ depends on which extension scheme is used. Multiscale window transform and synthesis can be introduced with this decomposition. As before, denote $[0,\varrho) \times [0,\varrho)$ as $\mathcal{D}(\underline{\mathbf{x}})$. For $f \in V_{\varrho,j_2}$, the 2-D scaling transform of f at level j is

$$\widehat{f}^{j}_{\underline{\mathbf{k}}} = \iint_{\mathcal{D}(\underline{\mathbf{x}})} f(\underline{\mathbf{x}})\ \phi^{\varrho,j}_{\underline{\mathbf{k}}}\ d\underline{\mathbf{x}}. \tag{2.155}$$

On basis of this we define the large-scale, meso-scale, and sub-mesoscale window syntheses, respectively,

$$\begin{aligned}
f^{\sim 0}(\underline{\mathbf{x}}) &= \sum_{\underline{\mathbf{k}} \in \mathcal{V}_{j_0+2\lambda}} \widehat{f}^{j_0}_{\underline{\mathbf{k}}}\ \phi^{\varrho,j_0}_{\underline{\mathbf{k}}}(\underline{\mathbf{x}}) && (2.156)\\
f^{\sim 1}(\underline{\mathbf{x}}) &= \sum_{\underline{\mathbf{k}} \in \mathcal{V}_{j_1+2\lambda}} \widehat{f}^{j_1}_{\underline{\mathbf{k}}}\ \phi^{\varrho,j_1}_{\underline{\mathbf{k}}}(\underline{\mathbf{x}}) - f^{\sim 0}(\underline{\mathbf{x}}) && (2.157)\\
f^{\sim 2}(\underline{\mathbf{x}}) &= f(\underline{\mathbf{x}}) - \sum_{\underline{\mathbf{k}} \in \mathcal{V}_{j_1+2\lambda}} \widehat{f}^{j_1}_{\underline{\mathbf{k}}}\ \phi^{\varrho,j_1}_{\underline{\mathbf{k}}}(\underline{\mathbf{x}}), && (2.158)
\end{aligned}$$

and accordingly:

$$\widehat{f}^{\sim w}_{\underline{\mathbf{k}}} = \iint_{\mathcal{D}(\underline{\mathbf{x}})} f^{\sim w}(\underline{\mathbf{x}})\ \phi^{\varrho,j_2}_{\underline{\mathbf{k}}}(\underline{\mathbf{x}})\ d\underline{\mathbf{x}}, \tag{2.159}$$

$$f^{\sim w}(\underline{\mathbf{x}}) = \sum_{\underline{\mathbf{k}} \in \mathcal{V}_{j_2+2\lambda}} \widehat{f}^{\sim w}_{\underline{\mathbf{k}}}\ \phi^{\varrho,j_2}_{\underline{\mathbf{k}}}(\underline{\mathbf{x}}), \tag{2.160}$$

for all the three windows $w = 0, 1, 2$. Eqs. (2.159) and (2.160) form the desired 2-D transform-synthesis pair.

2.7.6 Computational issues

From the definition, the key to the multiscale transform and synthesis is the computation of scaling transforms. In what follows, we discuss how these transforms may be efficiently found, with the aid of the low-pass filter introduced before.

For a function $f(\underline{\mathbf{x}})$ evaluated discretely on a square mesh grid with 2^{j_2} points in either dimension, the highest scale level is j_2, and hence the problem is studied in V_{ϱ,j_2} for some $\varrho = 2^\lambda$ (0 or 1 depending on whether a periodic or symmetric extension is adopted). The computation of $\widehat{f}^{j_2}_{\underline{\mathbf{k}}}$ is to finding the solution of equation set (2.154), and as in the 1-D case, this is identical to applying a weak filter to $f(\underline{\mathbf{x}})$. In other words, the function must be pre-filtered before entering the filter bank (see below).

Once the "pre-filtering" at scale level j_2 is done, the transforms below that level can be efficiently obtained. In fact, we have a fast algorithm for the computation. Given $f(\underline{\mathbf{x}}) \in V_{\varrho,j_2}$, the projection of it onto the subspace $V_{\varrho,j}$, $\forall\ 0 \le j < j_2$, which is generated by

$$\{\phi^{\varrho,j}_{\underline{\mathbf{k}}}\}_{\underline{\mathbf{k}}\in\mathcal{V}_{j+2\lambda}} = \{N^{j/2}\sum_{\underline{\mathbf{l}}\in\mathbb{Z}^2}\phi(\underline{\underline{\mathbf{M}}}^j(\underline{\mathbf{x}}+\varrho\underline{\mathbf{l}})-\underline{\mathbf{k}}), \quad \underline{\mathbf{k}}\in\mathcal{V}_{j+2\lambda}\},$$

can be represented as

$$f_j(\underline{\mathbf{x}}) = \sum_{\underline{\mathbf{k}}\in\mathcal{V}_{j+2\lambda}} \widehat{f}^j_{\underline{\mathbf{k}}}\phi^{\varrho,j}_{\underline{\mathbf{k}}}.$$

Since we have chosen an $\underline{\underline{\mathbf{M}}}$ such that $N = 2$, $f_j(\underline{\mathbf{x}})$ can be analyzed as a sum of two terms, i.e.,

$$\sum_{\underline{\mathbf{m}}\in\mathcal{V}_{j+2\lambda}} \widehat{f}^j_{\underline{\mathbf{m}}}\phi^{\varrho,j}_{\underline{\mathbf{m}}}(\underline{\mathbf{x}}) = \sum_{\underline{\mathbf{k}}\in\mathcal{V}_{j-1+2\lambda}} \widehat{f}^{j-1}_{\underline{\mathbf{k}}}\phi^{\varrho,j-1}_{\underline{\mathbf{k}}} + f'_{j-1}(\underline{\mathbf{x}}) \tag{2.161}$$

where

$$f'_{j-1}(\underline{\mathbf{x}}) \in V^{\perp}_{\varrho,j-1},$$

the orthogonal complement of $V_{\varrho,j-1}$ in $V_{\varrho,j}$. Multiply the above equation by $\phi^{\varrho,j-1}_{\underline{\mathbf{k}}}(\underline{\mathbf{x}})$ and integrate with respect to $\underline{\mathbf{x}}$ over $\mathcal{D}(\underline{\mathbf{x}}) = [0,\varrho) \times [0,\varrho)$. Because of the orthogonality between $V_{\varrho,j-1}$ and $V^{\perp}_{\varrho,j-1}$, it results in

$$\widehat{f}^{j-1}_{\underline{\mathbf{k}}} = \iint_{\mathcal{D}(\underline{\mathbf{x}})} \sum_{\underline{\mathbf{m}}\in\mathcal{V}_{j+2\lambda}} \widehat{f}^j_{\underline{\mathbf{m}}}\ \phi^{\varrho,j}_{\underline{\mathbf{m}}}(\underline{\mathbf{x}})\ \phi^{\varrho,j-1}_{\underline{\mathbf{k}}}(\underline{\mathbf{x}})\ d\underline{\mathbf{x}}$$

$$
\begin{aligned}
&= \sum_{\underline{\mathbf{m}}\in\mathcal{V}_{j+2\lambda}} \widehat{f}^{j}_{\underline{\mathbf{m}}} \iint_{\mathcal{D}(\underline{\mathbf{x}})} \sum_{\underline{\mathbf{l}}\in\mathbb{Z}^2} \phi^{j}_{\underline{\mathbf{m}}-\varrho\underline{\underline{\mathbf{M}}}^j\underline{\mathbf{l}}}(\underline{\mathbf{x}}) \sum_{\underline{\mathbf{l}}'\in\mathbb{Z}^2} \phi^{j}_{\underline{\mathbf{k}}-\varrho\underline{\underline{\mathbf{M}}}^j\underline{\mathbf{l}}'}(\underline{\mathbf{x}})\, d\underline{\mathbf{x}} \\
&= \sum_{\underline{\mathbf{m}}\in\mathcal{V}_{j+2\lambda}} \widehat{f}^{j}_{\underline{\mathbf{m}}} \sum_{\Delta l\in\mathbb{Z}^2} \iint_{\mathbb{R}^2} \phi^{j}_{\underline{\mathbf{m}}-\varrho\underline{\underline{\mathbf{M}}}^j\Delta\underline{\mathbf{l}}}(\underline{\mathbf{x}}')\, \phi^{j-1}_{\underline{\mathbf{k}}}(\underline{\mathbf{x}}')\, d\underline{\mathbf{x}}' \\
&\qquad\qquad (\underline{\mathbf{x}}' = \underline{\mathbf{x}} + \varrho\underline{\mathbf{l}}', \quad \Delta\underline{\mathbf{l}} = \underline{\mathbf{l}} - \underline{\mathbf{l}}') \\
&= \sum_{\underline{\mathbf{m}}\in\mathcal{V}_{j+2\lambda}} \widehat{f}^{j}_{\underline{\mathbf{m}}} \sum_{\Delta l\in\mathbb{Z}^2} h_0(\underline{\mathbf{m}} - \varrho\underline{\underline{\mathbf{M}}}^j\Delta\underline{\mathbf{l}} - \underline{\underline{\mathbf{M}}}\,\underline{\mathbf{k}}) \\
&\qquad\qquad \text{(by the proof of Theorem 2.7.3)} \\
&\equiv \sum_{\underline{\mathbf{m}}\in\mathcal{V}_{j+2\lambda}} \widehat{f}^{j}_{\underline{\mathbf{m}}}\, h_0^{\varrho,j}(\underline{\mathbf{m}} - \underline{\underline{\mathbf{M}}}\,\underline{\mathbf{k}}),
\end{aligned} \tag{2.162}
$$

for all $j = 0, 1, 2, ..., j_2 - 1$. Again, the convergence of $h_0^{\varrho,j}(\underline{\mathbf{m}} - \underline{\underline{\mathbf{M}}}\,\underline{\mathbf{k}}) = \sum_{\Delta l\in\mathbb{Z}^2} h_0(\underline{\mathbf{m}} - \varrho\underline{\underline{\mathbf{M}}}^j\Delta\underline{\mathbf{l}} - \underline{\underline{\mathbf{M}}}\,\underline{\mathbf{k}})$ is not an issue with the ϕ as constructed. Substitute these computed scaling transforms into (2.159) and (2.160) and the 2-D multiscale window transform and synthesis follow immediately.

2.7.7 Testing examples

Two functions

$$
f_1(x,y) = e^{-\frac{1}{4}(x^2+y^2)} + 2\, e^{-64(x^2+y^2)} \sin(20\pi x)\ \sin(20\pi y) \tag{2.163}
$$

$$
f_2(x,y) = (x+y+2) + 2\, e^{-64(x^2+y^2)} \sin(20\pi x)\ \sin(20\pi y) \tag{2.164}
$$

are chosen to test the capability of the multiscale window analysis just introduced. They are the 2-D version of the testing functions used in §2.6.3. In Fig. 2.13, their distributions are contoured (top row), over a region $\{(x,y) \mid |x| \le 1, \text{y} \le 1\}$. The function values are mapped on a 64×64 mesh grid (hence $j_2 = 12$), and a two-scale window analysis with $j_0 = 5$ is performed on the mapped data. In doing so, the definition domain is first transformed into $[0,1) \times [0,1)$, and then a periodic (symmetric) extension is adopted for f_1 (f_2). The testing result is very satisfactory. Displayed on the middle row of Fig. 2.13 are the large-scale reconstructions, which agree very well with the expected features (not shown here, as the difference between the reconstructed and idealized maps is almost indiscernible), except that for the f_2 case, there exists some wiggle close to the boundaries. This anomaly has been expected, as the extension scheme may cause some mismatch there. But even with this boundary effect, the error is limited within the tolerance (less than 5%). Our multiscale window transform-synthesis is therefore capable of handling highly localized events, with the precision needed satisfactorily met.

2.8 Summary

An analytical tool called multiscale window transform (MWT) has been developed, both on a 1-D field and on a 2-D field . With this MWT phenomena observed in GFD energetics, such as the energy intermittence, scale windowing, and self-similarity, can be faithfully represented, so long as the orthonormal shift invariant basis with which it is defined is properly localized, and the events of interest are not within some crucial distance immediately close to either of the two ends of the signal.

In the MWT development, we have particularly focused on the transform which can be readily applied to real problems. In the 1-D case, all functions are assumed to be (1) defined on a finite domain $[0, 1)$ (pre-scaled with the duration), and (2) belonging to a space containing scales larger than 2^{-j_2}, for some finite j_2. The first assumption is merely the reality. It motivates a study of the two no-energy leak schemes of extension: the periodic exention and symmetric extension. The second assumption is justifiable, with either of these extensions, and the justification is based on the fact that, in practice, signals are discretely defined on the finite interval [0,1).

Properties of the 1-D MWT have been examined. A notable such property is the marginalization over the domain $[0, 1)$. It identifies the sum of the phase space representation of a quadratic product to its duration average. This property also allows a simple representation of energy with proper units.

The 1-D multiscale window analysis can be conveniently realized. It is required that a scaling function ϕ be found which is (1) polynomially localized up to order $\gamma > 1$, (2) symmetric about the origin and maximized at zero, and (3) shift invariantly orthonormal. We have constructed such a ϕ and obtained its corresponding low-pass filter h_0 from a spline function. With this ϕ and h_0, scaling transforms at all the scale levels that are available can be efficiently computed. The computation involves first an application of a "pre-filter" at the highest level, and then a fast scaling transform for each level below. The 1-D multiscale transform and synthesis follow

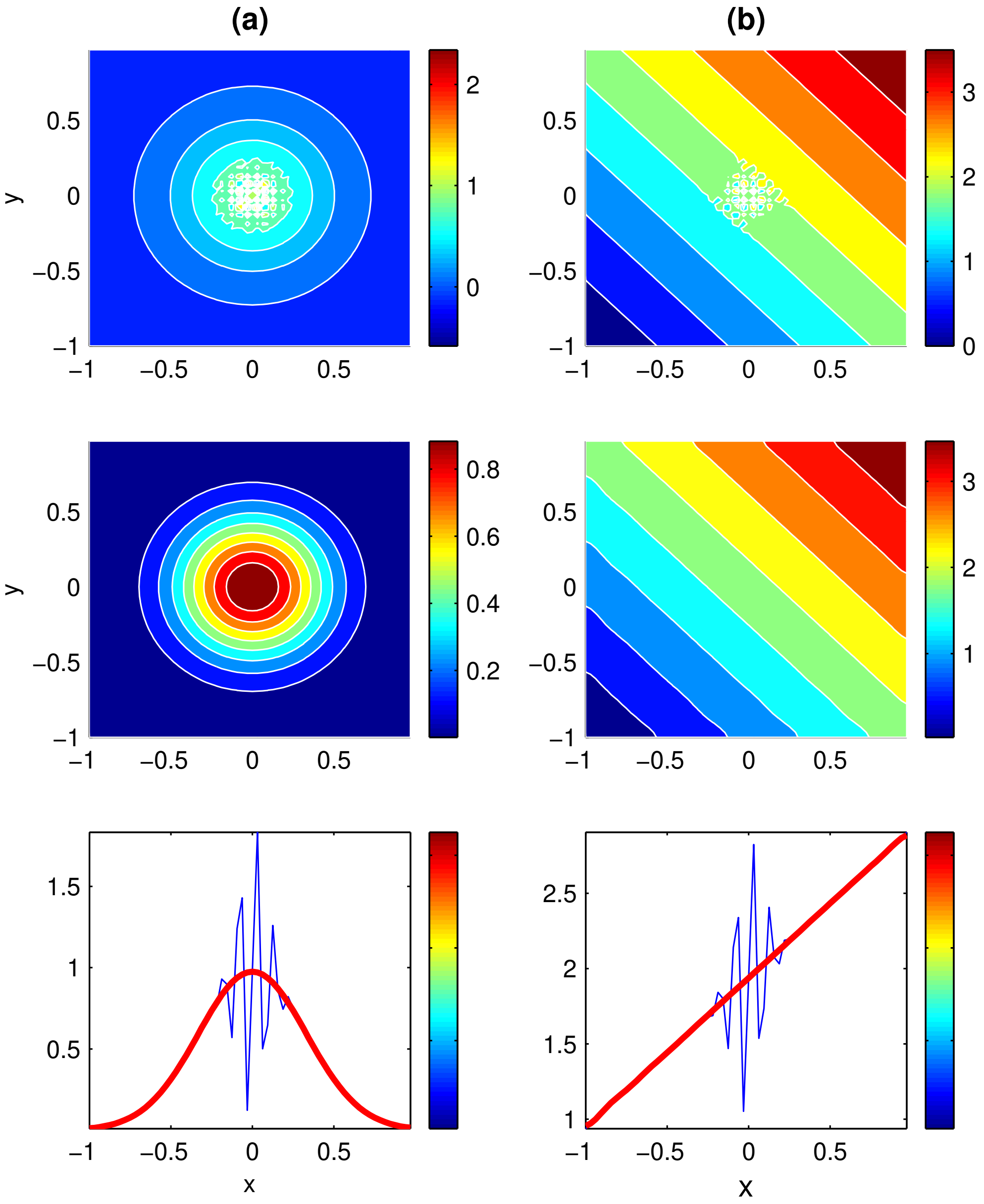

Figure 2.13: Examples testing the 2-D multiscale window transform/synthesis with functions $f_1(x, y)$ (top, left panel) and $f_2(x, y)$ (top, right panel), which are defined as shown in the text. Displayed on the middle row are the features reconstructed respectively with f_1 and f_2 on the large-scale window. The bottom-row plots show correspondingly the sectional distributions along $y = 0$, together with their noise parts that have been separated.

immediately by their respective definitions.

The 1-D results can be conveniently extended to a multi-dimensional case, especially a 2-D case. In §2.7, we have constructed a planar MWT by following a procedure similar to that for the 1-D transform. The tool construction is now completed, and we are ready to proceed to the next chapter, the MS-EVA formulation.

Appendix to Chapter 2

A2.1 Proof of Theorem 2.3.1

Proof

We have already shown that $\{\phi_n^{\varrho,j}\}_{n\in\mathcal{N}_\varrho^j}$ belongs to $L_2(\mathbb{T}_\varrho)$. We need only to show that

$$\int_0^\varrho \phi_n^{\varrho,j}(t)\ \phi_{n'}^{\varrho,j}(t)\ dt = \delta(n-n'), \qquad n,n' = 0,1,2,...,2^j\varrho - 1.$$

By the definition of periodized bases,

$$\begin{aligned}
&\int_0^\varrho \phi_n^{\varrho,j}(t)\ \phi_{n'}^{\varrho,j}(t)\ dt \\
&\quad = \sum_{\ell\in\mathbb{Z}}\sum_{\ell'\in\mathbb{Z}} \int_0^\varrho \phi_n^j(t+\varrho\ell)\ \phi_{n'}^j(t+\varrho\ell')\ dt \\
&\quad = \sum_{\Delta\ell\in\mathbb{Z}}\sum_{\ell'\in\mathbb{Z}} \int_{\varrho\ell'}^{\varrho(\ell'+1)} \phi_n^j(t'+\varrho\Delta\ell)\ \phi_{n'}^j(t')\ dt' \\
&\qquad\qquad\qquad (t' = t+\varrho\ell',\ \Delta\ell = \ell-\ell') \\
&\quad = \sum_{\Delta\ell\in\mathbb{Z}} \int_{\mathbb{R}} \phi_{n-2^j\varrho\Delta\ell}^j(t')\ \phi_{n'}^j(t')\ dt' \\
&\quad = \int_{\mathbb{R}} \phi_n^j(t')\ \phi_{n'}^j(t')\ dt' + \sum_{\Delta\ell\neq 0} \int_{\mathbb{R}} \phi_{n-2^j\varrho\Delta\ell}^j(t')\ \phi_{n'}^j(t')\ dt'.
\end{aligned}$$

In the derivation we have used term-by-term integration, which is made possible by the localization assumption for ϕ (polynomially localized up to an order $\gamma > 1$) we made in the beginning. Since $n, n' \in \mathcal{N}_\varrho^j$, an integer $\Delta\ell$ which is different from zero will not give any chance for $n - 2^j\varrho\Delta\ell$ to be equal to n'. By the orthonormality of $\{\phi_n^j\}_{n\in\mathbb{Z}}$, the second term of the above equation is thence zero, and the whole equation is simply $\delta(n-n')$. This finishes the proof of Theorem 2.3.1.

A2.2 Proof of Theorem 2.3.2

Proof

We have assumed that ϱ is a power of 2. Let it be 2^λ, and λ a positive integer. This yields the following two useful identities:

$$\phi_n^{1,j+\lambda}(\frac{t}{\varrho}) = \sqrt{\varrho}\phi_n^{\varrho,j}(t), \tag{A2.165}$$

$$\mathcal{N}_\varrho^j = \mathcal{N}_1^{j+\lambda}. \tag{A2.166}$$

For any $p \in V_{\varrho,j}$, it has a representation like

$$p(t) = \sum_{n\in\mathcal{N}_\varrho^j} \alpha_n \phi_n^{\varrho,j}(t), \tag{A2.167}$$

with α_n some expansion coefficients. Make a transformation

$$t' = t/\varrho, \qquad t = \varrho t.$$

By (A2.165) and (A2.166),

$$\begin{aligned} p(\varrho t') &= \sum_{n\in\mathcal{N}_\varrho^j} \alpha_n \phi_n^{\varrho,j}(\varrho t') \\ &= \sum_{n\in\mathcal{N}_1^{j+\lambda}} \frac{\alpha_n}{\sqrt{\varrho}} \phi_n^{1,j+\lambda}(t'). \end{aligned}$$

This is to say,

$$p_\varrho(t') \equiv p(\varrho t') \in V_{1,j+\lambda}.$$

As proved by Hernández and Weiss (1996),

$$V_{1,j+\lambda} \subset V_{1,j+1+\lambda}.$$

So

$$\begin{aligned} p(\varrho t') = p_\varrho(t') &= \sum_{n\in\mathcal{N}_1^{j+\lambda+1}} \beta_n \phi_n^{1,j+\lambda+1}(t') \\ &= \sum_{n\in\mathcal{N}_\varrho^{j+1}} \sqrt{\varrho}\beta_n \phi_n^{\varrho,j+1}(\varrho t'), \end{aligned}$$

where β_n are some expansion coefficients. Transformed back to t, this is

$$p(t) = \sum_{n \in \mathcal{N}_\varrho^{j+1}} \beta_n \phi_n^{\varrho,j+1}(t),$$

which means $p \in V_{\varrho,j+1}$. Since p is chosen arbitrarily from $V_{\varrho,j}$, we have

$$V_{\varrho,j} \subseteq V_{\varrho,j+1}.$$

But dim $V_{\varrho,j+1} >$ dim $V_{\varrho,j}$, so

$$V_{\varrho,j} \subset V_{\varrho,j+1}.$$

A2.3 Proof of Theorem 2.3.3

Proof

Because of the inclusion property we have just proved in Theorem A2.2, what we need to show here is that $\lim_{j\to\infty} V_{\varrho,j}$ is dense in $L_2(\mathbb{T}_\varrho)$. The proof of the 1-periodic case of this theorem has been done by Hernández and Weiss (1996). Moreover, they have shown that, for any function $g \in L_2(\mathbb{T}_1)$,

$$\lim_{j\to\infty} \|g - P_1^j g\|_{C_\infty} = 0, \tag{A2.168}$$

where P_1^j denotes the projection operator from $L_1(\mathbb{T}_1)$ onto $V_{1,j}$ ($C_\infty[a,b]$ is the normed space of continuous functions who have finite extrema over the definition domain). Now consider an arbitrary vector $f \in \mathrm{L}_2(\mathbb{T}_\varrho)$. We want to examine the performance of

$$\|f - P_\varrho^j f\|_{C_\infty}$$

when j is very large. Here $P_\varrho^j : L_2(\mathbb{T}_\varrho) \longrightarrow V_{\varrho,j}$ projects vectors in $L_2(\mathbb{T}_\varrho)$ onto $V_{\varrho,j}$, (Notice $V_{\varrho,j} \subset C_\infty(\mathbb{R})$ for $j \geq 0$) and ϱ is by assumption a power of 2. Make a transformation of variable:

$$t' = t/\varrho, \quad t = \varrho t',$$

and let

$$g(t') = f(\varrho t'),$$

then $g \in L_2(\mathbb{T}_1)$. By (A2.168), for any $\varepsilon > 0$, there always exists a $J = J(\varepsilon) > 0$, such that

$$\|g - P_1^{j+\lambda} g\|_{C_\infty} < \varepsilon \tag{A2.169}$$

as $j > J$. Here $\lambda = \log_2 \varrho$ is a positive integer, and

$$P_1^{j+\lambda} g(t') = \sum_{n\in\mathcal{N}_1^{j+\lambda}} \alpha_n \phi_n^{1,j+\lambda}(t')$$

with the Fourier coefficients being

$$\begin{aligned}
\alpha_n &= \left\langle g, \phi_n^{1,j+\lambda} \right\rangle \\
&= \int_0^1 g(t')\ \phi_n^{1,j+\lambda}(t')\ dt'
\end{aligned}$$

$$\begin{aligned}
&= \int_0^1 f(\varrho t') \cdot \sqrt{\varrho}\phi_n^{\varrho,j}(\varrho t')\ dt' && \text{(Eq. (A2.165) applied)} \\
&= \frac{1}{\sqrt{\varrho}} \int_0^\varrho f(t)\ \phi_n^{\varrho,j}(t)\ dt \\
&= \frac{1}{\sqrt{\varrho}} \left\langle f, \phi_n^{\varrho,j} \right\rangle \equiv \frac{1}{\sqrt{\varrho}} \beta_n,
\end{aligned}$$

where $\{\beta_n\}$ are the Fourier coefficients of f with respect to basis $\{\phi_n^{\varrho,j}\}$. So

$$\begin{aligned}
P_1^{j+\lambda} g(t') &= \sum_{n \in \mathcal{N}_1^{j+\lambda}} \frac{1}{\sqrt{\varrho}} \beta_n \cdot \sqrt{\varrho}\phi_n^{\varrho,j}(\varrho t') && \text{(Eq. (A2.165) applied)} \\
&= \sum_{n \in \mathcal{N}_\varrho^j} \beta_n \phi_n^{\varrho,j}(t) && \text{(Eq. (A2.166) applied)} \\
&= P_\varrho^j f(t).
\end{aligned}$$

But we also know $g(t') = f(\varrho t') = f(t)$, what Eq. (A2.169) actually states is thus

$$\|f - P_\varrho^j f\|_{C_\infty} < \varepsilon$$

as $j > J$. That is to say, given any $\varepsilon > 0$, when $j > J(\varepsilon)$, $V_{\varrho,j}$ is an ε-net of $L_2(\mathbb{T}_\varrho)$, and hence $\lim_{j \to \infty} V_{\varrho,j}$ is dense in $L_2(\mathbb{T}_\varrho)$.

A2.4 Proof of Theorem 2.3.6

Proof

First observe that

$$\int_0^2 p(t)\ q(t)\ dt = \sum_{n=0}^{2N-1} \alpha_n \beta_n \tag{A2.170}$$

holds for $p, q \in V_{2,j}$ by the generalized Parseval relation. Since both p and q are obtained by symmetric extension, we have

$$p(2-t) = p(t), \qquad q(2-t) = q(t). \tag{A2.171}$$

So the left hand side of (A2.170) is

$$\begin{aligned}
\int_0^2 p(t)\ q(t)\ dt &= \int_0^1 p(t)\ q(t)\ dt + \int_1^2 p(t)\ q(t)\ dt \\
&= \int_0^1 p(t)\ q(t)\ dt - \int_1^0 p(2-t')\ q(2-t')\ dt' \qquad (t' = 2-t) \\
&= \int_0^1 p(t)\ q(t)\ dt + \int_0^1 p(t')\ q(t')\ dt' \\
&= 2\int_0^1 p(t)\ q(t)\ dt.
\end{aligned} \tag{A2.172}$$

To perform a similar decomposition for the right hand side of (A2.170), notice

$$\begin{aligned}
\alpha_n &= \int_0^2 p(t)\ \phi_n^{2,j}(t)\ dt \\
&= \int_0^2 p(t) \sum_{\ell \in \mathbb{Z}} \phi^j_{n-2N\ell}(t)\ dt \\
&= \int_0^2 p(t) \sum_{\ell \in \mathbb{Z}} \sqrt{N}\phi(Nt - n + 2N\ell)\ dt \\
&= -\int_2^0 p(2-t') \sum_{\ell \in \mathbb{Z}} \sqrt{N}\phi(N(2-t') - n + 2N\ell)\ dt' \qquad (t' = 2-t) \\
&= \int_0^2 p(t')\ \sum_{\ell \in \mathbb{Z}} \sqrt{N}\phi(Nt' - 2N + n - 2N\ell)\ dt' \qquad (\text{p } \& \ \phi \text{ symm.}) \\
&= \int_0^2 p(t')\ \phi^{2,j}_{2N-n}(t')\ dt' \\
&= \alpha_{2N-n},
\end{aligned}$$

and this is true for all $n \in \mathcal{N}_2^j$. Likewise,

$$\beta_n = \beta_{2N-n}, \qquad \forall n = 0, 1, 2, ..., 2N-1.$$

So

$$\sum_{n=0}^{2N-1} \alpha_n \beta_n = 2\left(\sum_{n=1}^{N-1} \alpha_n \beta_n + \frac{1}{2}(\alpha_0 \beta_0 + \alpha_N \beta_N)\right). \tag{A2.173}$$

Eqs. (A2.172) and (A2.173) substituted back to (A2.170) yield

$$\int_0^1 p(t)\ q(t)\ dt = \sum_{n=1}^{N-1} \alpha_n \beta_n + \frac{1}{2}(\alpha_0 \beta_0 + \alpha_N \beta_N).$$

Q.E.D.

A2.5 Proof of Theorem 2.4.1

We begin the proof with a lemma.

Lemma A2.5.1

$$\sum_{m\geq m_0>0} \frac{1}{m^\gamma} \leq \frac{m_0^{1-\gamma}}{1-2^{1-\gamma}}, \qquad \forall \gamma > 1.$$

Proof

Let p_0 and p_1 be positive integers such that

$$2^{p_0-1} \leq m_0 < 2^{p_0}, \qquad 2^{p_1-1} \leq N_m < 2^{p_1}, \tag{A2.174}$$

then

$$\begin{aligned}
\sum_{m=m_0}^{N_m} \frac{1}{m^\gamma} &\leq \sum_{m=2^{p_0-1}}^{2^{p_1}-1} \frac{1}{m^\gamma} \\
&= \left[\frac{1}{(2^{p_0-1})^\gamma} + ... + \frac{1}{(2^{p_0}-1)^\gamma}\right] + ... + \left[\frac{1}{(2^{p_1-1})^\gamma} + ... + \frac{1}{(2^{p_1}-1)^\gamma}\right] \\
&\leq 2^{p_0-1} \cdot \frac{1}{2^{(p_0-1)\gamma}} + ... + 2^{p_1-1} \cdot \frac{1}{2^{(p_1-1)\gamma}} \\
&= \frac{1}{2^{(p_0-1)(\gamma-1)}} + ... + \frac{1}{2^{(p_1-1)(\gamma-1)}} + ... + \\
&= \frac{1}{1-2^{1-\gamma}} \cdot \left[2^{-(p_0-1)(\gamma-1)}\left(1-2^{-(p_1-p_0)(\gamma-1)}\right)\right].
\end{aligned}$$

This summation converges when $\gamma > 1$. Taking the limit as $N_m \to \infty$, we get

$$\sum_{m\geq m_0>0} \frac{1}{m^\gamma} \leq \frac{2^{-(p_0-1)(\gamma-1)}}{1-2^{1-\gamma}}, \tag{A2.175}$$

As $2^{p_0-1} \leq m_0$, this is exactly what we want.

Now let's get to the proof of Theorem 2.4.1. By assumption, the localized function $\phi(t)$ attains its maximum at $t = 0$. Let this maximum be C, then

$$\phi(0) = C.$$

We also know, for $n \in \mathbb{Z}$,

$$|\phi(n)| \leq \frac{C}{(1+n^2)^{\gamma/2}}, \qquad \gamma > 1. \tag{A2.176}$$

This "localization" can be used to estimate the bounds for the entries of matrix $\underline{\underline{\mathbf{H}}}$:

$$
\begin{aligned}
H_{nm} &= \sum_{l\in\mathbb{Z}} \phi^{j_2}_{m+lN}(t_n) \\
&= \sum_{l\in\mathbb{Z}} \sqrt{N}\phi\left(N\frac{n}{N} - m - lN\right) \\
&= \sum_{l\in\mathbb{Z}} \sqrt{N}\phi(n-m-lN), \qquad n,m = 0,1,...,N-1.
\end{aligned}
\tag{A2.177}
$$

Obviously, $\underline{\underline{\mathbf{H}}} = \{H_{nm}\}_{N\times N}$ forms a circulant matrix (e.g., Davis, 1979). On the diagonal,

$$
H_{nn} = \sqrt{N}\sum_{l\in\mathbb{Z}} \phi(-lN) = C\sqrt{N} + r_{nn}, \tag{A2.178}
$$

where

$$
\begin{aligned}
|r_{nn}| &= \sqrt{N} \sum_{l\in\mathbb{Z},|l|\neq 0} |\phi(lN)| \\
&\leq \sqrt{N}\cdot\sum_{|l|\neq 0} \frac{C}{(1+|lN|^2)^{\gamma/2}} \\
&< C\sqrt{N}\cdot\sum_{|l|>0} \frac{1}{|lN|^\gamma} \\
&= 2C\sqrt{N}\frac{1}{N^\gamma}\sum_{l=1}^{\infty}\frac{1}{l^\gamma} \\
&\leq \frac{2C}{1-2^{1-\gamma}}\sqrt{N}\frac{1}{N^\gamma} \qquad \text{by Lemma A2.5.1.}
\end{aligned}
\tag{A2.179}
$$

The off-diagonal elements, on the other hand, equal to

$$
H_{nm} = \sum_{l\in\mathbb{Z}} \sqrt{N}\phi(n-m-lN), \qquad \text{where } 0 < |n-m| < N. \tag{A2.180}
$$

We now examine $\sum_{m\neq n}|H_{nm}|$. As $\underline{\underline{\mathbf{H}}}$ is circulant, this sum is the same for all $0 \leq n < N$. It suffices to consider the case $n = 0$: $\sum_{m=1}^{N-1}|H_{0m}|$. For convenience, split the summation into two parts:

$$
\begin{aligned}
|H_{0m}| &\leq \sqrt{N}\sum_{l\in\mathbb{Z}} |\phi(m+lN)| \\
&\leq \sqrt{N}\sum_{l\in\mathbb{Z}} \frac{C}{[1+(m+lN)^2]^{\gamma/2}} \\
&= \underbrace{C\sqrt{N}\left\{\frac{1}{[1+m^2]^{\gamma/2}} + \frac{1}{[1+(m-N)^2]^{\gamma/2}}\right\}}_{(\mathrm{PRIN})_\mathrm{m}} \\
&\quad + \ (\mathrm{OTHER})_\mathrm{m},
\end{aligned}
\tag{A2.181}
$$

where

$$(\mathrm{OTHER})_{\mathrm{m}} = C\sqrt{N}\underbrace{\frac{1}{[1+(m+N)^2]^{\gamma/2}}}_{\text{(I)}} + \underbrace{C\sqrt{N}\sum_{|l|\geq 2}\frac{1}{[1+(m+lN)^2]^{\gamma/2}}}_{\text{(II)}}. \qquad \text{(A2.182)}$$

As $m > 0$, we have

$$\text{(I)} < \frac{1}{N^\gamma},$$

and

$$\begin{aligned}
\text{(II)} &= \sum_{l\geq 2}\frac{1}{[1+(m+lN)^2]^{\gamma/2}} + \sum_{l\geq 2}\frac{1}{[1+(m+lN)^2]^{\gamma/2}} \\
&< \sum_{l\geq 2}\frac{1}{(N-1-lN)^\gamma} + \sum_{l\geq 2}\frac{1}{[-(N-1)+lN]^\gamma} \\
&= 2\sum_{l\geq 2}\frac{1}{[(l-1)N+1]^\gamma} < 2\sum_{l'\geq 1}\frac{1}{(l'N)^\gamma} \\
&= \frac{2}{N^\gamma}\sum_{l'\geq 1}\frac{1}{l'^\gamma} \leq \frac{2}{1-2^{1-\gamma}}\cdot\frac{1}{N^\gamma} \qquad \text{for } \gamma > 1 \qquad \text{(A2.183)}
\end{aligned}$$

by Lemma A2.5.1. Hence, in (A2.181),

$$(\mathrm{OTHER})_{\mathrm{m}} < C\sqrt{N}\left[1 + \frac{2}{1-2^{1-\gamma}}\right]\frac{1}{N^\gamma}. \qquad \text{(A2.184)}$$

So

$$\begin{aligned}
\sum_{m=1}^{N-1}(\mathrm{OTHER})_{\mathrm{m}} &< C\sqrt{N}\left[1 + \frac{2}{1-2^{1-\gamma}}\right]\frac{1}{N^\gamma}\cdot(N-1) \\
&< C\sqrt{N}\left[1 + \frac{2}{1-2^{1-\gamma}}\right]\frac{1}{N^{\gamma-1}}. \qquad \text{(A2.185)}
\end{aligned}$$

The other part, $(\mathrm{PRIN})_{\mathrm{m}}$, sums to be

$$\begin{aligned}
\sum_{m=1}^{N-1}(\mathrm{PRIN})_{\mathrm{m}} &= \sum_{m=1}^{N-1} C\sqrt{N}\left\{\frac{1}{(1+m^2)^{\gamma/2}} + \frac{1}{[1+(m-N)^2]^{\gamma/2}}\right\} \\
&= 2C\sqrt{N}\sum_{m=1}^{N-1}\frac{1}{(1+m^2)^{\gamma/2}} \\
&= 2C\sqrt{N}\left[\frac{1}{2^{\gamma/2}} + \sum_{m=2}^{N-1}\frac{1}{m^\gamma}\right] \\
&< 2C\sqrt{N}\left[\frac{1}{2^{\gamma/2}} + \frac{2^{1-\gamma}}{1-2^{1-\gamma}}\right] \qquad \text{(A2.186)}
\end{aligned}$$

by Lemma A2.5.1. Putting $(\mathrm{PRIN})_{\mathrm{m}}$ and $(\mathrm{OTHER})_{\mathrm{m}}$ together, we have

$$\sum_{m=1}^{N-1}|H_{0m}| < C\sqrt{N}\left[2\left(\frac{1}{2^{\gamma/2}} + \frac{1}{2^{\gamma-1}(1-2^{1-\gamma})}\right) + \frac{1}{N^{\gamma-1}}\left(1 + \frac{2}{1-2^{1-\gamma}}\right)\right] \qquad \text{(A2.187)}$$

We are now able to make a comparison between $|H_{00}|$ and $\sum_{m=1}^{N-1} |H_{0m}|$.

$$\begin{aligned}
|H_{00}| - \sum_{m=1}^{N-1} |H_{0m}| &= \left| C\sqrt{N} + r_{00} \right| - \sum_{m=1}^{N-1} |H_{0m}| \\
&> C\sqrt{N} - |r_{00}| - \sum_{m=1}^{N-1} |H_{0m}|.
\end{aligned}$$

Substitution of (A2.179) and (A2.187) for $|r_{00}|$ and $\sum_{m=1}^{N-1} |H_{0m}|$ gives

$$\begin{aligned}
&|H_{00}| - \sum_{m=1}^{N-1} |H_{0m}| \\
&\quad > C\sqrt{N} \left\{ 1 - \frac{2}{1-2^{1-\gamma}} \frac{1}{N^\gamma} - \left(\frac{2}{2^{\gamma/2}} + \frac{2}{2^{\gamma-1}(1-2^{1-\gamma})} \right) - \frac{1}{N^{\gamma-1}} \left(1 + \frac{2}{1-2^{1-\gamma}} \right) \right\} \\
&\quad > C\sqrt{N} \left\{ 1 - \frac{2}{2^{\gamma/2}} - \frac{2}{2^{\gamma-1}(1-2^{1-\gamma})} - \left(1 + \frac{4}{1-2^{1-\gamma}} \right) \frac{1}{N^{\gamma-1}} \right\} \qquad \text{(A2.188)}
\end{aligned}$$

which is greater than zero given that (2.78) is satisfied. This is to say, $\underline{\underline{\mathbf{H}}}$ is strictly diagonally dominant under condition (2.78). Nonsingularity follows immediately from this diagonal dominance (e.g., Ortega, 1987).

Bibliography

[1] Anderson, Donald G. M., 1996: *Elementary Functional Analysis*, Applied Mathematics 210, Harvard University, Cambridge, MA 02138.

[2] Anderson, Donald G. M., 2001: *Partial Differential and Integral Equations*, Applied Mathematics 202, Harvard University, Cambridge, MA 02138.

[3] Bastiaans, M. J., 1980: Gabor's expansion of a signal into Gaussian elementary signals. *Proc. IEEE*, 68, 538-545.

[4] Cohen, A., and I. Daubechies, 1993: Nonseparable bidimensional wavelet bases. *Rev. Matematica Iberoamericanna*, vol. 9, No. 1, pp. 51-137.

[5] Daubechies, I., 1992: *Ten Lectures on Wavelets.* CBMS-NSF Regional Conf. Series in Appl. Math., Vol. 61. Society for Industrial and Applied Mathematics, Philadelphia, PA, 1992.

[6] Davis, Philip J., 1979: *Circulant Matrices.* John Wiley & Sons, New York. 250 pp.

[7] Gabor, D., 1946: Theory of communications. *J. Inst. Eletr. Engrg.*, London, 93 (III), 439-457.

[8] Hallatschek, K., M. Zilker, 1998: Real time data acquisition with transputes and Power PCs using the wavelet transform for event detection. *IEEE Transaction on Nuclear Science*, 45, p. 1872.

[9] Hernández, Eugenio, and Guido Weiss, 1996: *A First Course on Wavelets.* CRC Press. 489 pp.

[10] Holschneider, M., 1995: *Wavelets: An Analysis Tool.* Clarendon Press. Oxford. 423pp.

[11] Huang, Norden E., Z. Shen, and S. R. Long, 1999: A new view of nonlinear water waves: The Hilbert spectrum. *Annu. Rev. Fluid Mech.*, 31, 417-57.

[12] Istratescu, Vasile I., 1981: *Introduction to Linear Operator Theory.* Marcel Dekker, Inc.. 579pp.

[13] Kovacevic, J., and M. Vetterli, 1992: Nonseparable multidimensional perfect reconstruction filter banks and wavelet bases for $\mathcal{R}^n$. *IEEE Trans. Info. Theory.* Vol. 38, No. 2, 533-555.

[14] Kovacevic, J., and M. Vetterli, 1993: Nonseparable two- and three-dimensional wavelets, *IEEE Trans. on Signal Processing*, vol. 43, No. 5, 1269-1273.

[15] Kreyszig, Erwin, 1989: *Functional Analysis and its Applications.* John Wiley & Sons. 688 pp.

[16] Kumar, Praveen, and Efi Foufoula-Georgiou, 1997: Wavelet analysis for geophysical applications. *Rev. Geophys.*, 35, 4, 385-412.

[17] Louis, A. K., P. Maaβ, and A. Rieder, 1997: *Wavelets: Theory and Applications.* John Wiley & Sons. 324 pp.

[18] Mallat, S. G., 1989: Multiresolution approximations and wavelet orthonormal bases of $L^2(\mathbb{R})$. *Trans. Amer. Math. Soc.*, 315(1):69-87.

[19] Massopust, Peter R., 1994: *Fractal Functions, Fractal Surfaces, and Wavelets.* Academic Press. 383 pp.

[20] J. H. McClellan, 1973: The design of two-dimensional filters by transformations. *Proc. 7th Annual Princeton Conf. on Information Sciences and Systems*, Princeton (March),

[21] Meyer, Y., 1990: *Ondelettes et Opèrateurs I, II, and III*, Hermann, Paris.

[22] Meyer, Yves, 1992: *Wavelets and Operators.* Cambridge University Press. 223 pp.

[23] Nicolier, F., O. Laligant, and F. Truchetet, 1998: B-spline quincunx wavelet transform and implementation in Fourier domain. *Proceedings of the SPIE: Conference on Intelligent Robots and Computer Vision XVII.* Boston, Mass.. 223-234.

[24] Nicolier, F., O. Laligant, F. Truchetet, A.-C. Legrand, and S. Kohler, 1999: Human cells texture analysis with quincunx spline wavelet transform. *Proceedings of the SPIE: Conference on Human Vision and Electronic Imaging IV.* San Jose, Cal.. 606-615.

[25] Oden, J. Tinsley, and Leszek F. Demkowicz, 1996: *Applied Functional Analysis.* CRC Press. 653 pp.

[26] Ortega, James M., 1987: *Matrix Theory.* Plenum Press, NY, 262 pp.

[27] Press, W., H., S. A. Teukolsky, and W. T. Vetterling, 1993: *Numerical Recipes in C - The Art of Scientific Computing.* Cambridge University Press, 2nd eidtion. 965 pp.

[28] Shapiro, J. M., 1994: Adaptative McClellan transformations for quincunx filter banks. *IEEE Trans. on Signal Processing*, vol. 42, No. 3, 642-648.

[29] Strang, G., and T. Nguyen, 1997: *Wavelets and Filter Banks.* Wellesley-Cambridge Press. 520 pp.

[30] Sweldens, W., and P. Schröder, 1996: Building your own wavelets at Home. In: *Wavelets in Computer Graphics*, ACM SIGGRAPH Course Notes. (Available on http://cm.bell-labs.com/who/wim/)

[31] Truchetet, F., 1998: *Ondeletters pour le signal numérique.* coll. traitement du signal, Paris: Hermes, April 1998.

[32] van den Berg, J. C.: *Wavelets in Physics.* Cambridge University Press. 453 pp.

[33] Wojtaszczyk, P., 1997: *Mathematical Introduction to Wavelets.* Cambridge University Press. 261pp.

[34] Yosida, K., 1974: *Functional Analysis.* 4th ed. Springer, Berlin. 496 pp.

[35] Young, Robert M., 1980: *An Introduction to Nonharmonic Fourier Series.* Academic Press. 246pp.

Chapter 3

Formulation of the Multiscale Energy and Vorticity Analysis

With the multiscale window transform introduced in the preceding chapter, we proceed to formulate the MS-EVA, multiscale energy and vorticity analysis. The first section is about the primitive equation (PE) model from which the formulation starts. Concepts of kinetic energy and available potential energy are introduced and their evolution laws presented (same as those in a classical EVA setup). The large-, meso-, and sub-meso-scale energy equations are derived in respectively Sections 3.2, 3.3, and 3.4. Connections to their classical counterparts are also investigated (§3.5). The following two sections deal with the process separation, in order to substantiate the decomposition adopted in the energetics just developed (§3.6), and supplement it with interaction analyses to unravel the inter-scale information which would otherwise be deeply hidden in the "polluted" energetic maps (§3.7). In Sections 3.8 and 3.9, we present the vorticity equation and the multiscale enstrophy analysis. This chapter is closed with a brief summary and a discussion on how energy can be redistributed between locations in the physical space and between scale windows in the phase space, and what a role the mechanisms such as rotation, advection, buoyancy, and wave propagation may play in the redistribution.

In deriving the MS-EVA equations, all the time series are assumed to be defined on domain [0,1). They have to be extended to the whole real line before scale window transforms can be applied. Throughout the formulation, the symmetric extension is adopted except in the periodic case, where periodization yields the exact result. As shown in Chapter 2, either way energy will not leak outside the domain at any scale level.

Recall in Chapter 2 there is a constant factor 2^{j_2} for the multiscale energy definition [Eq. (2.73)], where j_2 is the maximal permissible scale level for a series with finite length. In this chapter, however, we will drop that factor in order to avoid cumbersome expressions. Therefore, **all the energetic terms hereafter, including the definitions of multiscale kinetic and available potential energies, unless otherwise indicated, should be multiplied by 2^{j_2} before physically interpreted**.

3.1 Three-dimensional primitive equation model and energy equations

As conventional, let P stand for pressure, ρ for density, and (x, y, z, t) for the space-time coordinates. The horizontal velocity is written either in a vector or a component form like

$$\underline{\mathbf{v}} = (u, v).$$

Note we use the vector form $\underline{\mathbf{v}}$ only for the *horizontal* components. The vertical velocity, w, is treated separately. This is because, for a geofluid, the vertical dimension usually has a size much smaller than its horizontal counterparts, and as a result, w can be neglected compared to u or v. The negligible w allows a definition of kinetic energy (KE) K with only u and v involved,

$$K = \frac{1}{2}(\underline{\mathbf{v}} \cdot \underline{\mathbf{v}}) = \frac{1}{2}(u^2 + v^2). \tag{3.1}$$

For the same reason, the horizontal and vertical gradients are also written separately. The gradient notation ∇ is used only for the horizontal components:

$$\nabla = \underline{\mathbf{i}}\frac{\partial}{\partial x} + \underline{\mathbf{j}}\frac{\partial}{\partial y}, \tag{3.2}$$

with the third component $\underline{\mathbf{k}}\ \partial/\partial z$ excluded. For a flow with velocity $\underline{\mathbf{v}}$, the quantity

$$\nabla \wedge \underline{\mathbf{v}} \cdot \underline{\mathbf{k}} \equiv \zeta \tag{3.3}$$

is referred to as a *vorticity* field, though in fact ζ measures only the rotation around the upward direction.

All the variables thus defined, either dependent or independent, are dimensional. With them the 3-D primitive governing equations are written as[1],

$$\frac{\partial \underline{\mathbf{v}}}{\partial t} = -\nabla K - w\frac{\partial \underline{\mathbf{v}}}{\partial z} - (f+\zeta)\underline{\mathbf{k}} \wedge \underline{\mathbf{v}} - \frac{1}{\rho_0}\nabla P + \underline{\mathbf{F}}_{mz} + \underline{\mathbf{F}}_{mh} \tag{3.4}$$

$$0 = \nabla \cdot \underline{\mathbf{v}} + \frac{\partial w}{\partial z} \tag{3.5}$$

$$0 = -\frac{\partial P}{\partial z} - \rho g \tag{3.6}$$

$$\frac{\partial \rho}{\partial t} = -\underline{\mathbf{v}} \cdot \nabla\rho - w\frac{\partial \rho}{\partial z} + F_{hz} + F_{hh}, \tag{3.7}$$

where the F-terms stand for subgrid process parameterizations, subscripts h and z signifying (x, y) and z directions, respectively. In the above equations, the flow is assumed to be incompressible. Also invoked are the hydrostatic and Bousinesq approximations.

A geofluid system is rather an inefficient heat engine. A huge part of its potential energy is not available for release or conversion. In order to define an energy of position, which will be called *available potential energy* henceforth, to better describe the system's capacity to do work, decompose the density field ρ into three parts,

$$\rho = \rho_0 + \bar{\rho}(z) + \rho'(x, y, z, t) \tag{3.8}$$

with $\rho' \ll \bar{\rho} \ll \rho_0$ the *density perturbation.* Here we do not scale $\bar{\rho}$ and ρ' as usual (e.g., Pinardi and Robinson, 1986), since we want to make all the variables dimensional, in accordance to the Harvard Ocean Prediction System. This decomposition allows the pressure to be integrated out, from the hydrostatic equation,

$$P = P_{atm} - \int_{\eta}^{z} \rho g dz$$

[1]The momentum equation can also be written in a form

$$\frac{\partial \underline{\mathbf{v}}}{\partial t} = -\nabla K - (f+\zeta)\underline{\mathbf{k}} \wedge \underline{\mathbf{v}} - \underline{\zeta}^* \wedge (w\underline{\mathbf{k}}) - \frac{1}{\rho_0}\nabla P + \underline{\mathbf{F}}_{mz} + \underline{\mathbf{F}}_{mh},$$

where

$$\underline{\zeta}^* = \zeta_1^* \underline{\mathbf{i}} + \zeta_2^* \underline{\mathbf{j}}$$

and

$$\zeta_1^* = -\frac{\partial v}{\partial z}, \quad \zeta_2^* = \frac{\partial u}{\partial z},$$

to facilitate the multiscale energetic equation derivation in later sections.

$$\approx \rho_0 g\eta - g\int_0^z \rho' dz + P_{atm} - \rho_0 gz - \int_0^z \rho(z) g dz, \tag{3.9}$$

where η is the sea surface elevation and P_{atm} the atmospheric pressure at the sea surface. Suppose P_{atm} is a constant. The only pressure part that accounts for horizontal motions is then the sum of the first two terms on the right hand side, or the **dynamic pressure**,

$$P' = \rho_0 g\eta - g\int_0^z \rho' dz. \tag{3.10}$$

The governing equations can be written with P' in place of P as the field variable. Likewise, ρ' may also be used to replace ρ in the mass continuity equation. In doing so, a large part of pressure and density that do not really affect in the motion is removed, and the resulting equation set naturally leads to the definition of *available potential energy* and will be seen soon. For convenience, Eqs. (3.4)-(3.7) are now recast as follows, with the primes of P' and ρ' dropped hereafter,

$$\begin{aligned}
\frac{\partial \underline{\mathbf{v}}}{\partial t} &= -\nabla K - w\frac{\partial \underline{\mathbf{v}}}{\partial z} - (f+\zeta)\underline{\mathbf{k}}\wedge\underline{\mathbf{v}} - \frac{1}{\rho_0}\nabla P + \underline{\mathbf{F}}_{mz} + \underline{\mathbf{F}}_{mh}, && (3.11)\\
0 &= \nabla\cdot\underline{\mathbf{v}} + \frac{\partial w}{\partial z}, && (3.12)\\
P &= \rho_0 g\eta - g\int_0^z \rho dz, && (3.13)\\
\frac{\partial \rho}{\partial t} &= -\underline{\mathbf{v}}\cdot\nabla\rho - w\frac{\partial\rho}{\partial z} + \frac{N^2\rho_0}{g}w + F_{\rho z} + F_{\rho h}. && (3.14)
\end{aligned}$$

In these equations, P and ρ are now understood as the dynamic pressure and density perturbation, and

$$N^2 = -\frac{g}{\rho_0}\frac{\partial\bar\rho}{\partial z}$$

is the **buoyancy frequency**, which varies with z only. For a stably stratified geofluid, N^2 is positive.

The kinetic energy equation is obtained by taking the dot product of $\underline{\mathbf{v}}$ with Eq. (3.11). With the incompressibility assumption and hydrostatic approximation [(3.12) and (3.13)], it reduces to

$$\frac{\partial K}{\partial t} = -\nabla\cdot(\underline{\mathbf{v}}K) - \frac{\partial(wK)}{\partial z} - \left[\nabla\cdot\left(\underline{\mathbf{v}}\frac{P}{\rho_0}\right) + \frac{\partial}{\partial z}\left(w\frac{P}{\rho_0}\right)\right] - b + \underline{\mathbf{v}}\cdot\underline{\mathbf{F}}_{mz} + \underline{\mathbf{v}}\cdot\underline{\mathbf{F}}_{mh}, \tag{3.15}$$

where

$$b = \frac{g}{\rho_0}w\rho \tag{3.16}$$

is the work rate by the buoyancy force. We will see soon that it plays an important role in mediating between the kinetic and potential energies.

The evolution of available potential energy (APE)

$$A = \frac{1}{2}\frac{g^2}{\rho_0^2 N^2}\rho^2 \tag{3.17}$$

is derived from equation of density perturbation. Multiplying Eq. (3.14) by $\rho g^2/(\rho_0^2 N^2)$, it is easy to get

$$\frac{\partial A}{\partial t} = -\nabla \cdot (\underline{\mathbf{v}}A) - \frac{\partial}{\partial z}(wA) + b - wA\frac{\partial}{\partial z}(\log N^2) + \frac{g^2}{\rho_0^2 N^2}\rho(F_{\rho,z} + F_{\rho,h}). \tag{3.18}$$

The fourth term on the right hand side has been called *apparent source or sink due to the stationary shear* (Pinardi and Robinson, 1986; Spall, 1989)[2]. Note that the buoyancy force work term bears a positive sign here, in opposite to that in (3.15). Summation of (3.15) and (3.18) cancels this term out. It is, therefore, merely the rate of conversion between the two energy forms (KE and APE). This kind of process, which converts energy from one form to another without causing net change in the total quantity, is called a **conversion process**. Here, if $b > 0$, K is converted to A; if $b < 0$, then A is extracted to feed the kinetic motion.

Equations (3.15) and (3.18) are the kinetic energy and potential energy evolutions for a geophysical fluid flow. A similar procedure will be followed to derive the equations for our multiscale energetics.

3.2 Large-scale energetics

3.2.1 Large-scale kinetic energy equation

The formulation of large-scale energetics generally follows from the previous derivation for the evolutions of K and A. The difference lies in that here we consider our problem in the phase space. Since the basis function $\phi^{\varrho,j}$, for any $0 \le j \le j_2$, is time dependent, and the derivative of $\phi^{\varrho,j}$ does not in general form an orthogonal pair with $\phi^{\varrho,j}$ itself, the local time change terms in the primitive equations need to be pre-treated specially before the energy equations can be formulated. Similar problems also exist in Harrison and Robinson (1978)'s

[2]It seems that the factor $\frac{1}{2}$ in Spall (1989)'s formulation in front of this term should be removed.

formalism. Appearing on the left hand side of their kinetic energy equation is $\bar{\underline{\mathbf{v}}} \cdot \frac{\partial \bar{\underline{\mathbf{v}}}}{\partial t}$, not in a form of time change of $K^L = \frac{1}{2}\bar{\underline{\mathbf{v}}} \cdot \bar{\underline{\mathbf{v}}}$.

To start, first consider $\frac{\partial \underline{\mathbf{v}}}{\partial t}$. Recall that our objective is to develop a diagnostic tool for an existing dataset. Thus every differential term has to be replaced eventually by its difference counterpart. That is to say, we actually do not need to deal with $\partial \underline{\mathbf{v}}/\partial t$ itself. Rather, it is the discretized form (space-dependence suppressed for clarity)

$$\frac{\underline{\mathbf{v}}(t+\Delta t) - \underline{\mathbf{v}}(t-\Delta t)}{2\Delta t} \equiv \delta_t \underline{\mathbf{v}}$$

that we should pay attention to (Δt is the time step size). Viewed as functions of t, $\underline{\mathbf{v}}(t+\Delta t)$ and $\underline{\mathbf{v}}(t-\Delta t)$ make two different series and may be transformed separately. Let

$$\int_0^\varrho \underline{\mathbf{v}}^{\sim 0}(t+\Delta t)\ \phi_n^{\varrho, j_2}(t)\ dt \quad \equiv \quad \widehat{\underline{\mathbf{v}}}_{n+}^{\sim 0}, \tag{3.19}$$

$$\int_0^\varrho \underline{\mathbf{v}}^{\sim 0}(t-\Delta t)\ \phi_n^{\varrho, j_2}(t)\ dt \quad \equiv \quad \widehat{\underline{\mathbf{v}}}_{n-}^{\sim 0}, \tag{3.20}$$

where ϱ is the periodicity of extension ($\varrho = 1$ and 2 for extensions by periodization and refection, respectively), and define an operator $\hat{\delta}_n$ such that

$$\hat{\delta}_n \widehat{\underline{\mathbf{v}}}_n^{\sim 0} = \frac{\widehat{\underline{\mathbf{v}}}_{n+}^{\sim 0} - \widehat{\underline{\mathbf{v}}}_{n-}^{\sim 0}}{2\Delta t}. \tag{3.21}$$

$\hat{\delta}_n \widehat{\underline{\mathbf{v}}}_n^{\sim 0}$ is actually the large-scale transform of $\delta_t \underline{\mathbf{v}}$, or the rate of change of $\widehat{\underline{\mathbf{v}}}_n^{\sim 0}$ on its corresponding scale window. Similarly, define difference operators of the second order as follows:

$$\delta_{t^2}^2 \underline{\mathbf{v}} \quad \equiv \quad \frac{\underline{\mathbf{v}}(t+\Delta t) - 2\underline{\mathbf{v}}(t) + \underline{\mathbf{v}}(t-\Delta t)}{(\Delta t)^2}, \tag{3.22}$$

$$\hat{\delta}_{n^2}^2 \widehat{\underline{\mathbf{v}}}_n^{\sim 0} \quad \equiv \quad \int_0^\varrho \delta_{t^2}^2 \underline{\mathbf{v}}^{\sim 0} \quad \phi_n^{\varrho, j_2}(t) dt. \tag{3.23}$$

Now take the dot product of $\widehat{\underline{\mathbf{v}}}_n^{\sim 0}$ with $\hat{\delta}_n \widehat{\underline{\mathbf{v}}}_n^{\sim 0}$,

$$\begin{aligned}
\widehat{\underline{\mathbf{v}}}_n^{\sim 0} \cdot \hat{\delta}_n \widehat{\underline{\mathbf{v}}}_n^{\sim 0} &= \left(-\frac{\widehat{\underline{\mathbf{v}}}_{n+}^{\sim 0} - 2\widehat{\underline{\mathbf{v}}}_n^{\sim 0} + \widehat{\underline{\mathbf{v}}}_{n-}^{\sim 0}}{2} + \frac{\widehat{\underline{\mathbf{v}}}_{n+}^{\sim 0} + \widehat{\underline{\mathbf{v}}}_{n-}^{\sim 0}}{2} \right) \cdot \frac{\widehat{\underline{\mathbf{v}}}_{n+}^{\sim 0} - \widehat{\underline{\mathbf{v}}}_{n-}^{\sim 0}}{2\Delta t} \\
&= \frac{1}{2\Delta t} \left(\frac{1}{2} \widehat{\underline{\mathbf{v}}}_{n+}^{\sim 0} \cdot \widehat{\underline{\mathbf{v}}}_{n+}^{\sim 0} - \frac{1}{2} \widehat{\underline{\mathbf{v}}}_{n-}^{\sim 0} \cdot \widehat{\underline{\mathbf{v}}}_{n-}^{\sim 0} \right) - (\Delta t)^2 \left(\hat{\delta}_{n^2}^2 \widehat{\underline{\mathbf{v}}}_n^{\sim 0} \cdot \hat{\delta}_n \widehat{\underline{\mathbf{v}}}_n^{0} \right) \\
&= \hat{\delta}_n K_n^L - (\Delta t)^2 \left(\hat{\delta}_{n^2}^2 \widehat{\underline{\mathbf{v}}}_n^{\sim 0} \cdot \hat{\delta}_n \widehat{\underline{\mathbf{v}}}_n^{\sim 0} \right),
\end{aligned} \tag{3.24}$$

where

$$K_n^L = \frac{1}{2} \widehat{\underline{\mathbf{v}}}_n^{\sim 0} \cdot \widehat{\underline{\mathbf{v}}}_n^{\sim 0} \tag{3.25}$$

is the kinetic energy at location n (in the phase space) for the large-scale window (the factor 2^{j_2} omitted). Note that K_n^L is different from $\widehat{K}_n^{\sim 0}$. The latter is the large-scale window transform of K, not a concept of "energy". Another quantity that might be confused with K_n^L is $K^{\sim 0}$, or the large-scale part of the field variable K. $K^{\sim 0}$ is a property in the physical space. It is conceptually different from the phase-space defined K_n^L for flow $\underline{\mathbf{v}}$.

Observe that the first term on the right hand side of Eq. (3.24) is the time change (in difference form) of the large-scale kinetic energy at time $2^{-j_2}n$ (scaled by the series length). The second term, which is proportional to $(\Delta t)^2$, is in general very small (of order $O[(\Delta t)^2]$ compared to $\hat{\delta}_n K_n^L$). As shown in the appendix, it could be significant only when processes with scales of grid size are concerned. Besides, it is expressed in a form of discretized Laplacian. We may thereby view it indistinguishably as a kind of subgrid parameterization and merge it into the dissipation terms. The term $\widehat{\underline{\mathbf{v}}}_n^{\sim 0} \cdot \hat{\delta}_n \widehat{\underline{\mathbf{v}}}_n^{\sim 0}$, which is akin to Harrison and Robinson's $\bar{\underline{\mathbf{v}}} \cdot \frac{\partial \bar{\underline{\mathbf{v}}}}{\partial t}$, is thus merely the change rate of the large-scale kinetic energy, with a small correction of order $(\Delta t)^2$ (t scaled by the series duration).

Above is how the time change term $\partial_t \underline{\mathbf{v}}$ is treated for the momentum equation. Similar treatment can also be applied to the density equation and we will see it later in this section.

Terms other than $\partial_t \underline{\mathbf{v}}$ and $\partial_t \rho$ in a 3-D primitive equation system do not have time derivatives involved. Multiscale window transforms can be applied directly to every field variable in spite of the spatial gradient operators, if any. To continue the derivation, first take a large-scale transform of (3.12),

$$\frac{\partial \widehat{w}_n^{\sim 0}}{\partial z} + \nabla \cdot \widehat{\underline{\mathbf{v}}}_n^{\sim 0} = 0. \tag{3.26}$$

This form-conserved continuity equation will be useful later. It relates the horizontal velocity to its vertical component.

Dot product of the momentum equation (3.11) with $\widehat{\underline{\mathbf{v}}}_n^{\sim 0} \phi_n^{\varrho, j_2}(t)$, followed by an integration with respect to t over the domain $[0, 1)$, gives the large-scale kinetic energy equation. The left hand side of this resulting equation is the time change term, $\widehat{\underline{\mathbf{v}}}_n^{\sim 0} \cdot \frac{\partial \underline{\mathbf{v}}}{\partial t}$, which we have already taken care of. Our task in the next step is to manipulate the right hand side into a summation of physically meaningful terms.

Look at the pressure work first. By Eq. (3.26), it is

$$\begin{aligned}
\int_0^{\varrho} - \hat{\underline{\mathbf{v}}}_n^{\sim 0} \cdot \frac{\nabla P^{\sim 0}}{\rho_0} \phi_n^{\varrho, j_2}(t)\, dt &= -\hat{\underline{\mathbf{v}}}_n^{\sim 0} \cdot \frac{\nabla \hat{P}_n^{\sim 0}}{\rho_0} \\
&= -\frac{1}{\rho_0}\left[\nabla \cdot (\hat{P}_n^{\sim 0} \hat{\underline{\mathbf{v}}}_n^{\sim 0}) + \frac{\partial}{\partial z}(\hat{P}_n^{\sim 0} \hat{w}_n^{\sim 0})\right] + \hat{w}_n^{\sim 0} \frac{\partial \hat{P}_n^{\sim 0}}{\partial z} \\
&= -\frac{1}{\rho_0}\left[\nabla \cdot (\hat{P}_n^{\sim 0} \hat{\underline{\mathbf{v}}}_n^{\sim 0}) + \frac{\partial}{\partial z}(\hat{P}_n^{\sim 0} \hat{w}_n^{\sim 0})\right] - \frac{g}{\rho_0} \hat{w}_n^{\sim 0} \hat{\rho}_n^{\sim 0} \\
&\equiv \Delta_h Q_{P^L} + \Delta_z Q_{P^L} - b_n^L,
\end{aligned} \tag{3.27}$$

where $\Delta_h Q_{P^L}$ and $\Delta_z Q_{P^L}$ (Q_P the pressure flux) are respectively the horizontal and vertical pressure working rates (Q stands for flux, a convention in many fluid mechanics textbooks). The third term, $-b_n^L = -\frac{g}{\rho_0}\hat{w}_n^{\sim 0}\hat{\rho}_n^{\sim 0}$, is the rate of buoyancy conversion between the large-scale kinetic and available potential energies.

For the term pre-multiplied by $f + \zeta$, the total vorticity, the integration reduces to:

$$\begin{aligned}
T_{K^L,\zeta} &= \int_0^{\varrho} -\hat{\underline{\mathbf{v}}}_n^{\sim 0} \phi_n^{\varrho, j_2}(t) \cdot [(f+\zeta)\underline{\mathbf{k}} \wedge \underline{\mathbf{v}}]^{\sim 0}\, dt \\
&= -f \hat{\underline{\mathbf{v}}}_n^{\sim 0} \cdot \underline{\mathbf{k}} \wedge \hat{\underline{\mathbf{v}}}_n^{\sim 0} - \hat{\underline{\mathbf{v}}}_n^{\sim 0} \cdot \int_0^{\varrho} [\zeta \underline{\mathbf{k}} \wedge \underline{\mathbf{v}}]^{\sim 0} \phi_n^{\varrho, j_2}(t)\, dt \\
&= -\hat{\underline{\mathbf{v}}}_n^{\sim 0} \cdot \int_0^{\varrho} [\zeta \underline{\mathbf{k}} \wedge \underline{\mathbf{v}}]^{\sim 0} \phi_n^{\varrho, j_2}(t)\, dt \\
&= -\hat{\underline{\mathbf{v}}}_n^{\sim 0} \cdot \underline{\mathbf{k}} \wedge \widehat{(\zeta \underline{\mathbf{v}})}_n^{\sim 0}.
\end{aligned} \tag{3.28}$$

It is of interest to observe that, $T_{K^L,\zeta}$ generally does not vanish, in contrast to that of Eq. (3.15), the classical EVA formulation. This term represents the contribution to the large-scale window through a kind of process called a **perfect transfer process** in phase space which we will define rigorously in §3.6.1. For a given flow, perfect transfer processes don't invoke net energy gain or net energy loss. What they do, in plain language, is to redistribute energy over the scale spectrum of that flow. In some sense, a perfect transfer process plays a role in energetics akin to what a conversion process does, and indeed, these two names have been used exchangably in the literature. We distinguish them by limiting *conversion* to processes between the two different forms of energy, while reserving the term *transfer process* for energy exchange between scale windows.[3]

Now we know rotation (f excluded) contributes to energy transfer, though it does no work to the whole system. Denote this rotation contribution as $T_{K^L,\zeta}$, which stands for "Transfer

[3]In this thesis, the term *transfer process* without the attribute "perfect" is used in a rather generic sense, referring to a general process of energy transfer in the phase space (energy could be gained or lost).

rate to large-scale kinetic energy K^L due to ζ".[4] Here ζ appears in the subscript to signify the mechanism. Since transfer rate is a natural measure of energy transfer, we will use the term "transfer" and "transfer rate" identically if no confusion exists in the context.

Among the remaining terms are the horizontal and vertical advections. After being dotted with $\hat{\underline{\mathbf{v}}}_n^{\sim 0}$, they total up to

$$\begin{aligned}
&\hat{\underline{\mathbf{v}}}_n^{\sim 0} \cdot \int_0^{\varrho} \left[(-\nabla K - w\frac{\underline{\mathbf{v}}}{\partial z}) \right]^{\sim 0} \phi_n^{\varrho, j_2}(t)\ dt \\
&= -\int_0^{\varrho} \left[\nabla \cdot (\hat{\underline{\mathbf{v}}}_n^{\sim 0} K^{\sim 0}) + K^{\sim 0} \frac{\partial \hat{w}_n^{\sim 0}}{\partial z} + \hat{\underline{\mathbf{v}}}_n^{\sim 0} \cdot \left[w \frac{\partial \underline{\mathbf{v}}}{\partial z} \right]^{\sim 0} \right] \phi_n^{\varrho, j_2}(t)\ dt \\
&= -\nabla \cdot (\hat{\underline{\mathbf{v}}}_n^{\sim 0} \widehat{K}_n^{\sim 0}) - \frac{\partial}{\partial z} (\hat{w}_n^{\sim 0} \widehat{K}_n^{\sim 0}) + \hat{w}_n^{\sim 0} \frac{\partial \widehat{K}_n^{\sim 0}}{\partial z} - \hat{\underline{\mathbf{v}}}_n^{\sim 0} \cdot \left(\widehat{w \frac{\partial \underline{\mathbf{v}}}{\partial z}} \right)_n^{\sim 0} \\
&\equiv \Delta_h Q_{K^L} + \Delta_z Q_{K^L} + T_{K^L, w\partial_z} + T_{K^L, \partial_z v}.
\end{aligned} \tag{3.29}$$

The Q-terms, $\Delta_h Q_{K^L}$ and $\Delta_z Q_{K^L}$, represent respectively the horizontal and vertical energy transports from surrounding spatial locations. The residual parts, $T_{K^L, \partial_z v}$ and $T_{K^L, w\partial_z}$, would sum up to zero if no transform were performed, as in the classical EVA formulation, Eq. (3.15). Here their sum in general does not vanish. Rather, they form another important type of energy transfer, the transfer due to vortex tube tilting. To see this, write up the part in its integral form. It is easy to show that[5]

$$\begin{aligned}
T_{K^L, \partial_z v} &= \int_0^{\varrho} \underline{\mathbf{k}} \cdot \left[w \underline{\zeta}^* \right]^{\sim 0} \wedge \hat{\underline{\mathbf{v}}}_n^{\sim 0}\ \phi_n^{\varrho, j_2}(t)\ dt \\
&= \hat{\underline{\mathbf{v}}}_n^{\sim 0} \cdot \underline{\mathbf{k}} \wedge (\widehat{\underline{\zeta}^* w})_n^{\sim 0},
\end{aligned} \tag{3.30}$$

where

$$\underline{\zeta}^* = \zeta_1^* \underline{\mathbf{i}} + \zeta_2^* \underline{\mathbf{j}},$$

and

$$\begin{aligned}
\zeta_1^* &= -\frac{\partial v}{\partial z}, \\
\zeta_2^* &= \frac{\partial u}{\partial z}.
\end{aligned}$$

Vector $\underline{\zeta}^*$ is actually the horizontal component of the real vorticity field, curl(u, v, w). Physically (and geometrically) it may be interpreted as the horizontal tilting of the vertical vortex

[4] A deeper look reveals that energy transfer is actually due to the evolution of ζ, rather than ζ itself.

[5] The following result can also be obtained (in fact, more easily) in a broader context where w is treated equally with $\underline{\mathbf{v}}$.

tubes. Since we have reserved the name *vorticity* and symbol ζ for the vertical component of curl(u, v, w), it is not tagged with that name. Instead, it is referred to as the vertical structure of horizontal flow $\underline{\mathbf{v}}$, and is denoted as $\underline{\zeta}^*$. By the thermal wind relation, this vertical structure is closely related to its horizontal density gradient. Therefore, an evolving density horizontal structure (a density front, for example) is also a cause of the *in situ* kinetic energy transfer.

The other transfer term, $T_{K^L,w\partial_z}$, may also be written in terms of $\underline{\zeta}^*$:

$$\begin{aligned} T_{K^L,w\partial_z} &= -\int_0^\varrho \underline{\mathbf{k}} \cdot \wedge(\widehat{w}_n^{\sim 0} \underline{\mathbf{v}}^{\sim 0})\ \phi_n^{\varrho, j_2}(t)\ dt \\ &= -\widehat{w}_n^{\sim 0} \underline{\mathbf{k}} \cdot (\widehat{\underline{\zeta}^* \wedge \underline{\mathbf{v}}})_n^{\sim 0}. \end{aligned} \tag{3.31}$$

In a three-dimensional fluid flow with z in a position identical to x and y, a form like (3.31) is a convenience for comparison's sake. In a geophysical fluid flow, however, it just complicates the physical interpretation. In the derivation that follows, we will stick to its original form:

$$T_{K^L,w\partial_z} = \widehat{w}_n^{\sim 0} \frac{\partial \widehat{K}_n^{\sim 0}}{\partial z}. \tag{3.32}$$

It should be pointed out, in general, that either $T_{K^L,w\partial_z}$ or $T_{K^L,\partial_z v}$ alone does not represent a perfect transfer process defined before. It is the sum of these two terms that can be put in this process category, as will be clear in §3.6.2. We write them separately because we need to consider them later individually when the two well-known processes, barotropic and baroclinic instabilities, are examined (see Chapter 4).

In Eq. (3.15), the friction terms, $\underline{\mathbf{F}}_{mz}$ and $\underline{\mathbf{F}}_{mh}$, stand for the effect of unresolved sub-grid processes. An explicit expression of them is problem-specific. They depend on the sub-grid parameterization that is adopted. For the commonly used Laplacian formulation, they are

$$\begin{aligned} F_{mz} &= \frac{\partial}{\partial z}\left(A_v \frac{\partial \underline{\mathbf{v}}}{\partial z}\right), \\ F_{mh} &= \nabla \cdot (A_h \nabla \cdot \underline{\mathbf{v}}), \end{aligned}$$

where A_h and A_v are the horizontal and vertical eddy viscosities, respectively. A large-scale window transform followed by an application of dot product with $\widehat{\underline{\mathbf{v}}}_n^{\sim 0}$ gives

$$\widehat{\underline{\mathbf{v}}}_n^{\sim 0} \cdot (\widehat{\underline{\mathbf{F}}_{mz}})_n^{\sim 0} = \widehat{\underline{\mathbf{v}}}_n^{\sim 0} \cdot \frac{\partial}{\partial z}\left(A_v \frac{\partial \widehat{\underline{\mathbf{v}}}_n^{\sim 0}}{\partial z}\right) = \frac{\partial}{\partial z}\left(A_v \frac{\partial K^L}{\partial z}\right) - A_v \left(\frac{\partial \widehat{\underline{\mathbf{v}}}_n^{\sim 0}}{\partial z}\right) \cdot \left(\frac{\partial \widehat{\underline{\mathbf{v}}}_n^{\sim 0}}{\partial z}\right) \tag{3.33}$$

$$\widehat{\underline{\mathbf{v}}}_n^{\sim 0} \cdot (\widehat{\underline{\mathbf{F}}_{mh}})_n^{\sim 0} = \widehat{\underline{\mathbf{v}}}_n^{\sim 0} \cdot \nabla \cdot (A_h \nabla \widehat{\underline{\mathbf{v}}}_n^{\sim 0}) = \nabla \cdot \left(A_h \nabla K_n^L\right) - A_h \nabla \widehat{\underline{\mathbf{v}}}_n^{\sim 0} : \left(\nabla \widehat{\underline{\mathbf{v}}}_n^{\sim 0}\right)^T. \tag{3.34}$$

These are what we can get from a Laplacian type dissipation. But we do not really use them in our MS-EVA. We will simply write these two terms as $F_{K^L,z}$ and $F_{K^L,h}$ (location index n suppressed for clarity). For one reason, in HOPS the horizontal sub-grid effect is parameterized with a Shapiro filter, so the above formula does not quite apply; for another reason, one may want to include into $F_{K^L,h}$ the afore-derived correction to $\hat{\delta}_n K_n^L$, as it behaves like a kind of horizontal dissipation. In short, different users of MS-EVA may modify these friction terms differently to meet their specific needs. It is better not to make the formula too particular here.

In a summary, the kinetic energy evolution on the large-scale window is governed by

$$\begin{aligned}
\hat{\delta}_n K_n^L &= -\nabla\cdot(\underline{\hat{\mathbf{v}}}_n^{\sim 0}\widehat{K}_n^{\sim 0}) - \frac{\partial}{\partial z}(\widehat{w}_n^{\sim 0}\widehat{K}_n^{\sim 0}) \\
&\quad -\underline{\hat{\mathbf{v}}}_n^{\sim 0}\cdot\underline{\mathbf{k}}\wedge\widehat{(\zeta\underline{\mathbf{v}})}_n^{\sim 0} + \widehat{w}_n^{\sim 0}\frac{\partial\widehat{K}_n^{\sim 0}}{\partial z} + \underline{\hat{\mathbf{v}}}_n^{\sim 0}\cdot\underline{\mathbf{k}}\wedge\widehat{(\underline{\zeta}^* w)}_n^{\sim 0} \\
&\quad -\nabla\cdot\left(\underline{\hat{\mathbf{v}}}_n^{\sim 0}\frac{\widehat{P}_n^{\sim 0}}{\rho_0}\right) - \frac{\partial}{\partial z}\left(\widehat{w}_n^{\sim 0}\frac{\widehat{P}_n^{\sim 0}}{\rho_0}\right) - \frac{g}{\rho_0}\widehat{w}_n^{\sim 0}\widehat{\rho}_n^{\sim 0} \\
&\quad +F_{K^L,z} + F_{K^L,h},
\end{aligned} \tag{3.35}$$

or, symbolically,

$$\begin{aligned}
\dot{K}_n^L &= \Delta_h Q_{K^L} + \Delta_z Q_{K^L} + T_{K^L,\zeta} + T_{K^L,w\partial_z} + T_{K^L,\partial_z v} \\
&\quad +\Delta_h Q_{P^L} + \Delta_z Q_{P^L} - b_n^L + +F_{K^L,z} + F_{K^L,h}.
\end{aligned} \tag{3.36}$$

3.2.2 Large-scale available potential energy equation

To arrive at the large-scale available potential energy equation, introduce for the time being a shorthand $c \equiv g^2/(\rho_0^2 N^2)$ to simplify the otherwise cumbersome derivation (Note that c is z-dependent). Take the large-scale window transform of the time-discretized version of Eq. (3.14) and multiply it by $c\widehat{\rho}_n^{\sim 0}$. The left hand side becomes, as before,

$$\begin{aligned}
c\widehat{\rho}_n^{\sim 0}\widehat{(\delta_t\rho)}_n^{\sim 0} &= c\widehat{\rho}_n^{\sim 0}\hat{\delta}_n\widehat{\rho}_n^{\sim 0} \\
&= \hat{\delta}_n A_n^L - (\Delta t)^2 c\left(\hat{\delta}_{n^2}^2\widehat{\rho}_n^{\sim 0}\cdot\hat{\delta}_n\widehat{\rho}_n^{\sim 0}\right),
\end{aligned}$$

where

$$A_n^L = \frac{1}{2}c\left(\widehat{\rho}_n^{\sim 0}\right)^2 = \frac{1}{2}\frac{g^2}{\rho_0^2 N^2}\left(\widehat{\rho}_n^{\sim 0}\right)^2 \tag{3.37}$$

(constant multiplier 2^{j_2} omitted) is the available potential energy at location n in the phase space (corresponding to the scaled time $2^{j_2}n$) for the large-scale window. Compared to $\hat{\delta}_n A_n^L$, the correction is of order $(\Delta t)^2$, and could be significant only at small scales, as argued for the kinetic energy case.

For the advection-related terms, the transform followed by a multiplication with $c\widehat{\rho}_n^{\sim 0}$ yields

$$(\mathrm{AD}) = c\hat{\rho}_n^0 \int_0^{\varrho} \left(-\nabla \cdot (\underline{\mathbf{v}}\rho)^{\sim 0} - \frac{\partial (w\rho)^{\sim 0}}{\partial z} \right) \phi_n^{\varrho, j_2}(t)\; dt = -c\widehat{\rho}_n^{\sim 0} \nabla \cdot \widehat{(\underline{\mathbf{v}}\rho)}_n^{\sim 0} - c\widehat{\rho}_n^{\sim 0} \frac{\partial}{\partial z} \widehat{(w\rho)}_n^{\sim 0}.$$

As will be explained later, we need to collect flux-like terms. In the phase space, these terms are:

$$\Delta_h Q_{A^L} \equiv -\nabla \cdot \left[\frac{1}{4} c\widehat{\rho}_n^{\sim 0} \widehat{(\underline{\mathbf{v}}\rho)}_n^{\sim 0} + \frac{1}{4} c\widehat{(\rho^2)}_n^{\sim 0} \widehat{\underline{\mathbf{v}}}_n^{\sim 0} \right] \tag{3.38}$$

$$\Delta_z Q_{A^L} \equiv -\frac{\partial}{\partial z} \left[\frac{1}{4} c\widehat{\rho}_n^{\sim 0} \widehat{(w\rho)}_n^{\sim 0} + \frac{1}{4} c\widehat{(\rho^2)}_n^{\sim 0} \widehat{w}_n^{\sim 0} \right]. \tag{3.39}$$

In §3.6.1, we shall explain why they should be written in such a form. With this flux representation, (AD) is decomposed as

$$\begin{aligned} (\mathrm{AD}) \quad = \quad & \Delta_h Q_{A^L} + \Delta_z Q_{A^L} \\ & - \left[c\widehat{\rho}_n^{\sim 0} \nabla \cdot \widehat{(\underline{\mathbf{v}}\rho)}_n^{\sim 0} + \Delta_h Q_{A^L} \right] \\ & - \left[c\widehat{\rho}_n^{\sim 0} \frac{\partial}{\partial z} \widehat{(w\rho)}_n^{\sim 0} + \Delta_z Q_{A^L} \right]. \end{aligned}$$

Making use of Eq. (3.26), and denoting

$$TS_{A^L} \equiv \frac{1}{4} \left(\widehat{\rho}_n^{\sim 0} \widehat{(w\rho)}_n^{\sim 0} + \widehat{(\rho^2)}_n^{\sim 0} \widehat{w}_n^{\sim 0} \right) \frac{\partial c}{\partial z}, \tag{3.40}$$

this can be further written as

$$\begin{aligned} (\mathrm{AD}) \quad = \quad & \Delta_h Q_{A^L} + \Delta_z Q_{A^L} \\ & - \left[c\widehat{\rho}_n^{\sim 0} \nabla \cdot \widehat{(\underline{\mathbf{v}}\rho)}_n^{\sim 0} + \Delta_h Q_{A^L} - \frac{1}{4} c \left(\widehat{\rho}_n^{\sim 0} \widehat{(\rho \nabla \cdot \underline{\mathbf{v}})}_n^{\sim 0} + \widehat{(\rho^2)}_n^{\sim 0} \nabla \cdot \widehat{\underline{\mathbf{v}}}_n^{\sim 0} \right) \right] \\ & - \left[c\widehat{\rho}_n^{\sim 0} \frac{\partial}{\partial z} \widehat{(w\rho)}_n^{\sim 0} + \Delta_z Q_{A^L} + TS_{A^L} - \frac{1}{4} c \left(\widehat{\rho}_n^{\sim 0} \widehat{\left(\rho \frac{\partial w}{\partial z}\right)}_n^{\sim 0} + \widehat{(\rho^2)}_n^{\sim 0} \frac{\partial \widehat{w}_n^{\sim 0}}{\partial z} \right) \right] \\ & + TS_{A^L} \\ \equiv \quad & \Delta_h Q_{A^L} + \Delta_z Q_{A^L} + T_{A^L, \partial_h \rho} + T_{A^L, \partial_z \rho} + TS_{A^L}, \end{aligned} \tag{3.41}$$

where $\Delta_h Q_{A^L}$ and $\Delta_z Q_{A^L}$ are, as we already know, the horizontal and vertical transports. The other pair,

$$\begin{aligned} T_{A^L,\partial_h\rho} &\equiv -c\widehat{\rho}_n^{\sim 0}\nabla\cdot\widehat{(\underline{\mathbf{v}}\rho)}_n^{\sim 0} - \Delta_h Q_{A^L} \\ &\quad +\frac{1}{4}c\left(\widehat{\rho}_n^{\sim 0}\widehat{(\rho\nabla\cdot\underline{\mathbf{v}})}_n^{\sim 0} + \widehat{(\rho^2)}_n^{\sim 0}\nabla\cdot\widehat{\underline{\mathbf{v}}}_n^{\sim 0}\right) \end{aligned} \tag{3.42}$$

$$\begin{aligned} T_{A^L,\partial_z\rho} &\equiv -c\widehat{\rho}_n^{\sim 0}\frac{\partial}{\partial z}\widehat{(w\rho)}_n^{\sim 0} - \Delta_z Q_{A^L} - TS_{A^L} \\ &\quad +\frac{1}{4}c\left(\widehat{\rho}_n^{\sim 0}\widehat{(\rho\frac{\partial w}{\partial z})}_n^{\sim 0} + \widehat{(\rho^2)}_n^{\sim 0}\frac{\partial\widehat{w}_n^{\sim 0}}{\partial z}\right) \end{aligned} \tag{3.43}$$

represent two mechanisms for APE transfer across scale windows. (The explanation is put off to §3.6.2.) By some algebraic manipulation, they can be reduced to

$$\begin{aligned} T_{A^L,\partial_h\rho} &= \frac{c}{2}\left[\widehat{\rho}_n^{\sim 0}\widehat{(\underline{\mathbf{v}}\cdot\nabla\rho)}_n^{\sim 0} - \nabla\widehat{\rho}_n^{\sim 0}\cdot\widehat{(\underline{\mathbf{v}}\rho)}_n^{\sim 0}\right], \\ T_{A^L,\partial_z\rho} &= \frac{c}{2}\left[\widehat{\rho}_n^{\sim 0}\widehat{\left(w\frac{\partial\rho}{\partial z}\right)}_n^{\sim 0} - \frac{\partial\widehat{\rho}_n^{\sim 0}}{\partial z}\widehat{(w\rho)}_n^{\sim 0}\right]. \end{aligned}$$

But in the later development, we won't use these simplified expressions. The original equations, more complicated as they might be, preserve better the total energy during numerical computations.

If necessary, $\Delta_h Q_{A^L}$ and $T_{A^L,\partial_h\rho}$ can be further decomposed as

$$\Delta_h Q_{A^L} = \Delta_x Q_{A^L} + \Delta_y Q_{A^L}, \tag{3.44}$$

$$T_{A^L,\partial_h\rho} = T_{A^L,\partial_x\rho} + T_{A^L,\partial_y\rho}, \tag{3.45}$$

where $\Delta_x Q_{A^L}$ ($T_{A^L,\partial_x\rho}$) and $\Delta_y Q_{A^L}$ ($T_{A^L,\partial_y\rho}$) are given by the equation for $\Delta_h Q_{A^L}$ ($T_{A^L,\partial_h\rho}$) with the gradient operator ∇ replaced by $\partial/\partial x$ and $\partial/\partial y$, respectively.

Besides the above fluxes and transfers, there exists an extra term

$$\begin{aligned} TS_{A^L} &\equiv \frac{1}{4}\left(\widehat{\rho}_n^{\sim 0}\widehat{(w\rho)}_n^{\sim 0} + \widehat{(\rho^2)}_n^{\sim 0}\widehat{w}_n^{\sim 0}\right)\frac{\partial c}{\partial z} \\ &= \frac{1}{4}c\left(\widehat{\rho}_n^{\sim 0}\widehat{(w\rho)}_n^{\sim 0} + \widehat{(\rho^2)}_n^{\sim 0}\widehat{w}_n^{\sim 0}\right)\frac{\partial(\log c)}{\partial z} \\ &= -\frac{1}{4}c\left(\widehat{\rho}_n^{\sim 0}\widehat{(w\rho)}_n^{\sim 0} + \widehat{(\rho^2)}_n^{\sim 0}\widehat{w}_n^{\sim 0}\right)\frac{\partial\left(\log N^2\right)}{\partial z} \end{aligned} \tag{3.46}$$

in the (AD) decomposition (recall $c = g^2/\rho_0^2 N^2$). We will show in §3.6.2 that this term represents an apparent source/sink due to the stationary vertical shear of density [cf. (3.18)], as well as an energy transfer.

Next consider the term $w\frac{N^2\rho_0}{g}$. Recall that N^2 is a function of z only. It is thus immune to the transform. So

$$
\begin{aligned}
c\widehat{\rho}_n^{\sim 0}\frac{\rho_0}{g}\cdot\widehat{(wN^2)}_n^{\sim 0} &= c\frac{N^2\rho_0}{g}\widehat{\rho}_n^{\sim 0}\widehat{w}_n^{\sim 0} \\
&= \frac{g}{\rho_0}\widehat{w}_n^{\sim 0}\widehat{\rho}_n^{\sim 0} = b_n^L,
\end{aligned}
\tag{3.47}
$$

which is exactly the buoyancy conversion between large-scale available potential and kinetic energies.

The diffusion terms are treated the same way as before, they are merely denoted as

$$
F_{A^L,z} = c\widehat{\rho}_n^{\sim 0}\widehat{(F_{\rho,z})}_n^{\sim 0}, \tag{3.48}
$$

$$
F_{A^L,h} = c\widehat{\rho}_n^{\sim 0}\widehat{(F_{\rho,h})}_n^{\sim 0} + (\Delta t)^2 c\left(\hat{\delta}^2_{n^2}\widehat{\rho}_n^{\sim 0}\cdot\hat{\delta}_n\widehat{\rho}_n^{\sim 0}\right). \tag{3.49}
$$

Put all the above equations together (with the aid of notations (3.38), (3.39) and (3.46)),

$$
\begin{aligned}
\hat{\delta}_n A_n^L = &\ \Delta_h Q_{A^L} + \Delta_z Q_{A^L} \\
&+\left[-c\widehat{\rho}_n^{\sim 0}\nabla\cdot\widehat{(\underline{\mathbf{v}}\rho)}_n^{\sim 0} - \Delta_h Q_{A^L} + \frac{1}{4}c\left(\widehat{\rho}_n^{\sim 0}\widehat{(\rho\nabla\cdot\underline{\mathbf{v}})}_n^{\sim 0} + \widehat{(\rho^2)}_n^{\sim 0}\nabla\cdot\widehat{\underline{\mathbf{v}}}_n^{\sim 0}\right)\right] \\
&+\left[-c\widehat{\rho}_n^{\sim 0}\frac{\partial}{\partial z}\widehat{(w\rho)}_n^{\sim 0} - \Delta_z Q_{A^L} - TS_{A^L} + \frac{1}{4}c\left(\widehat{\rho}_n^{\sim 0}\widehat{(\rho\frac{\partial w}{\partial z})}_n^{\sim 0} + \widehat{(\rho^2)}_n^{\sim 0}\frac{\partial\widehat{w}_n^{\sim 0}}{\partial z}\right)\right] \\
&+TS_{A^L} + \frac{g}{\rho_0}\widehat{w}_n^{\sim 0}\widehat{\rho}_n^{\sim 0} + F_{A^L,z} + F_{A^L,h},
\end{aligned}
\tag{3.50}
$$

or, in a symbolic form,

$$
\dot{A}_n^L = \Delta_h Q_{A^L} + \Delta_z Q_{A^L} + T_{A^L,\partial_h\rho} + T_{A^L,\partial_z\rho} + TS_{A^L} + b_n^L + F_{A^L,z} + F_{A^L,h}. \tag{3.51}
$$

3.3 Meso-scale energetics

In a classical Reynolds formulation, the meso-scale energetics are usually quite different in form from their large-scale counterparts (e.g., Harrison and Robinson, 1979). Our multiscale window transform, however, will eliminate this difference and bring forth two meso-scale energy equations in a form exactly the same as (3.35) and (3.50). This will be clear upon closing this section.

Following the definition in §2.2.5, the meso-scale kinetic energy for location n (constant multiplier 2^{j_2} omitted), which is denoted by K_n^M hereafter, is

$$
K_n^M = \frac{1}{2}\widehat{\underline{\mathbf{v}}}_n^{\sim 1}\cdot\widehat{\underline{\mathbf{v}}}_n^{\sim 1}. \tag{3.52}
$$

Application of the meso-scale window transform to (3.12) and (3.13) gives a continuity equation and a hydrostatic equation in the phase space:

$$\frac{\partial \widehat{w}_n^{\sim 1}}{\partial z} + \nabla \cdot \widehat{\underline{\mathbf{v}}}_n^{\sim 1} = 0, \tag{3.53}$$

$$\frac{\partial \widehat{P}_n^{\sim 1}}{\partial z} = -\widehat{\rho}_n^{\sim 1} g. \tag{3.54}$$

Reconstructed back to the physical space, they become, in their original forms,

$$\frac{\partial w^{\sim 1}}{\partial z} + \nabla \cdot \underline{\mathbf{v}}^{\sim 1} = 0, \tag{3.55}$$

$$\frac{\partial P^{\sim 1}}{\partial z} = -\rho^{\sim 1} g. \tag{3.56}$$

This form-conservation stems from the fact that they involve only linear operations. Eqs. (3.55) and (3.56) are to be used later in derivations.

In order to obtain the evolution for K_n^M, one needs to take the meso-scale window transform of the momentum equation. As in the preceding section, this is somewhat cumbersome due to the nonlinearity. We will do it step by step. First apply the meso-scale synthesis to the time-discretized momentum equation once. What results is the meso-scale momentum equation

$$\delta_t \underline{\mathbf{v}}^{\sim 1} = -\nabla K^{\sim 1} - \left(w \frac{\partial \underline{\mathbf{v}}}{\partial z}\right)^{\sim 1} - f\underline{\mathbf{k}} \wedge \underline{\mathbf{v}}^{\sim 1} - (\zeta \underline{\mathbf{k}} \wedge \underline{\mathbf{v}})^{\sim 1} - \frac{1}{\rho_0} \nabla P^{\sim 1} + \underline{\mathbf{F}}_{mz}^{\sim 1} + \underline{\mathbf{F}}_{mh}^{\sim 1}. \tag{3.57}$$

Take a dot product with $\widehat{\underline{\mathbf{v}}}_n^{\sim 1} \phi_n^{\varrho, j_2}(t)$ followed by an integration over the $[0, \varrho)$. As before, the left hand side is equal to

$$\hat{\delta}_n K_n^M - (\Delta t)^2 \left(\hat{\delta}_{n^2}^2 \widehat{\underline{\mathbf{v}}}_n^{\sim 1} \cdot \hat{\delta}_n \widehat{\underline{\mathbf{v}}}_n^{\sim 1}\right),$$

the rate of change of the meso-scale energy corrected by a small term of order $(\Delta t)^2$. On the right hand side, the advection-related terms become

$$\begin{aligned}
&- \int_0^{\varrho} \widehat{\underline{\mathbf{v}}}_n^{\sim 1} \phi_n^{\varrho, j_2}(t) \cdot \left[\nabla K^{\sim 1} + \left(w \frac{\partial \underline{\mathbf{v}}}{\partial z}\right)^{\sim 1}\right] dt \\
&= -\widehat{\underline{\mathbf{v}}}_n^{\sim 1} \cdot \nabla \widehat{K}_n^{\sim 1} - \widehat{\underline{\mathbf{v}}}_n^{\sim 1} \cdot \widehat{\left(w \frac{\partial \underline{\mathbf{v}}}{\partial z}\right)}_n^{\sim 1} \\
&= -\nabla \cdot \left[\widehat{\underline{\mathbf{v}}}_n^{\sim 1} \widehat{K}_n^{\sim 1}\right] - \frac{\partial}{\partial z} \left(\widehat{w}_n^{\sim 1} \widehat{K}_n^{\sim 1}\right) \\
&\quad + \widehat{w}_n^{\sim 1} \frac{\partial \widehat{K}_n^{\sim 1}}{\partial z} - \widehat{\underline{\mathbf{v}}}_n^{\sim 1} \cdot \widehat{\left(w \frac{\partial \underline{\mathbf{v}}}{\partial z}\right)}_n^{\sim 1}.
\end{aligned}$$

Recall that

$$\frac{\partial K}{\partial z} = -\underline{\mathbf{k}} \cdot \underline{\zeta}^* \wedge \underline{\mathbf{v}},$$

$$w\frac{\partial \mathbf{v}}{\partial z} = -\underline{\mathbf{k}} \wedge (w\underline{\zeta}^*).$$

The above energetic terms are thereby reduced to

$$\begin{aligned}
& - \nabla \cdot \left[\widehat{\underline{\mathbf{v}}}_n^{\sim 1} \widehat{K}_n^{\sim 1}\right] - \frac{\partial}{\partial z}\left(\widehat{w}_n^{\sim 1}\widehat{K}_n^{\sim 1}\right) \\
& + \widehat{w}_n^{\sim 1}\frac{\partial \widehat{K}_n^{\sim 1}}{\partial z} + \widehat{\underline{\mathbf{v}}}_n^{\sim 1} \cdot \underline{\mathbf{k}} \wedge \left(\widehat{\underline{\zeta}^* w}\right)_n^{\sim 1} \\
& \equiv \Delta_h Q_{K^M} + \Delta_z Q_{K^M} + T_{K^M, w\partial_z} + T_{K^M, \partial_z v},
\end{aligned} \tag{3.58}$$

where $T_{K^M,\partial_z v}$ together with $T_{K^M,w\partial_z}$ represents the transfer due to the vertical structure evolution of the horizontal velocity field.

The contribution from the vorticity field is obtained the same way we did before in deriving the large-scale energetics. The result is

$$\begin{aligned}
T_{K^M,\zeta} &= -\int_0^{\varrho} \widehat{\underline{\mathbf{v}}}_n^{\sim 1}\phi_n^{\varrho, j_2}(t) \cdot \left[f\underline{\mathbf{k}} \wedge \underline{\mathbf{v}}^{\sim 1} + \underline{\mathbf{k}} \wedge (\zeta \underline{\mathbf{v}})^{\sim 1}\right]\ dt \\
&= -\widehat{\underline{\mathbf{v}}}_n^{\sim 1} \cdot \left[f\underline{\mathbf{k}} \wedge \widehat{\underline{\mathbf{v}}}_n^{\sim 1} + \underline{\mathbf{k}} \wedge \widehat{(\zeta \underline{\mathbf{v}})}_n^{\sim 1}\right] \\
&= -\widehat{\underline{\mathbf{v}}}_n^{\sim 1} \cdot \underline{\mathbf{k}} \wedge \widehat{(\zeta \underline{\mathbf{v}})}_n^{\sim 1}.
\end{aligned} \tag{3.59}$$

Similar action performed on the pressure terms yields

$$\begin{aligned}
-\widehat{\underline{\mathbf{v}}}_n^{\sim 1} \cdot \nabla \frac{\widehat{P}_n^{\sim 1}}{\rho_0} &= -\nabla \cdot \left(\widehat{\underline{\mathbf{v}}}_n^{\sim 1}\frac{\widehat{P}_n^{\sim 1}}{\rho_0}\right) - \frac{\partial}{\partial z}\left(\widehat{w}_n^{\sim 1}\frac{\widehat{P}_n^{\sim 1}}{\rho_0}\right) + \widehat{w}_n^{\sim 1}\frac{\partial}{\partial z}\left(\frac{\widehat{P}_n^{\sim 1}}{\rho_0}\right) \\
&= -\nabla \cdot \left(\widehat{\underline{\mathbf{v}}}_n^{\sim 1}\frac{\widehat{P}_n^{\sim 1}}{\rho_0}\right) - \frac{\partial}{\partial z}\left(\widehat{w}_n^{\sim 1}\frac{\widehat{P}_n^{\sim 1}}{\rho_0}\right) - \frac{g}{\rho_0}\widehat{w}_n^{\sim 1}\widehat{\rho}_n^{\sim 1} && (3.60) \\
&\equiv \Delta_h Q_{P^M} + \Delta_z Q_{P^M} - b_n^M. && (3.61)
\end{aligned}$$

All the above terms, together with the dissipation which we just write symbolically as $F_{K^M,z}$ and $F_{K^M,h}$, form an evolution law for the meso-scale kinetic energy

$$\begin{aligned}
\hat{\delta}_n K_n^M = & -\nabla \cdot \left[\widehat{\underline{\mathbf{v}}}_n^{\sim 1} \widehat{K}_n^{\sim 1}\right] - \frac{\partial}{\partial z}\left(\widehat{w}_n^{\sim 1}\widehat{K}_n^{\sim 1}\right) \\
& + \widehat{w}_n^{\sim 1}\frac{\partial \widehat{K}_n^{\sim 1}}{\partial z} + \widehat{\underline{\mathbf{v}}}_n^{\sim 1} \cdot \underline{\mathbf{k}} \wedge \left(\widehat{\underline{\zeta}^* w}\right)_n^{\sim 1} \\
& - \widehat{\underline{\mathbf{v}}}_n^{\sim 1} \cdot \underline{\mathbf{k}} \wedge \widehat{(\zeta \underline{\mathbf{v}})}_n^{\sim 1} \\
& - \nabla \cdot \left(\widehat{\underline{\mathbf{v}}}_n^{\sim 1}\frac{\widehat{P}_n^{\sim 1}}{\rho_0}\right) - \frac{\partial}{\partial z}\left(\widehat{w}_n^{\sim 1}\frac{\widehat{P}_n^{\sim 1}}{\rho_0}\right) \\
& - \frac{g}{\rho_0}\widehat{w}_n^{\sim 1}\widehat{\rho}_n^{\sim 1} + F_{K^M,z} + F_{K^M,h}.
\end{aligned} \tag{3.62}$$

Symbolically, this is

$$\dot{K}_n^M = \Delta_h Q_{K^M} + \Delta_z Q_{K^M} + T_{K^M,w\partial_z} + T_{K^M,\partial_z v} + T_{K^M,\zeta}$$

$$+ \quad \Delta_h Q_{P^M} + \Delta_z Q_{P^M} - b_n^M + F_{K^M,z} + F_{K^M,h}. \tag{3.63}$$

To get the meso-scale available potential energy equation, recall the time-discretized perturbation density equation is

$$\delta_t \rho = -\underline{\mathbf{v}} \cdot \nabla\rho - w\frac{\partial \rho}{\partial z} + w\frac{N^2\rho_0}{g} + F_{\rho z} + F_{\rho h},$$

which, after being applied with the meso-scale synthesis, becomes

$$\delta_t \rho^{\sim 1} = -\nabla \cdot (\rho \underline{\mathbf{v}})^{\sim 1} - \frac{\partial (w\rho)^{\sim 1}}{\partial z} + w^{\sim 1}\frac{N^2\rho_0}{g} + F_{\rho z}^{\sim 1} + F_{\rho h}^{\sim 1}. \tag{3.64}$$

Transforming it with $\phi_n^{\varrho, j_2}(t)$, one gets

$$\hat{\delta}_n \widehat{\rho}_n^{\sim 1} = -\nabla \cdot \widehat{(\rho \underline{\mathbf{v}})}_n^{\sim 1} - \frac{\partial}{\partial z}\widehat{(w\rho)}_n^{\sim 1} + \frac{N^2\rho_0}{g}\widehat{w}_n^{\sim 1} + \widehat{F}_{\rho,z}^{\sim 1} + \widehat{F}_{\rho,h}^{\sim 1}. \tag{3.65}$$

The available potential energy evolution is obtained by multiplying this equation by $c\widehat{\rho}_n^{\sim 1}$ (recall the temporary notation $c \equiv g^2/\rho_0^2 N^2$). The resulting left hand side,

$$\hat{\delta}_n A_n^M - (\Delta t)^2 c (\hat{\delta}_{n^2}^2 \widehat{\rho}_n^{\sim 1} \cdot \hat{\delta}_n \widehat{\rho}_n^{\sim 1}),$$

is dominated by the rate of change of

$$A_n^M = \frac{1}{2}c\left(\widehat{\rho}_n^{\sim 1}\right)^2 = \frac{1}{2}\frac{g^2}{\rho_0^2 N^2}\left(\widehat{\rho}_n^{\sim 1}\right)^2, \tag{3.66}$$

i.e., the evolution of meso-scale available potential energy. The other term, as before, is significant only when the grid-scale processes are of interest and might be put in the sub-grid process parameterization.

To further the derivation, apply the transport-transfer decomposition technique used before to the advection-related terms. Let

$$\Delta_h Q_{A^M} \equiv -\nabla \cdot \left[\frac{1}{4}c\widehat{\rho}_n^{\sim 1}\widehat{(\underline{\mathbf{v}}\rho)}_n^{\sim 1} + \frac{1}{4}c\widehat{(\rho^2)}_n^{\sim 1}\widehat{\underline{\mathbf{v}}}_n^{\sim 1}\right]; \tag{3.67}$$

$$\Delta_z Q_{A^M} \equiv -\frac{\partial}{\partial z}\left[\frac{1}{4}c\widehat{\rho}_n^{\sim 1}\widehat{(w\rho)}_n^{\sim 1} + \frac{1}{4}c\widehat{(\rho^2)}_n^{\sim 1}\widehat{w}_n^{\sim 1}\right]; \tag{3.68}$$

$$TS_{A^M} \equiv \frac{1}{4}\left(\widehat{\rho}_n^{\sim 1}\widehat{(w\rho)}_n^{\sim 1} + \widehat{(\rho^2)}_n^{\sim 1}\widehat{w}_n^{\sim 1}\right)\frac{\partial c}{\partial z}. \tag{3.69}$$

This yields,

$$\Delta_h Q_{A^M} + \Delta_z Q_{A^M}$$

$$
\begin{aligned}
& - \left[c\widehat{\rho}_n^{\sim 1} \nabla \cdot \widehat{(\underline{\mathbf{v}}\rho)}_n^{\sim 1} + \Delta_h Q_{A^M} - \frac{1}{4} c \left(\widehat{\rho}_n^{\sim 1} (\rho \widehat{\nabla \cdot \underline{\mathbf{v}}})_n^{\sim 1} + \widehat{(\rho^2)}_n^{\sim 1} \nabla \cdot \widehat{\underline{\mathbf{v}}}_n^{\sim 1} \right) \right] \\
& - \left[c\widehat{\rho}_n^{\sim 1} \frac{\partial}{\partial z} \widehat{(w\rho)}_n^{\sim 1} + \Delta_z Q_{A^M} + TS_{A^M} - \frac{1}{4} c \left(\widehat{\rho}_n^{\sim 1} \widehat{(\rho \frac{\partial w}{\partial z})}_n^{\sim 1} + \widehat{(\rho^2)}_n^{\sim 1} \frac{\partial \widehat{w}_n^{\sim 1}}{\partial z} \right) \right] \\
& + TS_{A^M} \\
\equiv \; & \Delta_h Q_{A^M} + \Delta_z Q_{A^M} + T_{A^M, \partial_h \rho} + T_{A^M, \partial_z \rho} + TS_{A^M},
\end{aligned}
$$

The next mechanism that accounts for the evolution of A_n^M is the buoyancy conversion:

$$
\begin{aligned}
c\widehat{\rho}_n^{\sim 1} \cdot \frac{N^2 \rho_0}{g} \widehat{w}_n^{\sim 1} & = \frac{g}{\rho_0} \widehat{w}_n^{\sim 1} \widehat{\rho}_n^{\sim 1} \\
& \equiv b_n^M.
\end{aligned} \tag{3.70}
$$

It is equal in magnitude to its counterpart in the meso-scale kinetic equation, but with an opposite sign.

All the above terms, plus the diffusion, add up to form the meso-scale available potential energy equation [with the aid of notations (3.67), (3.68), and (3.69)]:

$$
\begin{aligned}
\hat{\delta}_n A_n^M = \; & \Delta_h Q_{A^M} + \Delta_z Q_{A^M} \\
& + \left[-c\widehat{\rho}_n^{\sim 1} \nabla \cdot \widehat{(\underline{\mathbf{v}}\rho)}_n^{\sim 1} - \Delta_h Q_{A^M} + \frac{1}{4} c \left(\widehat{\rho}_n^{\sim 1} (\rho \widehat{\nabla \cdot \underline{\mathbf{v}}})_n^{\sim 1} + \widehat{(\rho^2)}_n^{\sim 1} \nabla \cdot \widehat{\underline{\mathbf{v}}}_n^{\sim 1} \right) \right] \\
& + \left[-c\widehat{\rho}_n^{\sim 1} \frac{\partial}{\partial z} \widehat{(w\rho)}_n^{\sim 1} - \Delta_z Q_{A^M} - TS_{A^M} + \frac{1}{4} c \left(\widehat{\rho}_n^{\sim 1} \widehat{(\rho \frac{\partial w}{\partial z})}_n^{\sim 1} + \widehat{(\rho^2)}_n^{\sim 1} \frac{\partial \widehat{w}_n^{\sim 1}}{\partial z} \right) \right] \\
& + TS_{A^M} + \frac{g}{\rho_0} \widehat{w}_n^{\sim 1} \widehat{\rho}_n^{\sim 1} + F_{A^M, z} + F_{A^M, h},
\end{aligned} \tag{3.71}
$$

which is also written as

$$
\dot{A}_n^M = \Delta_h Q_{A^M} + \Delta_z Q_{A^M} + T_{A^M, \partial_h \rho} + T_{A^M, \partial_z \rho} + TS_{A^M} + b_n^M + F_{A^M, z} + F_{A^M, h} \tag{3.72}
$$

for the sake of notational simplicity.

3.4 Sub-mesoscale energetics

Arbitrarily many scale windows and their energetics can been obtained in a way as the meso-scale energy equations are obtained. We usually consider only one window more: the sub-mesoscale window. The derivation will be the same as before. What is needed is to change the superscript, which should be switched from ~ 1 to ~ 2. On this window, the kinetic and available potential energies are

$$
\begin{aligned}
K_n^S &= \frac{1}{2}\widehat{\underline{\mathbf{v}}}_n^{\sim 2} \cdot \widehat{\underline{\mathbf{v}}}_n^{\sim 2}, & (3.73)\\
A_n^S &= \frac{1}{2}\frac{g^2}{\rho_0 N^2}\left(\widehat{\rho}_n^{\sim 2}\right)^2. & (3.74)
\end{aligned}
$$

In analogy to their meso-scale counterparts, the equations governing the evolutions of K_n^S and A_n^S are

$$
\begin{aligned}
\hat{\delta}_n K_n^S = & -\nabla\cdot\left[\widehat{\underline{\mathbf{v}}}_n^{\sim 2}\widehat{K}_n^{\sim 2}\right] - \frac{\partial}{\partial z}\left(\widehat{w}_n^{\sim 2}\widehat{K}_n^{\sim 2}\right)\\
& +\widehat{w}_n^{\sim 2}\frac{\partial \widehat{K}_n^{\sim 2}}{\partial z} + \widehat{\underline{\mathbf{v}}}_n^{\sim 2}\cdot\underline{\mathbf{k}}\wedge\left(\widehat{\underline{\zeta}^* w}\right)_n^{\sim 2}\\
& -\widehat{\underline{\mathbf{v}}}_n^{\sim 2}\cdot\underline{\mathbf{k}}\wedge\left(\widehat{\zeta\underline{\mathbf{v}}}\right)_n^{\sim 2}\\
& -\nabla\cdot\left(\widehat{\underline{\mathbf{v}}}_n^{\sim 2}\frac{\widehat{P}_n^{\sim 2}}{\rho_0}\right) - \frac{\partial}{\partial z}\left(\widehat{w}_n^{\sim 2}\frac{\widehat{P}_n^{\sim 2}}{\rho_0}\right)\\
& -\frac{g}{\rho_0}\widehat{w}_n^{\sim 2}\widehat{\rho}_n^{\sim 2} + F_{K^S,z} + F_{K^S,h}. \qquad (3.75)
\end{aligned}
$$

and

$$
\begin{aligned}
\hat{\delta}_n A_n^S = \; & \Delta_h Q_{A^S} + \Delta_z Q_{A^S}\\
& + \left[-c\widehat{\rho}_n^{\sim 2}\nabla\cdot\left(\widehat{\underline{\mathbf{v}}\rho}\right)_n^{\sim 2} - \Delta_h Q_{A^S} + \frac{1}{4}c\left(\widehat{\rho}_n^{\sim 2}\left(\rho\widehat{\nabla\cdot\underline{\mathbf{v}}}\right)_n^{\sim 2} + \left(\widehat{\rho^2}\right)_n^{\sim 2}\nabla\cdot\widehat{\underline{\mathbf{v}}}_n^{\sim 2}\right)\right]\\
& + \left[-c\widehat{\rho}_n^{\sim 2}\frac{\partial}{\partial z}\left(\widehat{w\rho}\right)_n^{\sim 2} - \Delta_z Q_{A^S} - TS_{A^S} + \frac{1}{4}c\left(\widehat{\rho}_n^{\sim 2}\left(\rho\widehat{\frac{\partial w}{\partial z}}\right)_n^{\sim 2} + \left(\widehat{\rho^2}\right)_n^{\sim 2}\frac{\partial \widehat{w}_n^{\sim 2}}{\partial z}\right)\right]\\
& + TS_{A^S} + \frac{g}{\rho_0}\widehat{w}_n^{\sim 2}\widehat{\rho}_n^{\sim 2} + F_{A^S,z} + F_{A^S,h}, \qquad (3.76)
\end{aligned}
$$

where

$$
\begin{aligned}
\Delta_h Q_{A^S} &\equiv -\nabla\cdot\left[\frac{1}{4}c\widehat{\rho}_n^{\sim 2}\left(\widehat{\underline{\mathbf{v}}\rho}\right)_n^{\sim 2} + \frac{1}{4}c\left(\widehat{\rho^2}\right)_n^{\sim 2}\widehat{\underline{\mathbf{v}}}_n^{\sim 2}\right]; & (3.77)\\
\Delta_z Q_{A^S} &\equiv -\frac{\partial}{\partial z}\left[\frac{1}{4}c\widehat{\rho}_n^{\sim 2}\left(\widehat{w\rho}\right)_n^{\sim 2} + \frac{1}{4}c\left(\widehat{\rho^2}\right)_n^{\sim 2}\widehat{w}_n^{\sim 2}\right]; & (3.78)\\
TS_{A^S} &\equiv \frac{1}{4}\left(\widehat{\rho}_n^{\sim 2}\left(\widehat{w\rho}\right)_n^{\sim 2} + \left(\widehat{\rho^2}\right)_n^{\sim 2}\widehat{w}_n^{\sim 2}\right)\frac{\partial c}{\partial z}. & (3.79)
\end{aligned}
$$

As before, these equations are also symbolically written as

$$
\begin{aligned}
\dot{K}_n^S &= \Delta_h Q_{K^S} + \Delta_z Q_{K^S} + T_{K^S,w\partial_z} + T_{K^S,\partial_z v} + T_{K^S,\zeta}\\
&\quad + \Delta_h Q_{P^S} + \Delta_z Q_{P^S} - b_n^S + F_{K^S,z} + F_{K^S,h}, \qquad (3.80)\\
\dot{A}_n^S &= \Delta_h Q_{A^S} + \Delta_z Q_{A^S} + T_{A^S,\partial_h \rho} + T_{A^S,\partial_z \rho}
\end{aligned}
$$

$$+TS_{A^S} + b_n^S + F_{A^S,z} + F_{A^S,h} \tag{3.81}$$

for later use.

3.5 Connection to the classical energetic formulation

The connection between our MS-EVA to the classical energetic equations is easily seen if we begin the formulation with the momentum equation (dissipation ignored) in a form

$$\frac{\partial \underline{\mathbf{v}}}{\partial t} = -\nabla \cdot (\underline{\mathbf{v}}\ \underline{\mathbf{v}}) - \frac{\partial}{\partial z}(w\underline{\mathbf{v}}) - f\underline{\mathbf{k}} \wedge \underline{\mathbf{v}} - \frac{1}{\rho_0}\nabla P \tag{3.82}$$

rather than (3.11). A large-scale transform followed by a dot product with $\widehat{\underline{\mathbf{v}}}_n^{\sim 0}$ yields

$$\widehat{\underline{\mathbf{v}}}_n^{\sim 0} \cdot \left(\widehat{\frac{\partial \underline{\mathbf{v}}}{\partial t}}\right)_n^{\sim 0} = \underbrace{\widehat{\underline{\mathbf{v}}}_n^{\sim 0} \cdot \left[-\nabla \cdot \widehat{(\underline{\mathbf{v}}\ \underline{\mathbf{v}})}_n^{\sim 0} - \frac{\partial}{\partial z}\widehat{(w\underline{\mathbf{v}})}_n^{\sim 0}\right]}_{(\mathbf{I})} \underbrace{- \widehat{\underline{\mathbf{v}}}_n^{\sim 0} \cdot \nabla \frac{\widehat{P}_n^{\sim 0}}{\rho_0}}_{(\mathbf{II})}, \tag{3.83}$$

where term **(I)** is equal to the sum of all the transport and transfer terms in our previous large-scale KE equation, and

$$(\mathbf{II}) = \Delta_h Q_{P^L} + \Delta_z Q_{P^L} - b_n^L.$$

We will see what this equation reduces to if $j_1 = j_2$ (that is to say, only two scale windows are considered), $j_0 = 0$, and a periodic extension is employed.

Let q be any field variable (u, v, w, or P). A two-scale window decomposition means

$$q = q^{\sim 0} + q^{\sim 1}. \tag{3.84}$$

With the choice of zero j_0 and periodic extension, we know from the preceding chapter that $q^{\sim 0}$ is constant in time and is equal to $\bar{q}$ or $2^{j_2/2}\widehat{q}_n^{\sim 0}$ in magnitude, that is,

$$q^{\sim 0} = \bar{q} = 2^{j_2/2}\ \widehat{q}_n^{\sim 0}, \tag{3.85}$$

and hence

$$q^{\sim 1} = q - q^{\sim 0} = q - \bar{q} = q'. \tag{3.86}$$

In this case, Eq. (3.84) is simply

$$q = \bar{q} + q', \tag{3.87}$$

with $\bar{q}$ and q' satisfying

$$\widehat{(\bar{q})}_n^{\sim 0} = \widehat{(q^{\sim 0})}_n^{\sim 0} = q^{\sim 0} = 2^{-j_2/2}\ \bar{q}, \tag{3.88}$$

$$\widehat{(q')}_n^{\sim 0} = \widehat{(q^{\sim 1})}_n^{\sim 0} = 0, \tag{3.89}$$

by the properties of multiscale window transform.

Substituting $\underline{\mathbf{v}}$ and w for the q in (3.87), the velocity field is decomposed as

$$\underline{\mathbf{v}} = \bar{\underline{\mathbf{v}}} + \underline{\mathbf{v}}', \qquad w = \bar{w} + w'.$$

Let $K^L \equiv \frac{1}{2}\bar{\underline{\mathbf{v}}} \cdot \bar{\underline{\mathbf{v}}}$. The equivalence between the large-scale transform and duration average allows an expression of the large-scale kinetic energy K_n^L in terms of K^L. In fact,

$$K_n^L = 2^{j_2}\left(\frac{1}{2}\hat{\underline{\mathbf{v}}}_n^{\sim 0} \cdot \hat{\underline{\mathbf{v}}}_n^{\sim 0}\right) = \frac{1}{2}\bar{\underline{\mathbf{v}}} \cdot \bar{\underline{\mathbf{v}}} = K^L. \tag{3.90}$$

Note here we have taken into account the multiplier 2^{j_2}. These facts are now used to simplify the term **(I)** of Eq. (3.82). With the two-scale decomposition, the dyad $(\underline{\mathbf{v}}\ \underline{\mathbf{v}})$ after transforming is expanded as

$$\begin{aligned}
\widehat{(\underline{\mathbf{v}}\ \underline{\mathbf{v}})}_n^{\sim 0} &= \widehat{(\bar{\underline{\mathbf{v}}}\ \bar{\underline{\mathbf{v}}})}_n^{\sim 0} + \widehat{(\bar{\underline{\mathbf{v}}}\ \underline{\mathbf{v}}')}_n^{\sim 0} + \widehat{(\underline{\mathbf{v}}'\ \bar{\underline{\mathbf{v}}})}_n^{\sim 0} + \widehat{(\underline{\mathbf{v}}'\ \underline{\mathbf{v}}')}_n^{\sim 0} &(3.91)\\
&= \bar{\underline{\mathbf{v}}}\ \hat{\bar{\underline{\mathbf{v}}}}_n^{\sim 0} + \bar{\underline{\mathbf{v}}}\ \hat{\underline{\mathbf{v}}}'^{\sim 0}_n + \hat{\underline{\mathbf{v}}}'^{\sim 0}_n\ \bar{\underline{\mathbf{v}}} + \widehat{(\underline{\mathbf{v}}'\underline{\mathbf{v}}')}_n^{\sim 0} &(3.92)\\
&= \bar{\underline{\mathbf{v}}}\hat{\underline{\mathbf{v}}}_n^{\sim 0} + \widehat{(\underline{\mathbf{v}}'\underline{\mathbf{v}}')}_n^{\sim 0}. &(3.93)
\end{aligned}$$

Likewise,

$$\widehat{(w\underline{\mathbf{v}})}_n^{\sim 0} = \bar{w}\hat{\underline{\mathbf{v}}}_n^{\sim 0} + \widehat{(w'\underline{\mathbf{v}}')}_n^{\sim 0}. \tag{3.94}$$

These allow term **(I)** to be written as

$$\begin{aligned}
\textbf{(I)} &= \hat{\underline{\mathbf{v}}}_n^{\sim 0} \cdot \left[-\nabla \cdot (\bar{\underline{\mathbf{v}}}\ \hat{\underline{\mathbf{v}}}_n^{\sim 0}) - \frac{\partial}{\partial z}(\bar{w}\hat{\underline{\mathbf{v}}}_n^{\sim 0})\right] + \hat{\underline{\mathbf{v}}}_n^{\sim 0} \cdot \left[-\nabla \cdot \widehat{(\underline{\mathbf{v}}'\underline{\mathbf{v}}')}_n^{\sim 0} - \frac{\partial}{\partial z}\widehat{(w'\underline{\mathbf{v}}')}_n^{\sim 0}\right]\\
&= 2^{-j_2}\left\{-\nabla \cdot \left(\bar{\underline{\mathbf{v}}}K^L\right) - \frac{\partial}{\partial z}\left(\bar{w}K^L\right) + \bar{\underline{\mathbf{v}}} \cdot \left[-\nabla \cdot \overline{(\underline{\mathbf{v}}'\underline{\mathbf{v}}')} - \frac{\partial}{\partial z}\overline{(w'\underline{\mathbf{v}}')}\right]\right\}\\
&= 2^{-j_2}\left\{-\nabla \cdot \left(\bar{\underline{\mathbf{v}}}K^L\right) - \frac{\partial}{\partial z}\left(\bar{w}K^L\right) + \bar{\underline{\mathbf{v}}} \cdot \nabla_3 \cdot \underline{\underline{\mathbf{T}}}\right\}, &(3.95)
\end{aligned}$$

where

$$\nabla_3 = \underline{\mathbf{i}}\frac{\partial}{\partial x} + \underline{\mathbf{j}}\frac{\partial}{\partial y} + \underline{\mathbf{k}}\frac{\partial}{\partial z},$$

and

$$\underline{\underline{\mathbf{T}}} = \begin{bmatrix} -\overline{(u'u')} & -\overline{(u'v')} & -\overline{(u'w')} \\ -\overline{(v'u')} & -\overline{(v'v')} & -\overline{(v'w')} \\ -\overline{(w'u')} & -\overline{(w'v')} & -\overline{(w'w')} \end{bmatrix}. \tag{3.96}$$

Term **(II)** is treated as that in (3.27). As stated before, it is equal to $\Delta_h Q_{PL} + \Delta_z Q_{PL} - b_n^L$, which is, in the present setting,

$$\textbf{(II)} = 2^{-j_0} \left\{ -\frac{1}{\rho_0} \nabla \cdot (\bar{P} \bar{\underline{\mathbf{v}}}) - \frac{1}{\rho_0} \frac{\partial}{\partial z} (\bar{P} \bar{w}) - \frac{g}{\rho_0} \bar{w} \bar{\rho}. \right\} \tag{3.97}$$

Substitute **(I)** and **(II)** back to Eq. (3.83). Considering that the left hand side is now $2^{-j_0} \bar{\underline{\mathbf{v}}} \cdot \overline{\left(\frac{\partial \underline{\mathbf{v}}}{\partial t}\right)}$, we have, with the common factor 2^{-j_0} cancelled out,

$$\begin{aligned} \bar{\underline{\mathbf{v}}} \cdot \overline{\left(\frac{\partial \underline{\mathbf{v}}}{\partial t}\right)} \;=\; & -\nabla \cdot \left(\bar{\underline{\mathbf{v}}} K^L\right) - \frac{\partial}{\partial z} \left(\bar{w} K^L\right) \\ & -\frac{1}{\rho_0} \nabla \cdot (\bar{P} \bar{\underline{\mathbf{v}}}) - \frac{1}{\rho_0} \frac{\partial}{\partial z} (\bar{P} \bar{w}) - \frac{g}{\rho_0} \bar{w} \bar{\rho}. \\ & +\bar{\underline{\mathbf{v}}} \cdot \nabla_3 \cdot \underline{\underline{\mathbf{T}}}. \end{aligned} \tag{3.98}$$

This is exactly what Harrison and Robinson (1978) have obtained for their large-scale kinetic energy, with $\underline{\underline{\mathbf{T}}}$ the Reynolds stress tensor in their formulation.

For the meso-scale window, the counterpart of Eq. (3.83) is

$$\hat{\underline{\mathbf{v}}}_n^{\sim 1} \cdot \widehat{\left(\frac{\partial \underline{\mathbf{v}}}{\partial t}\right)}_n^{\sim 1} = -\hat{\underline{\mathbf{v}}}_n^{\sim 1} \cdot \nabla \cdot \widehat{(\underline{\mathbf{v}}\ \underline{\mathbf{v}})}_n^{\sim 1} - \hat{\underline{\mathbf{v}}}_n^{\sim 1} \cdot \frac{\partial}{\partial z} \widehat{(w \underline{\mathbf{v}})}_n^{\sim 1} - \hat{\underline{\mathbf{v}}}_n^{\sim 1} \cdot \nabla \frac{\hat{P}_n^{\sim 1}}{\rho_0} \tag{3.99}$$

In order to make it comparable to the classical eddy KE equation, just $j_0 = 0$ and periodic extension are not enough, as now there no longer exists for field p a linear relation between $\hat{p}_n^{\sim 1}$ and p'. We have to marginalize (3.99) to the physical space to fulfill this mission. In this particular case, the marginalization equality (2.70) is simply

$$\mathcal{M}_n \hat{p}_n^{\sim 1} \hat{q}_n^{\sim 1} = \overline{p' q'}, \qquad \forall\ p, q \in V_{\varrho, j_2}, \tag{3.100}$$

since here the deviation operation (prime) and the meso-scale synthesis operator are identical. Marginalization of (3.99) yields

$$\overline{\underline{\mathbf{v}}' \cdot \frac{\partial \underline{\mathbf{v}}'}{\partial t}} = -\underbrace{\overline{\underline{\mathbf{v}}' \cdot \nabla \cdot (\underline{\mathbf{v}}\ \underline{\mathbf{v}})'}}_{\textbf{(I}'\textbf{)}} - \underbrace{\overline{\underline{\mathbf{v}}' \cdot \frac{\partial}{\partial z} (w \underline{\mathbf{v}})'}}_{\textbf{(II}'\textbf{)}} - \underbrace{\overline{\underline{\mathbf{v}}' \cdot \nabla \left(\frac{P'}{\rho_0}\right)}}_{\textbf{(III}'\textbf{)}}. \tag{3.101}$$

It is easy to show, as we did before,

$$(\mathbf{III}') = \nabla \cdot \overline{\left(\underline{\mathbf{v}}' \frac{P'}{\rho_0}\right)} + \frac{\partial}{\partial z}\overline{\left(w' \frac{P'}{\rho_0}\right)} + \frac{g}{\rho_0}\overline{w'\rho'}. \tag{3.102}$$

The other two terms on the right hand side are

$$\begin{aligned}
(\mathbf{I}') &= \overline{\underline{\mathbf{v}}' \cdot \nabla \cdot (\underline{\mathbf{v}}\ \underline{\mathbf{v}} - \overline{\underline{\mathbf{v}}\ \underline{\mathbf{v}}})} \\
&= \overline{\underline{\mathbf{v}}' \cdot \nabla \cdot (\bar{\underline{\mathbf{v}}}\ \underline{\mathbf{v}}' + \underline{\mathbf{v}}'\bar{\underline{\mathbf{v}}} + \underline{\mathbf{v}}'\underline{\mathbf{v}}')} \\
&= \nabla \cdot \overline{\left(\underline{\mathbf{v}}\frac{\underline{\mathbf{v}}' \cdot \underline{\mathbf{v}}'}{2}\right)} + \overline{\underline{\mathbf{v}}' \cdot \nabla \cdot \underline{\mathbf{v}}'\bar{\underline{\mathbf{v}}}} \\
&= \nabla \cdot \overline{\left(\underline{\mathbf{v}}\frac{\underline{\mathbf{v}}' \cdot \underline{\mathbf{v}}'}{2}\right)} + \overline{(\nabla \cdot \underline{\mathbf{v}}')\underline{\mathbf{v}}'} \cdot \bar{\underline{\mathbf{v}}} + \overline{\underline{\mathbf{v}}'\underline{\mathbf{v}}'} : \nabla\bar{\underline{\mathbf{v}}},
\end{aligned} \tag{3.103}$$

and likewise,

$$(\mathbf{II}') = \frac{\partial}{\partial z}\overline{\left(w\frac{\underline{\mathbf{v}}' \cdot \underline{\mathbf{v}}'}{2}\right)} + \overline{(\frac{\partial w'}{\partial z})\underline{\mathbf{v}}'} \cdot \bar{\underline{\mathbf{v}}} + \overline{\underline{\mathbf{v}}'w'} \cdot \frac{\partial \bar{\underline{\mathbf{v}}}}{\partial z}. \tag{3.104}$$

These two terms sum up to

$$\begin{aligned}
(\mathbf{I}') + (\mathbf{II}') \quad = \quad & \nabla \cdot \overline{\left(\underline{\mathbf{v}}\frac{\underline{\mathbf{v}}' \cdot \underline{\mathbf{v}}'}{2}\right)} + \frac{\partial}{\partial z}\overline{\left(w\frac{\underline{\mathbf{v}}' \cdot \underline{\mathbf{v}}'}{2}\right)} \\
& + \overline{\underline{\mathbf{v}}'\underline{\mathbf{v}}'} : \nabla\bar{\underline{\mathbf{v}}} + \overline{\underline{\mathbf{v}}'w'} \cdot \frac{\partial \bar{\underline{\mathbf{v}}}}{\partial z},
\end{aligned} \tag{3.105}$$

where the fact (3.55) has been used. The meso-scale kinetic energy equation hence becomes

$$\begin{aligned}
\overline{\underline{\mathbf{v}}' \cdot \frac{\partial \underline{\mathbf{v}}'}{\partial t}} \quad = \quad & -\nabla \cdot \overline{\left(\underline{\mathbf{v}}\frac{\underline{\mathbf{v}}' \cdot \underline{\mathbf{v}}'}{2}\right)} - \frac{\partial}{\partial z}\overline{\left(w\frac{\underline{\mathbf{v}}' \cdot \underline{\mathbf{v}}'}{2}\right)} \\
& -\nabla \cdot \overline{\left(\underline{\mathbf{v}}' \frac{P'}{\rho_0}\right)} - \frac{\partial}{\partial z}\overline{\left(w' \frac{P'}{\rho_0}\right)} - \frac{g}{\rho_0}\overline{w'\rho'} \\
& -\overline{\underline{\mathbf{v}}'\underline{\mathbf{v}}'} : \nabla\bar{\underline{\mathbf{v}}} - \overline{\underline{\mathbf{v}}'w'} \cdot \frac{\partial \bar{\underline{\mathbf{v}}}}{\partial z}.
\end{aligned} \tag{3.106}$$

Again, this is exactly the eddy KE equation obtained by Harrison and Robinson (1978).

With the same expansion and following the same procedure, a mean and an eddy APE equations in the classical sense can also been recovered. Therefore, in a classical Reynolds formulation, the mean energetics are just those of a very particular case (a two-scale decomposition with a zero j_0 and periodic extension) in the MS-EVA framework, and the eddy energetics may be obtained by marginalizing the corresponding MS-EVA meso-scale energetics for that particular case. In this sense, MS-EVA is a generalization of the classical formalism.

We have explained in Chapter 1 that the meaning of *analysis* in MS-EVA is two-fold: It means not only a multiscale analysis in the phase space, but also a process analysis in the physical space. The above comparison is made actually in the context of multiscale decomposition. It should be pointed out that, even in the above particular situation, MS-EVA and the classical Reynolds formalism are different in terms of analysis in the physical space. The classical formalism attributes all processes, except the mean flow advection, to the Reynolds stress without distinguishing the roles of rotation, vertical shear of horizontal flow, and eddy transport. In this sense, even at $j_0 = 0$ with a periodic extension, MS-EVA is expected to do a better job in process identification.

3.6 Transfer and transport

Process identification with the MS-EVA relies on a clear classification of the processes represented by the energetic terms. According to previous researchers, these processes can be categorized into four classes: transport, transfer, conversion, and dissipation (e.g., Harrison and Robinson, 1978; Fournier, 1999; Lesieur, 1990). Conversion processes occur between the two forms of energy and have been introduced before (cf. p. 120). Dissipation is important but is beyond the scope of this thesis. In this section, we discuss the transfer and transport processes only.

We have mentioned that, in the multiscale energy equations, terms symbolized with T represent the rate of energy transfer through phase space. A process responsible for this task is called a **transfer process**. Its counterpart in the physical space, distinguished symbolically with Q among the remaining terms, is the transport process. Strictly speaking, by **a transport process** we mean a process which vanishes under integration over a closed spatial domain. As we have seen before, there could be some ambiguity with the transfer-transport separation in the derivation of the APE equations. A purpose of this section is, therefore, to fix this ambiguity problem through a physical clarification of these two processes.

3.6.1 Perfect transfer and transfer-transport separation

In the class of transfer processes, there is a subclass which accounts for the phase space energy redistribution, but makes no contribution to the total energy evolution. These kind of processes are called **perfect transfers** (or perfect transfer processes). We have introduced the concept in deriving the large-scale kinetic energy equation (p. 123).

Mathematically, a perfect transfer is a process represented by a term $T_{EW,n}$, E standing for K or A, such that

$$\sum_W \mathcal{M}_n T_{EW,n} = 0, \tag{3.107}$$

where the window index W belongs to $\{L, M, S\}$, and n runs over the sampling space on the time duration. As the sampling space corresponds to time in our formalism, one may say that a perfect transfer process has a vanishing property in the time direction, in comparison to the space vanishing property of the transport process.

We now claim, with the definition (3.107), that the process represented by the advection-related term in the potential energy equation derivation (i.e., the (AD) on p. 127) can always be faithfully decomposed as a combination of a transport process and a perfect transfer process. Before showing how, we need some knowledge of the structure of the transfer terms (T terms) in the energetic equations.

Observe that all these T-terms are made of a more basic class of **transfer functions**, which are in the form

$$T(w,n) = \hat{\mathcal{Z}}_n^{\sim w} \cdot \widehat{(pq)}_n^{\sim w}, \tag{3.108}$$

for any functions $\mathcal{Z}$, p, and q in V_{ϱ,j_2}, with $w \in \{0,1,2\}$ corresponding to the large-, meso-, and submeso-scale window, respectively.[6] This is a direct consequence of the quadratic nonlinear interactions in the PE model equations. (We may re-define the transfer process as something represented by an energetic term which has a form as a transfer function.) Marginalization of $T(w,n)$, followed by a summation over windows $w \in \{0,1,2\}$, gives

$$\begin{aligned}\sum_w \mathcal{M}_n T(w,n) &= \sum_w \mathcal{M}_n \hat{\mathcal{Z}}_n^{\sim w} \widehat{(pq)}_n^{\sim w} \\ &= \int_0^1 \sum_w \mathcal{Z}^{\sim w} (pq)^{\sim w} \, dt.\end{aligned}$$

As the large-, meso-, and sub-meso-scale subspaces are mutually orthogonal, this is simply

$$\sum_w \mathcal{M}_n T(w,n) = \int_0^1 \mathcal{Z} \; (pq) \; dt,$$

[6] As a superscript, w denotes a scale window. We will distinguish this w from the notation for vertical velocity by reminding the reader of its scale window identity each time it is used.

or, in a more concise form,

$$\boxed{\sum_w \mathcal{M}_n T(w,n) = \overline{\mathcal{Z}\ (pq)}.} \tag{3.109}$$

With this equality, we are now able to give an explanation of why the transport in the APE equations should be written in a form like (3.38) and (3.39) (p. 127).

For simplicity, consider the 2-D, non-diffusive version of the density equation

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot (\underline{\mathbf{v}}\rho) \equiv -\nabla \cdot \underline{\mathbf{Q}}_\rho, \tag{3.110}$$

where $\underline{\mathbf{Q}}_\rho$ stands for the flux of density (with a nondivergent horizontal flow). Take a transform for the window w and location n, then multiply it by $\widehat{\rho}_n^{\sim w}$. We get, on the right hand side, the advection-related term

$$(\text{AD}) = -\widehat{\rho}_n^{\sim w}\ \nabla \cdot \widehat{(\underline{\mathbf{Q}}_\rho)}_n^{\sim w}. \tag{3.111}$$

The flux of $A = \frac{1}{2}\rho^2$ (the pre-multiplier c dropped for clarity) in the physical space is in a form of $\rho\underline{\mathbf{Q}}_\rho = \rho\rho\underline{\mathbf{v}}$. In the phase space, only quadratic quantities have physical interpretation by the property of marginalization, so the flux of A for window w, which has its root in (AD), should be in a form like

$$\kappa_1 \widehat{\rho}_n^{\sim w} \widehat{(\underline{\mathbf{Q}}_\rho)}_n^{\sim w} + \kappa_2 \widehat{(\rho^2)}_n^{\sim w} \widehat{\underline{\mathbf{v}}}_n^{\sim w},$$

with κ_1 and κ_2 some constants to be determined. We choose $\kappa_1 = \kappa_2 \equiv \kappa$ to reduce the number of unkonws to one.[7] In this way, (AD) is decomposed as

$$\begin{aligned}
(\text{AD}) \quad &= \quad -\nabla \cdot \left(\kappa \widehat{\rho}_n^{\sim w} \widehat{(\underline{\mathbf{Q}}_\rho)}_n^{\sim w} + \kappa \widehat{(\rho^2)}_n^{\sim w} \widehat{\underline{\mathbf{v}}}_n^{\sim w} \right) \\
&\quad\quad - \left[\widehat{\rho}_n^{\sim w} \nabla \cdot \widehat{(\underline{\mathbf{Q}}_\rho)}_n^{\sim w} - \nabla \cdot \left(\kappa \widehat{\rho}_n^{\sim w} \widehat{(\underline{\mathbf{Q}}_\rho)}_n^{\sim w} + \kappa \widehat{(\rho^2)}_n^{\sim w} \widehat{\underline{\mathbf{v}}}_n^{\sim w} \right) \right] \\
&\equiv \quad -\nabla \cdot \left(\kappa \widehat{\rho}_n^{\sim w} \widehat{(\underline{\mathbf{Q}}_\rho)}_n^{\sim w} + \kappa \widehat{(\rho^2)}_n^{\sim w} \widehat{\underline{\mathbf{v}}}_n^{\sim w} \right) + T_n^w.
\end{aligned} \tag{3.112}$$

[7]This relation is chosen for the time being in analogy to the Leibniz differentiation rule. It is supported by a later comparison of the transport thus-obtained to the classical energetics, and the validation result of the MS-EVA in Chapter 4 with the Eady instability problem. (Different κ_1 and κ_2 will lead to an asymmetric vertical structure which is not expected from the analytical solution.) A theoretical justification, however, is left for future work.

We want to make T_n^w represent a perfect transform process. It is then expected that

$$\sum_w \mathcal{M}_n T_n^w = 0.$$

By equality (3.109), this is to say (Recall the definition of $\underline{\mathbf{Q}}_\rho$ and the nondivergence of $\underline{\mathbf{v}}$.)

$$\int_0^1 \left[\rho\nabla\cdot(\rho\underline{\mathbf{v}}) - \kappa\nabla\cdot(\rho^2\underline{\mathbf{v}}) - \kappa\nabla\cdot(\rho^2\underline{\mathbf{v}})\right]\ dt$$
$$= (2\kappa - \frac{1}{2})\ \nabla\cdot\overline{(\rho^2\underline{\mathbf{v}})} = 0.$$

Obviously, κ can only have one value, $\frac{1}{4}$, in order for $\sum_w \mathcal{M}_n T_n^w = 0$ to hold for any scalar ρ and nondivergent flow $\underline{\mathbf{v}}$. The decomposition (3.112) is therefore unambiguously determined.

The above argument can be extended to a 3-D case straightforwardly. The result is similar and the κ is again unambiguously $\frac{1}{4}$. This explains why we chose the way as (3.38) and (3.39 to separate the transport and transfer processes for a multiscale APE equation, and why the chosen separation is unique in this sense.

Recall that in formulating the APE transfers, we did not just stop at this transport-transfer separation. Rather, we added to each of the transfers already obtained an extra term obtained by taking advantage of the continuity equation. The consequence of this addition is that, every T term of the APE equation, as well as the sum of all the T terms, now has a correspondence to a perfect transfer process. In the next subsection, we will give a proof of this remarkable property.

3.6.2 The T terms as perfect transfer processes

For each scale window, we have collected six energy transfer terms from the MS-EVA energetic equations. They are, symbolically,

$$T_{K^W,\zeta},\quad T_{K^W,w\partial_z},\quad T_{K^W,\partial_z v}$$
$$T_{A^W,\partial_x\rho},\quad T_{A^W,\partial_y\rho}, T_{A^W,\partial_z\rho},$$

with $W = L, M, S$ standing for the large-, meso-, or sub-meso-scale window. Using equality (3.109), we now show that these six T-terms actually represent five perfect transfer processes.

First look at $T_{A^W,\partial_x\rho}$ (note that the dependence on n is suppressed). Its expression is seen in the $T_{A^W,\partial_h\rho}$ of Eqs. (3.50), (3.71), and (3.76), with the operator $\nabla\cdot$ replaced by $\partial/\partial x$. Make the summation over $W = \{L, M, S\}$, followed by a marginalization over n. We have

$$\begin{aligned}
\sum_W \mathcal{M}_n T_{A^W,\partial_x\rho} &= \sum_w \mathcal{M}_n \left\{ -c\widehat{\rho}_n^{\sim w} \frac{\partial}{\partial x} \widehat{(u\rho)}_n^{\sim w} + \frac{\partial}{\partial x}\left[\frac{1}{4}c\widehat{\rho}_n^{\sim w}\widehat{(u\rho)}_n^{\sim w} + \frac{1}{4}c\widehat{(\rho^2)}_n^{\sim w}\widehat{u}_n^{\sim w}\right]\right. \\
&\quad \left. +\frac{1}{4}c\left[\widehat{\rho}_n^{\sim w}\widehat{\left(\rho\frac{\partial u}{\partial x}\right)}_n^{\sim w} + \widehat{(\rho^2)}_n^{\sim w}\frac{\partial \widehat{u}_n^{\sim w}}{\partial x}\right]\right\} \\
&= \int_0^1 \left\{ -c\rho\frac{\partial}{\partial x}(u\rho) + \frac{\partial}{\partial x}\left[\frac{1}{4}c\rho(u\rho) + \frac{1}{4}c\rho^2 u\right] + \frac{1}{4}c\rho\left(\frac{\partial u}{\partial x}\rho\right) + \frac{1}{4}c\rho^2\frac{\partial u}{\partial x}\right\}\, dt \\
&= \int_0^1 \left\{ -c\rho\frac{\partial}{\partial x}(u\rho) + \frac{\partial}{\partial x}(u\frac{c\rho^2}{2}) + \frac{c\rho^2}{2}\frac{\partial u}{\partial x}\right\}\, dt \\
&= 0.
\end{aligned}$$

Likewise, we also have

$$\sum_W \mathcal{M}_n T_{A^W,\partial_y\rho} = 0.$$

In the transfer due to $\partial_z\rho$, $T_{A^W,\partial_z\rho}$, the additional

$$\frac{1}{4}\left[\widehat{\rho}_n^{\sim w}\widehat{(w\rho)}_n^{\sim w} + \widehat{(\rho^2)}_n^{\sim w}\widehat{w}_n^{\sim w}\right]\frac{\partial c}{\partial z}$$

happens to cancel out the extra term in $\sum_W \mathcal{M}_n T_{A^W,\partial_z\rho}$ brought up by the z-dependence of c or N^2. The final result is, again,

$$\sum_W \mathcal{M}_n T_{A^W,\partial_z\rho} = 0.$$

For the kinetic energy equation, marginalization of the T-terms results in

$$\begin{aligned}
\sum_W \mathcal{M}_n T_{K^W,\zeta} &= -\int_0^1 \underline{\mathbf{v}}\cdot\underline{\mathbf{k}}\wedge\zeta\underline{\mathbf{v}}\, dt = 0; \\
\sum_W \mathcal{M}_n T_{K^W,w\partial_z} &= -\int_0^1 w\frac{\partial K}{\partial z}\, dt; \\
\sum_W \mathcal{M}_n T_{K^W,\partial_z v} &= -\int_0^1 \underline{\mathbf{v}}\cdot\underline{\mathbf{k}}\wedge(\underline{\zeta}^* w)\, dt \\
&= \int_0^1 \underline{\mathbf{v}}\cdot\underline{\mathbf{k}}\wedge\left\{-w\frac{\partial v}{\partial z}\underline{\mathbf{i}} + w\frac{\partial u}{\partial z}\underline{\mathbf{j}}\right\}\, dt \\
&= -\int_0^1 w\frac{\partial K}{\partial z}\, dt.
\end{aligned}$$

Clearly, $T_{K^W,\zeta}$ represents a perfect transfer process. $T_{K^W,w\partial_z}$ or $T_{K^W,\partial_z v}$ alone usually does not qualify for this category (unless $\overline{w\frac{\partial K}{\partial z}} = 0$), but their sum $T_{K^W,w\partial_z} + T_{K^W,\partial_z v}$ does. We write them separately because we need to use their individual expressions later.

Besides the six T-terms, there is another transfer term, TS_{A^W}, in the multiscale APE equations, which is related to the apparent source/sink in Eq. (3.18). This transfer is *imperfect*, as it generally does not sum to zero over the windows and locations in the sampling space. In fact,

$$\begin{aligned}\sum_W \mathcal{M}_n TS_{A^W} &= \sum_w \mathcal{M}_n \left\{ \frac{1}{2} \widehat{\rho}_n^{\sim w} \widehat{(w\rho)}_n^{\sim w} \frac{\partial c}{\partial z} \right\} \\ &= -\int_0^1 wA \frac{\partial}{\partial z} \left(\log N^2 \right) \, dt,\end{aligned}$$

which is the apparent source/sink term in Eq. (3.18) averaged over the time duration. It does not vanish, should there be a stationary vertical shear in the background density profile.

3.6.3 More about the transport processes

Transport processes are represented by these "Q-terms" in the multiscale energy equations. In this subsection we want to gain more understanding about these terms' physical significance. For each Q-term, the horizontal and vertical components are expressed in the same form. As a result, we only need to look at the former, and the physical interpretation thus obtained applies equally to the latter.

Let $\underline{\mathbf{Q}}_A^w$, $\underline{\mathbf{Q}}_K^w$, $\underline{\mathbf{Q}}_P^w$ denote, respectively, the fluxes (horizontal) of APE, KE, and pressure with scale window w, i.e.,

$$\underline{\mathbf{Q}}_A^w = \frac{1}{4} c \widehat{\rho}_n^{\sim w} \widehat{(\rho \underline{\mathbf{v}})}_n^{\sim w} + \frac{1}{4} c \widehat{(\rho^2)}_n^{\sim w} \widehat{\underline{\mathbf{v}}}_n^{\sim w}, \tag{3.113}$$

$$\underline{\mathbf{Q}}_P^w = \frac{1}{\rho_0} \widehat{\underline{\mathbf{v}}}_n^{\sim w} \widehat{P}_n^{\sim w}, \tag{3.114}$$

$$\underline{\mathbf{Q}}_K^w = \widehat{\underline{\mathbf{v}}}_n^{\sim w} \widehat{K}_n^{\sim w}, \tag{3.115}$$

the transport terms are then just

$$-\nabla \cdot \underline{\mathbf{Q}}_A^w, \quad -\nabla \cdot \underline{\mathbf{Q}}_P^w, \quad -\nabla \cdot \underline{\mathbf{Q}}_K^w. \tag{3.116}$$

Here again the location indices have been suppressed for simplicity. Throughout this subsection, only two scales (corresponding to windows $w = 0, 1$) are considered.

Of the three Q-terms, the one with an apparent physical interpretation is $\nabla \cdot \underline{\mathbf{Q}}_P^w$. By marginalization it yields $\nabla \cdot \frac{1}{\rho_0} \overline{(\underline{\mathbf{v}}^{\sim w} P^{\sim w})}$. In the particular case when $j_0 = 0$ and a periodic extension is used, this becomes, for windows $w = 0$ and $w = 1$ (cf. Theorem 2.3.8),

$$\frac{1}{\rho_0}\overline{\underline{\mathbf{v}}^{\sim 0} P^{\sim 0}} = \frac{1}{\rho_0}\bar{\underline{\mathbf{v}}}\bar{P}, \qquad \frac{1}{\rho_0}\overline{\underline{\mathbf{v}}^{\sim 1} P^{\sim 1}} = \frac{1}{\rho_0}\overline{\underline{\mathbf{v}}' P'},$$

which apparently represent the mean (former) and eddy (latter) pressure working rates.

The interpretation of the other two processes, $\nabla \cdot \underline{\mathbf{Q}}_A^w$ or $\nabla \cdot \underline{\mathbf{Q}}_K^w$, however, is far more complicated. Look at $\nabla \cdot \underline{\mathbf{Q}}_K^w$ first. The complexity of $\nabla \cdot \underline{\mathbf{Q}}_K^w$ over $\nabla \cdot \underline{\mathbf{Q}}_P^w$ stems from the fact that $\widehat{K}_n^{\sim w}$ is by no means equal to the energy on scale window w. To see this, notice that $\underline{\mathbf{v}} = \underline{\mathbf{v}}^{\sim 0} + \underline{\mathbf{v}}^{\sim 1}$, which gives

$$\begin{aligned} K &= \frac{1}{2}\underline{\mathbf{v}} \cdot \underline{\mathbf{v}} = \frac{1}{2}(\underline{\mathbf{v}}^{\sim 0} + \underline{\mathbf{v}}^{\sim 1}) \cdot (\underline{\mathbf{v}}^{\sim 0} + \underline{\mathbf{v}}^{\sim 1}) \\ &= \frac{1}{2}\underline{\mathbf{v}}^{\sim 0} \cdot \underline{\mathbf{v}}^{\sim 0} + \underline{\mathbf{v}}^{\sim 1} \cdot \underline{\mathbf{v}}^{\sim 0} + \frac{1}{2}\underline{\mathbf{v}}^{\sim 1} \cdot \underline{\mathbf{v}}^{\sim 1}. \end{aligned} \tag{3.117}$$

Taking a large-scale transform of it, we get

$$\widehat{K}_n^{\sim 0} = \frac{1}{2}(\widehat{\underline{\mathbf{v}}^{\sim 0} \cdot \underline{\mathbf{v}}^{\sim 0}})_n^{\sim 0} + (\widehat{\underline{\mathbf{v}}^{\sim 1} \cdot \underline{\mathbf{v}}^{\sim 0}})_n^{\sim 0} + \frac{1}{2}(\widehat{\underline{\mathbf{v}}^{\sim 1} \cdot \underline{\mathbf{v}}^{\sim 1}})_n^{\sim 0}. \tag{3.118}$$

If $j_0 = 0$, and a periodic extension is used, we know from Theorem 2.3.8 that the large-scale $\underline{\mathbf{v}}^{\sim 0}$ is just the mean over time $\bar{\underline{\mathbf{v}}}$ (constant in time) and is equal to $2^{j_2/2}\hat{\underline{\mathbf{v}}}_n^{\sim 0}$. Thus

$$\begin{aligned} (\widehat{\underline{\mathbf{v}}^{\sim 0} \cdot \underline{\mathbf{v}}^{\sim 0}})_n^{\sim 0} &= \bar{\underline{\mathbf{v}}} \cdot (\widehat{\underline{\mathbf{v}}^{\sim 0}})_n^{\sim 0} = 2^{j_2/2}\hat{\underline{\mathbf{v}}}_n^{\sim 0} \cdot \hat{\underline{\mathbf{v}}}_n^{\sim 0} = 2K_n^L \cdot 2^{j_2/2} \\ (\widehat{\underline{\mathbf{v}}^{\sim 1} \cdot \underline{\mathbf{v}}^{\sim 0}})_n^{\sim 0} &= \bar{\underline{\mathbf{v}}} \cdot (\widehat{\underline{\mathbf{v}}^{\sim 1}})_n^{\sim 0} = 0 \\ (\widehat{\underline{\mathbf{v}}^{\sim 1} \cdot \underline{\mathbf{v}}^{\sim 1}})_n^{\sim 0} &= 2^{-j_2/2}\overline{\underline{\mathbf{v}}' \cdot \underline{\mathbf{v}}'} = 2^{-j_2/2} \sum_{2^{-j_2}n \in [0,1)} \hat{\underline{\mathbf{v}}}_n^{\sim 1} \cdot \hat{\underline{\mathbf{v}}}_n^{\sim 1} \qquad \text{(marginalization)} \\ &= 2\left[K_n^M\right]^{\text{avg}} \cdot 2^{j_2/2}, \end{aligned}$$

where

$$\left[K_n^M\right]^{\text{avg}} = \frac{1}{2^{j_2}} \sum_{2^{-j_2}n \in [0,1)} K_n^M$$

is the meso-scale energy $K_n^M = \frac{1}{2}\hat{\underline{\mathbf{v}}}_n^{\sim 1} \cdot \hat{\underline{\mathbf{v}}}_n^{\sim 1}$ averaged over all locations. So

$$\widehat{K}_n^{\sim 0} = 2^{j_2/2}K_n^L + 2^{j_2/2}\left[K_n^M\right]^{\text{avg}}, \tag{3.119}$$

and

$$\hat{\underline{\mathbf{v}}}_n^{\sim 0}\widehat{K}_n^{\sim 0} = 2^{j_2/2}\hat{\underline{\mathbf{v}}}_n^{\sim 0}K_n^L + 2^{j_2/2}\hat{\underline{\mathbf{v}}}_n^{\sim 0}\left[K_n^M\right]^{\text{avg}}$$

$$= \quad \underline{\bar{\mathbf{v}}} K_n^L + \underline{\bar{\mathbf{v}}} \left[K_n^M \right]^{\text{avg}}. \tag{3.120}$$

This is to say, $\widehat{K}_n^{\sim 0}$ is not just related to the large-scale energy. It involves contributions from other scales also. In this particular case, those contributions are quantified in an amount of the averaged meso-scale energy over all locations in the sampling space. In general, Eq. (3.118) cannot be reduced to Eq. (3.119). The transport process $\nabla \cdot \underline{\mathbf{Q}}_{\mathbf{K}}^{\mathbf{0}}$ thereby cannot be interpreted simply as transport of the large-scale K, in contrast to that of $\nabla \cdot \underline{\mathbf{Q}}_P^0$.

For the meso-scale energetics, the counterpart of Eq. (3.118)

$$\widehat{K}_n^{\sim 1} = \frac{1}{2} (\widehat{\underline{\mathbf{v}}^{\sim 0} \cdot \underline{\mathbf{v}}^{\sim 0}})_n^{\sim 1} + (\widehat{\underline{\mathbf{v}}^{\sim 1} \cdot \underline{\mathbf{v}}^{\sim 0}})_n^{\sim 1} + \frac{1}{2} (\widehat{\underline{\mathbf{v}}^{\sim 1} \cdot \underline{\mathbf{v}}^{\sim 1}})_n^{\sim 1} \tag{3.121}$$

gives much more complex a flux form. Even in the simplest case when $j_0 = 0$, and a periodic extension is adopted, the flux can only be simplified to

$$\underline{\bar{\mathbf{v}}} \cdot \widehat{\underline{\mathbf{v}}}_n^{\sim 1} \widehat{\underline{\mathbf{v}}}_n^{\sim 1} + \frac{1}{2} \widehat{\underline{\mathbf{v}}}_n^{\sim 1} (\widehat{\underline{\mathbf{v}}' \cdot \underline{\mathbf{v}}'})_n^{\sim 1}, \tag{3.122}$$

which cannot be written in a form similar to $\widehat{\underline{\mathbf{v}}}_n^{\sim 0} \widehat{K}_n^{\sim 0}$, let alone $\underline{\mathbf{Q}}_P^w$.

The interpretation for $\nabla \cdot \underline{\mathbf{Q}}_A^w$ is also complex, By the marginalization property,

$$\begin{aligned}
\mathcal{M}_n \underline{\mathbf{Q}}_A^w &= \mathcal{M}_n \left[\frac{1}{4} c \widehat{\rho}_n^{\sim w} \widehat{(\rho \underline{\mathbf{v}})}_n^{\sim w} + \frac{1}{4} c \widehat{(\rho^2)}_n^{\sim w} \widehat{\underline{\mathbf{v}}}_n^{\sim w} \right] \\
&= \frac{c}{4} \overline{\rho^{\sim w} (\rho \underline{\mathbf{v}})^{\sim w}} + \frac{c}{4} \overline{(\rho^2)^{\sim w} \underline{\mathbf{v}}^{\sim w}}.
\end{aligned}$$

If $j_0 = 0$, and a periodic extension is used, then

$$\rho^{\sim 0} = \bar{\rho}, \quad \rho^{\sim 1} = \rho',$$

So when $w = 0$,

$$\begin{aligned}
\mathcal{M}_n \underline{\mathbf{Q}}_A^0 &= \frac{1}{4} c \overline{\bar{\rho} \overline{(\rho \underline{\mathbf{v}})}} + \frac{1}{4} c \overline{\overline{\rho^2} \underline{\bar{\mathbf{v}}}} = \frac{1}{4} c \bar{\rho} \overline{(\rho \underline{\mathbf{v}})} + \frac{1}{4} c \overline{\rho^2} \underline{\bar{\mathbf{v}}} \\
&= \frac{1}{4} c \bar{\rho} \ (\bar{\rho} \underline{\bar{\mathbf{v}}} + \overline{\rho' \underline{\mathbf{v}}'}) + \frac{1}{4} c (\bar{\rho}^2 + \overline{\rho'^2}) \underline{\bar{\mathbf{v}}} \\
&= \underline{\bar{\mathbf{v}}} \ (\frac{1}{2} c \bar{\rho}^2) + \frac{1}{4} c \ \left(\overline{\underline{\mathbf{v}}' \rho'} \ \bar{\rho} + \overline{\rho'^2} \underline{\bar{\mathbf{v}}} \right).
\end{aligned} \tag{3.123}$$

The first part is the flux of the large-scale APE by the large-scale flow (recall $\bar{\rho} = 2^{j_2/2} \widehat{\rho}_n^{\sim 0}$ in this case). Again, there exists an extra term which describes contributions from the other scale window. Similarly, when $w = 1$, it can be shown that the marginalization of $\underline{\mathbf{Q}}_A^w$ can be written as something like the flux of $\frac{1}{2}c\,\rho'^2$ (not the eddy APE, however), $\frac{1}{2}c\,\overline{\underline{\mathbf{v}}\rho'^2}$, plus a residual term. The details of the residual are not shown here. Generally speaking, a flux to a designated spot on a designated scale window involves processes not just for that window. It is a complex combination of the contributions from all scales over the spectrum.

3.7 Interaction analysis

Different from the classical energetics, a localized energy transfer involves not only interactions between scales, but also interactions between locations in the sampling space. We have already seen this in the definition of perfect transfer processes (3.107). A schematic is also seen in Fig. 3.1. The addition of the sampling space interaction compounds substantially the transfer

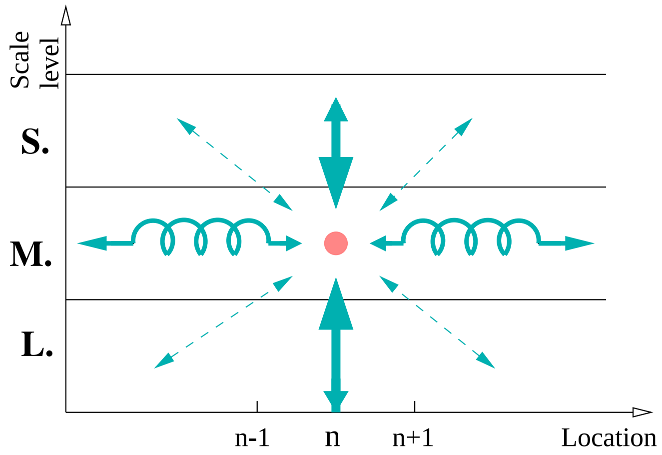

Figure 3.1: A schematic of the energy transfers toward a meso-scale process at location n. Depicted are the transfers from different scales at the same location (thick arrows), transfers from surrounding locations at the same scale level (coiled arrows), and transfers from different scales at different locations (dashed arrows).

problem, as it mingles the inter-scale communications with transfers from the same scale window, and as a result, useful information tends to be be disguised, especially for those processes such as instabilities. We must get rid of this part in order to have the desired information show up from under the camouflage.

3.7.1 Triad interaction and detailed analysis

The inter-location and inter-scale transfers can be distinguished with an analysis of the interactions involved. Observe that each interaction term in the PE model equations (3.4)-(3.7) adopts a quadratic form. This makes the interaction triad a basic unit for each of the transfer processes that arise (particularly, perfect transfer processes). By an **interaction triad** we mean three entities, called modes, with distinct location and scale identities which interact with each other in the phase space.[8] In our formulation, these modes are labeled with their corresponding location index n and scale window index w. We write a mode as $(n,\ w)$, and a triad as $(n,w;\ n_1,w_1;\ n_2,\ w_2)$. for n, n_1, n_2 = 0, ..., 2^{j_2}-1, and w, w_1, w_2 = 0,1,2. When a meso-scale window is concerned, such a mode has been identified as an eddy in real geophysical fluid flows.

Interactions between the triad modes, or **triad interactions**, are fundamental in energy transfer processes. We have examined before a basic transfer function, $T(w,n) = \widehat{\mathcal{Z}}_n^{\sim w} \cdot \widehat{(pq)}_n^{\sim w}$, for fields $\mathcal{Z}, p, q \in V_{\varrho,j_2}$, with $w = 0,1,2$, corresponding to the large-, meso-, and sub-meso-scale window, respectively. In this context, it just represents the cumulative transfer due to many of these irreducible interactions (while one or more of these basic transfers make(s) a (perfect) transfer process). Decomposition of a transfer term into a sum of triad interaction functions is called a **triad interaction analysis**, or simply **interaction analysis**. In this section, we will show how such an analysis works with the basic transfer function $T(w,n) = \widehat{\mathcal{Z}}_n^{\sim w} \widehat{(pq)}_n^{\sim w}$.

To begin, a representation of triad modes is necessary. Compared to the classical energetic studies, triad interaction analysis in the MS-EVA is compounded by the addition of location-location interactions in the phase space (cf. Fig. 3.1). Fournier (1999) and Khac (1998) also mentioned this fact in their wavelet formulations. For $p \in V_{\varrho,j_2}$, recall from Chapter 2 the reconstruction

$$p(t) = \sum_{w=0}^{2} \sum_{n \in \mathcal{N}_\varrho^{j_2}} \widehat{p}_n^{\sim w} \phi_n^{\varrho,j_2}(t).$$

This is to say, $\widehat{p}_n^{\sim w} \phi_n^{\varrho,j_2}(t)$ may be technically treated as the characterization of mode $(n,\ w)$.

[8]The term *triad* is borrowed from the Fourier (and wavelet) energetic analysis with respect to space, where a conservation rule can be derived (see Iima and Toh, 1995; Rajaee and Karlsson, 1992; Pedlosky, 1979; Waleffe, 1992; Fournier, 1999). Here we have not found such a rule and the usage of this terminology is rather loose.

Denote

$$Tr(n, w \mid n_1, w_1; n_2, w_2) \equiv \widehat{\mathcal{Z}}_n^{\sim w} \cdot \left[\widehat{p}_{n_1}^{\sim w_1} \widehat{q}_{n_2}^{\sim w_2} \; \widehat{(\phi_{n_1}^{\varrho, j_2} \phi_{n_2}^{\varrho, j_2})}_n^{\sim w} \right], \tag{3.124}$$

which stands for the rate of energy transferred to (n, w) from the interaction of (n_1, w_1) and (n_2, w_2). The basic transfer function $T(w, n)$ then takes the form

$$T(w, n) = \sum_{w_1, w_2} \sum_{n_1, n_2} Tr(n, w \mid n_1, w_1; n_2, w_2). \tag{3.125}$$

For convenience, $Tr(n, w \mid n_1, w_1; n_2, w_2)$ is called a **unit expression** of the interaction amongst the triad $(n, w;\ n_1, w_1;\ n_2, w_2)$, where (n_1, w_1) and (n_2, w_2) are the **giving modes** while (n, w) is the **receiving mode**, a naming convention after Iima and Toh (1995).

Equation (3.125) is the detailed analysis for the basic transfer function $T(w, n)$. Because of its "atomic" structure, virtually any kind of interaction analysis-related task can be fulfilled with it.

3.7.2 Modified interaction analysis

As just mentioned, theoretically Eq. (3.125) is enough for us to achieve the goal of inter-scale information unraveling. Practically, however, it is by no means an efficient way to serve our purpose. The computation with it is very expensive because of the huge number of mode combinations and hence the huge number of triads, while such a detailed analysis is not at all necessary for real problems. If (3.125) is modified such that some of the terms are combined the computational redundancy may be then greatly reduced whereas the physical interpretation could be even clearer. In this subsection, we will derive such a modified expression. For convenience, we still call it an interaction analysis though now it is not a "detailed" one any more.

Look at the meso-scale window ($w = 1$) first. It is of particular importance because it mediates between the large scales and sub-meso scales on a spectrum. For a field p, make the decomposition

$$\begin{aligned} p &= \widehat{p}_n^{\sim 1} \phi_n^{\varrho, j_2}(t) + p_{*^1} \\ &= p^{\sim 0} + \widehat{p}_n^{\sim 1} \phi_n^{\varrho, j_2}(t) + p_{*^1}^{\sim 1} + p^{\sim 2}, \end{aligned} \tag{3.126}$$

where

$$p_{*^1} = p - \widehat{p}_n^{\sim 1} \phi_n^{\varrho, j_2}(t) \tag{3.127}$$

and $p_{*^1}^{\sim 1}$ is the meso-scale part of p_{*^1},

$$\begin{aligned} p_{*^1}^{\sim 1} &= p^{\sim 1} - \widehat{p}_n^{\sim 1} \phi_n^{\varrho, j_2} \\ &= \sum_{i \in \mathcal{N}_\varrho^{j_2}, i \neq n} \widehat{p}_i^{\sim 1} \phi_i^{\varrho, j_2}. \end{aligned} \tag{3.128}$$

The new interaction analysis concerns the relationship between scales and locations, instead of between triads. The advantage of this is that we don't have to resort to those triad modes, which may not have physical correspondence in the large-scale window, to make interpretation. Note not any $\widehat{p}_n^{\sim 1} \phi_n^{\varrho, j_2}$ can convincingly characterize $p^{\sim 1}(t)$ at location n. But in this context, as the basis function $\phi_n^{\varrho, j_2}(t)$ is a very localized one (localization order $\gamma \geq 2$. cf. the preceding chapter), we expect the removal of $\widehat{p}_n^{\sim 1} \phi_n^{\varrho, j_2}$ will effectively (though not totally) eliminate from $p^{\sim 1}$ the contribution from location n. This has been evidenced in the example of of a meridional velocity series v (Fig. 3.2), where at $n = 384$, $v_{*^1}^{\sim 1}$ is only about 6% ($|\frac{-0.0106}{0.17}|$) of the $v^{\sim 1}$ in magnitude, while at other locations v and $v_{*^1}^{\sim 1}$ are almost the same (fluctuations negligible around n). Therefore, one may practically, albeit not perfectly, take $\widehat{p}_n^{\sim 1} \phi_n^{\varrho, j_2}$ as the meso-scale part of p with contribution from location n only (corresponding to $t = 2^{-j_2} n$), and $p_{*^1}^{\sim 1}$ the part from all locations other than n. Note $p_{*^1}^{\sim 1}$ has an n-dependence. For notational clarity, it is suppressed henceforth.

Likewise, for field $q \in V_{\varrho, j_2}$, it can also be decomposed as

$$\begin{aligned} q &= q^{\sim 0} + q^{\sim 1} + q^{\sim 2} && (3.129) \\ &= q^{\sim 0} + \widehat{q}_n^{\sim 1} \ \phi_n^{\varrho, j_2} + q_{*^1}^{\sim 1} + q^{\sim 2}, && (3.130) \end{aligned}$$

with interpretation analogous to that of $p_{*^1}^{\sim 1}$ for the starred term. The decompositions for p and q yield an analysis of the basic transfer function $T(1, n) = \widehat{\mathcal{Z}}_n^{\sim 1} \cdot \widehat{(pq)}_n^{\sim 1}$ into interactions as shown in the following:

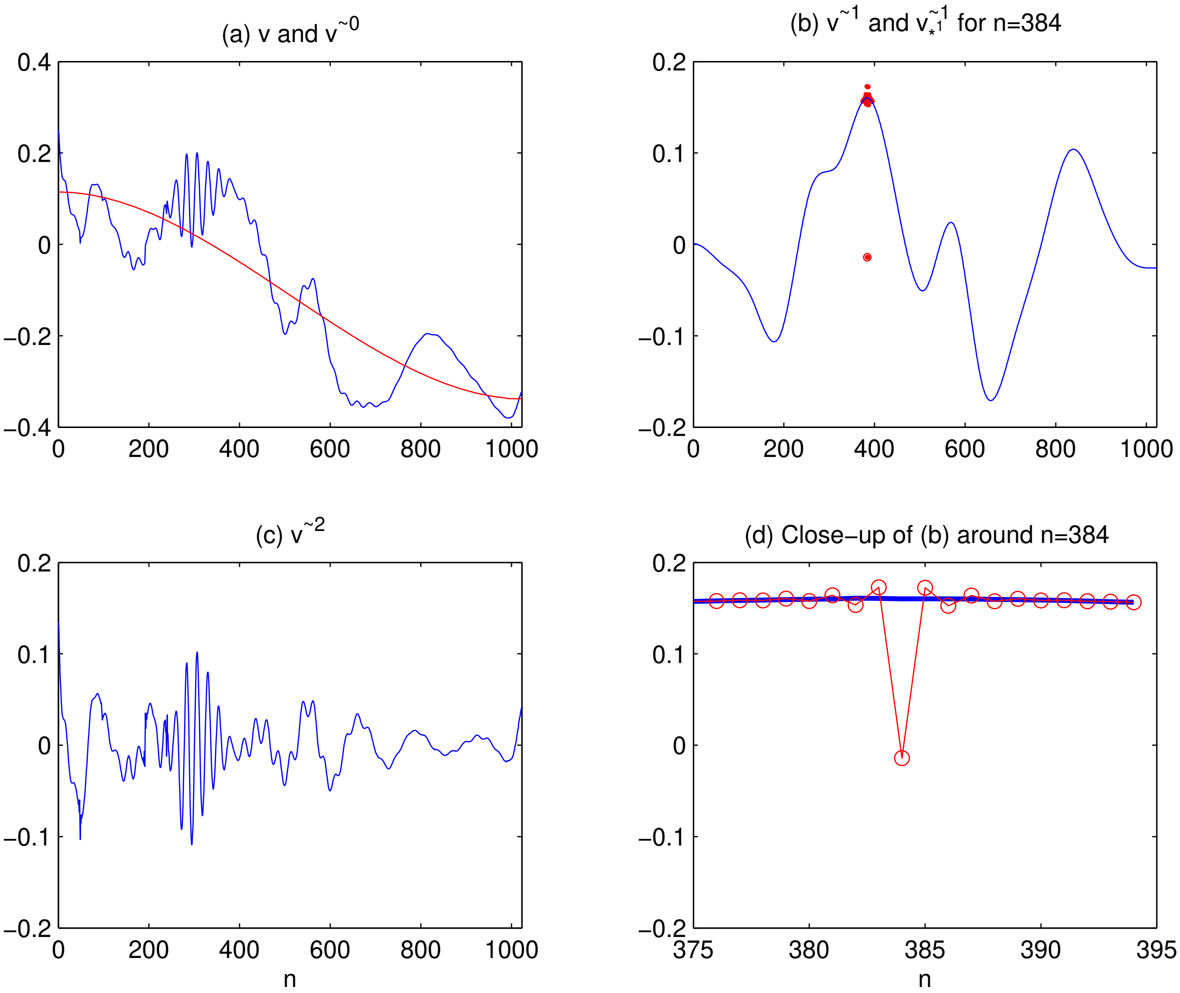

Figure 3.2: A typical time series of v (in cm/s) from the IFF forecast dataset (point (35, 43, 2). See Chapter 5.) and its derived series. There are $2^{j_2} = 1024$ data points, and scale windows are chosen such that $j_0 = 0$ and $j_1 = 4$. The original series v and its large-scale reconstruction $v^{\sim 0}$ are shown in (a), and the meso-scale and sub-mesoscale are plotted in (b) and (c) respectively. Also plotted in (b) is the "starred" series (dotted) $v_{*1}^{\sim 1}$ for location $n = 384$. (d) is the close-up of (b) around $n = 384$. Apparently, $v_{*1}^{\sim 1}$ is at least one order smaller than $v^{\sim 1}$ in size at that point, while these two are practically the same at other points. Location n corresponds to a scaled time $t = 2^{-j_2}n$ (here forecast day 8).

	$p^{\sim 0}$	$\widehat{p}_n^{\sim 1}\phi_n^{\varrho,j_2}$	$p_{*^1}^{\sim 1}$	$p^{\sim 2}$
$q^{\sim 0}$	L-L	L-$\mathrm{M_n}$	L-$\mathrm{M_*}$	L-S
$\widehat{q}_n^{\sim 1}\phi_n^{\varrho,j_2}$	$\mathrm{M_n}$-L	$\mathrm{M_n}$-$\mathrm{M_n}$	$\mathrm{M_n}$-$\mathrm{M_*}$	$\mathrm{M_n}$-S
$q_{*^1}^{\sim 1}$	$\mathrm{M_*}$-L	$\mathrm{M_*}$-$\mathrm{M_n}$	$\mathrm{M_*}$-$\mathrm{M_*}$	$\mathrm{M_*}$-S
$q^{\sim 2}$	S-L	S-$\mathrm{M_n}$	S-$\mathrm{M_*}$	S-S

In this table, L stands for large-scale window and S for sub-mesoscale window (all locations). $\mathrm{M_n}$ is used to denote the meso-scale contribution from location n, while $\mathrm{M_*}$ signifies the meso-scale contributions *other than* that location. Among these interactions, $\mathrm{M_n}$-$\mathrm{M_*}$ and $\mathrm{M_*}$-$\mathrm{M_*}$ contribute to $T(1,n)$ from the same scale window (meso-scale). There's no cross-scale transfers happening in there. Cross-scale energy transfer is fulfilled through other processes. We may sub-total all the resulting sixteen terms into five more meaningful terms:

$$\begin{aligned}
T_n^{0\to 1} &= \widehat{\mathcal{Z}}_n^{\sim 1}\cdot\Big[\widehat{(p^{\sim 0}q^{\sim 0})}_n^{\sim 1} + \widehat{q}_n^{\sim 1}\widehat{(p^{\sim 0}\phi_n^{\varrho,j_2})}_n^{\sim 1} + \widehat{(p^{\sim 0}q_{*^1}^{\sim 1})}_n^{\sim 1} \\
&\qquad + \widehat{p}_n^{\sim 1}\widehat{(\phi_n^{\varrho,j_2}q^{\sim 0})}_n^{\sim 1} + \widehat{(p^{\sim 1}q^{\sim 0})}_n^{\sim 1}\Big] \\
&= \widehat{\mathcal{Z}}_n^{\sim 1}\cdot\Big[\widehat{(p^{\sim 0}q^{\sim 0})}_n^{\sim 1} + \widehat{(p^{\sim 1}q^{\sim 0})}_n^{\sim 1} + \widehat{(p^{\sim 0}q^{\sim 1})}_n^{\sim 1}\Big] \qquad (3.131) \\
T_n^{2\to 1} &= \widehat{\mathcal{Z}}_n^{\sim 1}\cdot\Big[\widehat{p}_n^{\sim 1}\widehat{(\phi_n^{\varrho,j_2}q^{\sim 2})}_n^{\sim 1} + \widehat{(p_{*^1}^{\sim 1}q^{\sim 2})}_n^{\sim 1} + \widehat{q}_n^{\sim 1}\widehat{(p^{\sim 2}\phi_n^{\varrho,j_2})}_n^{\sim 1} \\
&\qquad + \widehat{(p^{\sim 2}q_{*^1}^{\sim 1})}_n^{\sim 1} + \widehat{(p^{\sim 2}q^{\sim 2})}_n^{\sim 1}\Big] \\
&= \widehat{\mathcal{Z}}_n^{\sim 1}\cdot\Big[\widehat{(p^{\sim 1}q^{\sim 2})}_n^{\sim 1} + \widehat{(p^{\sim 2}q^{\sim 2})}_n^{\sim 1} + \widehat{(p^{\sim 2}q^{\sim 1})}_n^{\sim 1}\Big] \qquad (3.132) \\
T_n^{0\oplus 2\to 1} &= \widehat{\mathcal{Z}}_n^{\sim 1}\cdot\Big[\widehat{(p^{\sim 2}q^{\sim 0})}_n^{\sim 1} + \widehat{(p^{\sim 0}q^{\sim 2})}_n^{\sim 1}\Big] \qquad (3.133) \\
T_{n\to n}^{1\to 1} &= \widehat{\mathcal{Z}}_n^{\sim 1}\cdot\Big[\widehat{p}_n^{\sim 1}\widehat{q}_n^{\sim 1}\widehat{\left(\phi_n^{\varrho,j_2}\right)^2}_n^{\sim 1}\Big] \qquad (3.134) \\
T_{\mathrm{other}\to n}^{1\to 1} &= \widehat{\mathcal{Z}}_n^{\sim 1}\cdot\Big[\widehat{(p^{\sim 1}q_{*^1}^{\sim 2})}_n^{\sim 1} + \widehat{q}_n^{\sim 1}\widehat{(p_{*^1}^{\sim 2}\phi_n^{\varrho,j_2})}_n^{\sim 1}\Big]. \qquad (3.135)
\end{aligned}$$

If necessary, $T_{n\to n}^{1\to 1}$ and $T_{\mathrm{other}\to n}^{1\to 1}$ may also be combined to one term. The result is denoted as $T_n^{1\to 1}$.

The physical interpretations of above five terms are clear. Their superscripts reveal how energy is transferred to mode $(1,n)$ from other scales. Specifically, $T_n^{0\to 1}$ and $T_n^{2\to 1}$ are transfer rates from windows 0 and 1, respectively, and $T_n^{0\oplus 2\to 1}$ is the contribution from the window 0-window 2 interaction over the meso-scale range. The last two terms, $T_{n\to n}^{1\to 1}$ and $T_{\mathrm{other}\to n}^{1\to 1}$, sum up to $T_n^{1\to 1}$, which represents the part of transfer from the same window.

Above are the interaction analysis for $T(1,n)$. Using the same technique, one can obtain a similar analysis for $T(0,n)$:

$$\begin{aligned} T(0,n) &= \widehat{\mathcal{Z}}_n^{\sim 0} \cdot \widehat{(pq)}_n^{\sim 0} \\ &= T_n^{1\to 0} + T_n^{2\to 0} + T_n^{1\oplus 2\to 0} + T_{n\to n}^{0\to 0} + T_{\text{other}\to n}^{0\to 0}, \end{aligned} \tag{3.136}$$

where

$$T_n^{1\to 0} = \widehat{\mathcal{Z}}_n^{\sim 0} \cdot \left[\widehat{(p^{\sim 1} q^{\sim 1})}_n^{\sim 0} + \widehat{(p^{\sim 1} q^{\sim 0})}_n^{\sim 0} + \widehat{(p^{\sim 0} q^{\sim 1})}_n^{\sim 0} \right] \tag{3.137}$$

$$T_n^{2\to 0} = \widehat{\mathcal{Z}}_n^{\sim 0} \cdot \left[\widehat{(p^{\sim 0} q^{\sim 2})}_n^{\sim 0} + \widehat{(p^{\sim 2} q^{\sim 2})}_n^{\sim 0} + \widehat{(p^{\sim 2} q^{\sim 0})}_n^{\sim 0} \right] \tag{3.138}$$

$$T_n^{1\oplus 2\to 0} = \widehat{\mathcal{Z}}_n^{\sim 0} \cdot \left[\widehat{(p^{\sim 2} q^{\sim 1})}_n^{\sim 0} + \widehat{(p^{\sim 1} q^{\sim 2})}_n^{\sim 0} \right] \tag{3.139}$$

$$T_{n\to n}^{0\to 0} = \widehat{\mathcal{Z}}_n^{\sim 0} \cdot \left[\widehat{p}_n^{\sim 0} \widehat{q}_n^{\sim 0} \widehat{\left(\phi_n^{\varrho, j_2}\right)^2}_n^{\sim 0} \right] \tag{3.140}$$

$$T_{\text{other}\to n}^{0\to 0} = \widehat{\mathcal{Z}}_n^{\sim 0} \cdot \left[\widehat{(p^{\sim 0} q_{*^0}^{\sim 0})}_n^{\sim 0} + \widehat{q}_n^{\sim 0} \widehat{(p_{*^0}^{\sim 0} \phi_n^{\varrho, j_2})}_n^{\sim 0} \right], \tag{3.141}$$

and $T(2,n)$:

$$\begin{aligned} T(2,n) &= \widehat{\mathcal{Z}}_n^{\sim 2} \cdot \widehat{(pq)}_n^{\sim 2} \\ &= T_n^{0\to 2} + T_n^{1\to 2} + T_n^{0\oplus 1\to 2} + T_{n\to n}^{2\to 2} + T_{\text{other}\to n}^{2\to 2}, \end{aligned} \tag{3.142}$$

where

$$T_n^{0\to 2} = \widehat{\mathcal{Z}}_n^{\sim 2} \cdot \left[\widehat{(p^{\sim 0} q^{\sim 0})}_n^{\sim 2} + \widehat{(p^{\sim 2} q^{\sim 0})}_n^{\sim 2} + \widehat{(p^{\sim 0} q^{\sim 2})}_n^{\sim 2} \right] \tag{3.143}$$

$$T_n^{1\to 2} = \widehat{\mathcal{Z}}_n^{\sim 2} \cdot \left[\widehat{(p^{\sim 1} q^{\sim 2})}_n^{\sim 2} + \widehat{(p^{\sim 1} q^{\sim 1})}_n^{\sim 2} + \widehat{(p^{\sim 2} q^{\sim 1})}_n^{\sim 2} \right] \tag{3.144}$$

$$T_n^{0\oplus 1\to 2} = \widehat{\mathcal{Z}}_n^{\sim 2} \cdot \left[\widehat{(p^{\sim 0} q^{\sim 1})}_n^{\sim 2} + \widehat{(p^{\sim 1} q^{\sim 0})}_n^{\sim 2} \right] \tag{3.145}$$

$$T_{n\to n}^{2\to 2} = \widehat{\mathcal{Z}}_n^{\sim 2} \cdot \left[\widehat{p}_n^{\sim 2} \widehat{q}_n^{\sim 2} \widehat{\left(\phi_n^{\varrho, j_2}\right)^2}_n^{\sim 2} \right] \tag{3.146}$$

$$T_{\text{other}\to n}^{2\to 2} = \widehat{\mathcal{Z}}_n^{\sim 2} \cdot \left[\widehat{(p^{\sim 2} q_{*^2}^{\sim 2})}_n^{\sim 2} + \widehat{q}_n^{\sim 2} \widehat{(p_{*^2}^{\sim 2} \phi_n^{\varrho, j_2})}_n^{\sim 2} \right]. \tag{3.147}$$

In these analyses, p_{*^0} and p_{*^2} are defined as

$$p_{*^0} = p - \widehat{p}_n^{\sim 0} \phi_n^{\varrho, j_2}(t), \tag{3.148}$$

$$p_{*^2} = p - \widehat{p}_n^{\sim 2} \phi_n^{\varrho, j_2}(t). \tag{3.149}$$

The physical meaning of the interaction terms is embedded in these mnemonic notations. In the superscripts, arrows signify the directions of energy transfer and the numbers 0, 1, 2 represent the large-scale, meso-scale, and sub-mesoscale windows, respectively. For easy reference, we give a complete list of the interaction analyses in Table 3.1 for the transfer function $T(w, n)$, for $w = 0, 1, 2$.

What merits mentioning is that these interaction analyses are by no means unique. One may obtain different analyses by making different sub-grouping for Eq. (3.125) according to the problem studied. The moral here is to try to avoid those starred terms as in Eq. (3.135), which makes the major overhead in computation (in terms of either memory or CPU usage). In the above analyses, say the meso-scale analysis, if a whole transfer process is calculated, the sum of those terms in the form of $T_{n \to n}^{1 \to 1}$ will vanish by the definition of transfer processes (note a transform does not affect the spatial structure of a field). This also implies that the sum of those transfer functions in the form of $T_{\text{other} \to n}^{1 \to 1}$ will be equal to the sum of terms in the same form but with all the stars dropped. Hence in performing interaction analysis for a transfer process, we may simply ignore the stars for the corresponding terms. But if it is an arbitrary transfer term which does not necessarily form a transfer process (e.g, $T_{K^M, w\partial_z}$), the starred-term-caused heavy computational overhead will be still a problem.

Fortunately, under certain circumstances, this overhead may be avoided in practice. Recall that we have built a highly localized scaling basis function ϕ. For any $p \in V_{\varrho, j_2}$, it yields a function $p(t)\phi_n^{\varrho, j_2}(t)$ with an effective support of the order of the grid size. The large- or meso-scale transform of this function is thence negligible, should j_1 be smaller than j_2 by some considerable number (3 is enough). Only when it is in the sub-mesoscale window need we really compute the starred term. An example with a typical time series of ρ and u is plotted in Fig. 3.3. Apparently, for the large-scale and meso-scale cases, $\widehat{\rho}_n^{\sim 0}\widehat{(u\phi_n^{\varrho, j_2})}_n^{\sim 0}$ and $\widehat{\rho}_n^{\sim 1}\widehat{(u\phi_n^{\varrho, j_2})}_n^{\sim 1}$ (red circles) are very small and hence $\widehat{(\rho_{*^0}^{\sim 0}u)}_n^{\sim 0}$ and $\widehat{(\rho_{*^1}^{\sim 1}u)}_n^{\sim 1}$ can be approximated by $\widehat{(\rho^{\sim 0}u)}_n^{\sim 0}$ and $\widehat{(\rho^{\sim 1}u)}_n^{\sim 1}$, respectively. This approximation fails only in the sub-mesoscale case, where the corresponding two parts are of the same order.

It is of interest to give an estimation of the relative importance of all these interaction terms obtained thus far. For the mesoscale transfer function $T(1, n)$, $T_n^{0 \oplus 2 \to 1}$ is generally not significant (compared to other terms). This is because, on a spectrum, if two processes

Table 3.1: Interaction analysis of the energy transfers for transfer functions $T(w, n)$, $w = 0, 1, 2$. (L., M. and S. denote large-scale, meso-scale and sub-mesoscale windows, respectively.)

Destination	Analysis term		Source
Large-scale window, location n $T(0,n)$	$T_n^{1\to 0}$		Meso-scale window
	$T_n^{2\to 0}$		Sub-mesoscale window
	$T_n^{1\oplus 2\to 0}$		M.-S. interaction
	$T_n^{0\to 0}$	$T_{n\to n}^{0\to 0}$	Same scale window (large-scale), same location (n)
		$T_{\text{other}\to n}^{0\to 0}$	Same scale window (large-scale), locations other than n
Meso-scale window, location n $T(1,n)$	$T_n^{0\to 1}$		Large-scale window
	$T_n^{2\to 1}$		Sub-mesoscale window
	$T_n^{0\oplus 2\to 1}$		L.-S. interaction
	$T_n^{1\to 1}$	$T_{n\to n}^{1\to 1}$	Same scale window (meso-scale), scame location (n)
		$T_{\text{other}\to n}^{1\to 1}$	Same scale window (meso-scale), locations other than n
Sub-mesoscale window, location n $T(2,n)$	$T_n^{0\to 2}$		Large-scale window
	$T_n^{1\to 2}$		meso-scale window
	$T_n^{0\oplus 1\to 2}$		L.-M. interaction
	$T_n^{2\to 2}$	$T_{n\to n}^{2\to 2}$	Same scale window (sub-mesoscale), scame location (n)
		$T_{\text{other}\to n}^{2\to 2}$	Same scale window (sub-mesoscale), locations other than n

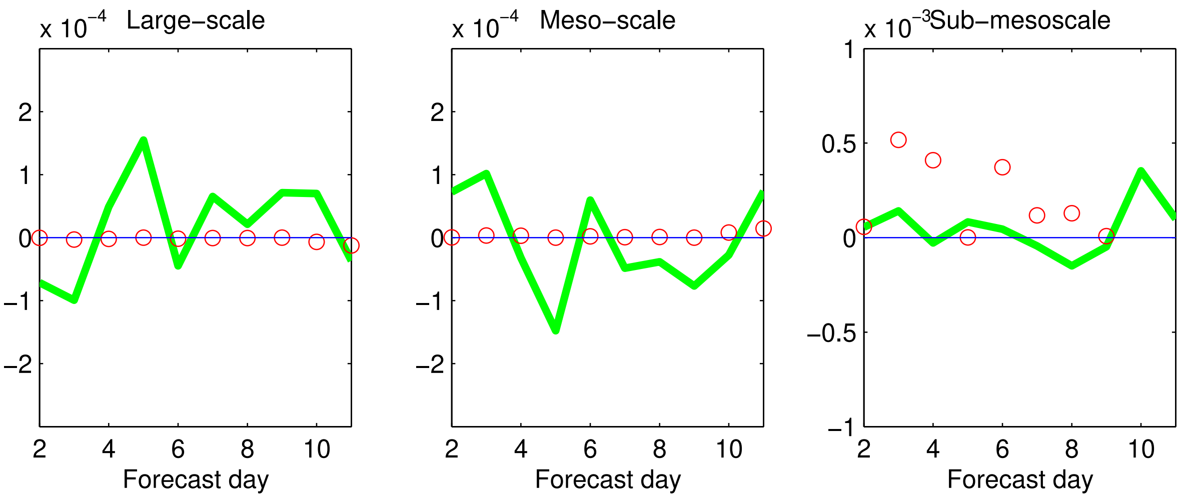

Figure 3.3: An example showing relative importance of the decomposed terms from $T_{A,\partial_h\rho}$. Data source: same as that in Fig. 3.2 (zonal velocity only). Unit: kg/m^{2}s.

Left: $\widehat{(\rho^{\sim 0}_{*^0} u)}_n^{\sim 0}$ (heavy solid line) and $\widehat{\rho}_n^{\sim 0}\left(\widehat{u\phi_n^{\varrho,j_2}}\right)_n^{\sim 0}$ (circle);

Middle: $\widehat{(\rho^{\sim 1}_{*^1} u)}_n^{\sim 1}$ (heavy solid line) and $\widehat{\rho}_n^{\sim 1}\left(\widehat{u\phi_n^{\varrho,j_2}}\right)_n^{\sim 1}$ (circle);

Right: $\widehat{(\rho^{\sim 1}_{*^2} u)}_n^{\sim 2}$ (heavy solid line) and $\widehat{\rho}_n^{\sim 2}\left(\widehat{u\phi_n^{\varrho,j_2}}\right)_n^{\sim 2}$ (circle).

Obviously, the $\widehat{(\rho^{\sim w}_{*^w} u)}_n^{\sim w}$ in the decomposition

$$\widehat{(\rho^{\sim w} u)}_n^{\sim w} = \widehat{(\rho^{\sim w}_{*^w} u)}_n^{\sim w} + \widehat{\rho}_n^{\sim w}\widehat{(u\phi_n^{\varrho,j_2})}_n^{\sim w}$$

can be well approximated by $\widehat{(\rho^{\sim w} u)}_n^{\sim w}$ for windows $w = 0, 1$.

are far away from each other (as is the case for large scale and sub-meso scale), they are usually separable and the interaction are accordingly very weak. Even if there exists some interaction, the spawned new processes generally stay in their original windows, seldom going into between. Apart from $T_n^{0\oplus 2\to 1}$, all the others are of comparable sizes, though more often than not $T_n^{0\to 1}$ dominates the rest (e.g., Fig. 3.4b).

For the large-scale window, things are a little different. This time it is term $T_n^{2\to 0}$ that is not significant, with the same reason as above. But term $T_n^{1\oplus 2\to 0}$ is in general not negligible. In this window, the dominant energy transfer is usually not from other scales, but from other locations at the same scale level. Mathematically this is to say, $T_{\text{other}\to n}^{0\to 0}$ usually dominates the other terms. This is understandable since a large-scale feature results from interactions with modes covering a large range of location on the time series. If each location contributes even a little bit, the grand total could be huge. This fact is seen in the example in Fig. 3.4a.

By the same argument as above, within the sub-mesoscale window, the dominant term is $T_n^{1\to 2}$. But $T_n^{0\oplus 1\to 2}$ could be of some importance also. In comparison to these two, $T_n^{0\to 2}$ and $T_n^{2\to 2} = T_{\text{other}\to n}^{2\to 2} + T_{n\to n}^{2\to 2}$ are not significant.

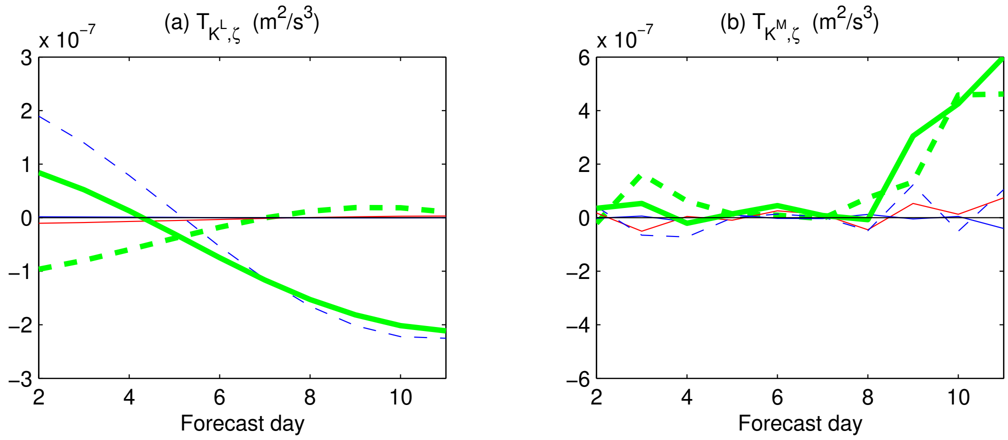

Figure 3.4: An example showing the relative importance of analytical terms of $T_{K,\zeta}$ at ten (time) locations. The data source and parameter choice are the same as that of Fig. 3.2. The factor 2^{j_2} has been multiplied.
(a) Analysis of $T_{K^L,\zeta}$ (thick green solid): $T_{K^L,\zeta}^{1\to 0}$ (thick green dashed), $T_{K^L,\zeta}^{2\to 0}$ (red solid), $T_{K^L,\zeta}^{1\oplus 2\to 0}$ (blue solid), and $T_{K^L,\zeta}^{0\to 0}$ (blue dashed). **(b)** Analysis of $T_{K^M,\zeta}$ (thick green solid): $T_{K^M,\zeta}^{0\to 1}$ (thick green dashed), $T_{K^M,\zeta}^{2\to 1}$ (red solid), $T_{K^M,\zeta}^{0\oplus 2\to 1}$ (blue solid), and $T_{K^M,\zeta}^{1\to 1}$ (blue dashed).

3.8 Vorticity equation

With the multiscale energetics formulated, the vorticity part of the MS-EVA now can be easily developed. The following two sections deal with this development.

First we need a governing equation for the vorticity field, $\zeta = \underline{\mathbf{k}} \cdot \nabla \wedge \underline{\mathbf{v}}$. It is obtained by crossing the momentum equation (3.4) followed by a dot product with $\underline{\mathbf{k}}$,

$$\frac{\partial \zeta}{\partial t} = \underline{\mathbf{k}} \cdot \nabla \wedge w \frac{\partial \underline{\mathbf{v}}}{\partial z} - \underline{\mathbf{k}} \cdot \nabla \wedge [(f+\zeta)\underline{\mathbf{k}} \wedge \underline{\mathbf{v}}] + F_{\zeta,z} + F_{\zeta,h}, \tag{3.150}$$

where $F_{\zeta,z}$ and $F_{\zeta,h}$ denote respectively the vertical and horizontal diffusion. Making use of the continuity equation (3.5), we get,

$$\frac{\partial \zeta}{\partial t} = \underbrace{-\nabla \cdot (\underline{\mathbf{v}}\zeta) - \frac{\partial}{\partial z}(w\zeta)}_{\text{(I)}} \underbrace{-\beta v}_{\text{(II)}} \underbrace{+(f+\zeta)\frac{\partial w}{\partial z}}_{\text{(III)}} \underbrace{+\underline{\mathbf{k}} \cdot \frac{\partial \underline{\mathbf{v}}}{\partial z} \wedge \nabla w}_{\text{(IV)}} \underbrace{+F_{\zeta,z} + F_{\zeta,h}}_{\text{(V)}}. \tag{3.151}$$

Here $\beta = \partial f / \partial y$ is a constant if a β-plane is approximation is assumed. But in general, it does not need to be so.

In Eq. (3.151), there are five mechanisms that contribute to the change of relative vorticity ζ. Apparently, term (I) is the advection of ζ by the flow, and term (V) the diffusion. β-effect comes into play through term (II). It is the advection of planetary vorticity f by meridional velocity v. Vortex tubes may stretch or shrink. The vorticity gain or loss due to stretching or shrinking is represented in term (III). Vortex tube may also tilt. Term (IV) results from such an effect.

In order to better understand the contributions to the evolution of ζ, Spall (1989) split the velocity field into a rotational component and a irrotational component.

$$\underline{\mathbf{v}} = \underline{\mathbf{k}} \wedge \nabla \Psi + \nabla \chi, \tag{3.152}$$

where Ψ is the streamfunction, and χ the velocity potential. As a geostrophic flow is horizontally nondivergent on an f-plane, this decomposition helps distinguish geostrophy from ageostrophy in a flow. With decomposition (3.152), ζ is related to Ψ as: $\zeta = \nabla \wedge \underline{\mathbf{v}} = \nabla^2 \Psi$. Following Spall (1989), denote terms due to geostrophy by superscript qg, terms appearing only in the linear balance by lb, and terms appearing in the full balance by fb. The evolution of ζ is then

$$\left[\dot{\zeta}\right]^{\mathrm{qg}} = -[J(\Psi,\zeta)]^{\mathrm{qg}} - [J(\Psi,f)]^{\mathrm{qg}} - \left[f\nabla^2\chi\right]^{\mathrm{qg}} - [\nabla\chi \cdot \nabla f]^{\mathrm{lb}}$$

$$
\begin{aligned}
& -\left[\nabla\chi\cdot\nabla\zeta\right]^{\mathrm{fb}} - \left[\zeta\nabla^2\chi\right]^{\mathrm{fb}} - \left[w\zeta_z\right]^{\mathrm{fb}} \\
& -\left[\nabla w\cdot\nabla\Psi_z\right]^{\mathrm{fb}} - J(w,\chi_z) + F_{\zeta,z} + F_{\zeta,h},
\end{aligned}
\tag{3.153}
$$

where J signifies the Jacobian operator ($J(a,b) = \frac{\partial a}{\partial x}\frac{\partial b}{\partial y} - \frac{\partial b}{\partial x}\frac{\partial a}{\partial y}$). This is the Eq. (15) of Spall (1989), and the interpretation for each term is seen in the following sections therein (also seen in Lorenz, 1960).

3.9 Multiscale enstrophy equations and their interpretations

3.9.1 Large-scale enstrophy equation

Enstrophy $Z = \frac{1}{2}\zeta^2$ is a positive definite measure of rotation. It is the *"energy" of vorticity.* According to Chapter 2, the large-scale enstrophy at location n is defined to be

$$
Z_n^L = \frac{1}{2}\left(\widehat{\zeta}_n^{\sim 0}\right)^2. \tag{3.154}
$$

We begin the derivation of evolution of Z_n^L with Eq. (3.151). Eq. (3.153) is too cumbersome, and unless necessary for some specific problems, it is not advised for multiscale analysis.

As before, first discretize the only time derivative term in Eq. (3.151), $\partial\zeta/\partial t$, to $\delta_t\zeta$. Take a large-scale transform of the resulting equation and then multiply it by $\widehat{\zeta}_n^{\sim 0}$. The left hand side results in the evolution $\hat{\delta}_n Z_n^L$ plus a correction term which is of the order Δt^2, Δt being the time spacing of the series. Merging the correction term into the horizontal diffusion, we get an equation

$$
\begin{aligned}
\dot{Z}_n^L \quad = \quad & \underbrace{-\widehat{\zeta}_n^{\sim 0}\left[\nabla\cdot\widehat{(\underline{\mathbf{v}}\zeta)}_n^{\sim 0} + \frac{\partial\widehat{(w\zeta)}_n^{\sim 0}}{\partial z}\right]}_{(\mathbf{AD})} - \beta\,\widehat{\zeta}_n^{\sim 0}\widehat{v}_n^{\sim 0} \\
& + f\,\widehat{\zeta}_n^{\sim 0}\widehat{\left(\frac{\partial w}{\partial z}\right)}_n^{\sim 0} + \widehat{\zeta}_n^{\sim 0}\,\widehat{\left(\zeta\frac{\partial w}{\partial z}\right)}_n^{\sim 0} \\
& + \widehat{\zeta}_n^{\sim 0}\,\underline{\mathbf{k}}\cdot\widehat{\left(\frac{\partial\underline{\mathbf{v}}}{\partial z}\wedge\nabla w\right)}_n^{\sim 0} + F_{Z^L,z} + F_{Z^L,h}.
\end{aligned}
$$

Again, $F_{Z^L,z}$ and $F_{Z^L,h}$ here are just symbolic representations of the vertical and horizontal diffusions. Following the practice in deriving the APE equation, the process represented by the advection-related terms (AD) can be decomposed into a sum of transport processes and transfer processes. Denote

$$\Delta_h Q_{Z^L} = -\nabla \cdot \left[\frac{1}{4} \widehat{\zeta}_n^{\sim 0} \widehat{(\underline{\mathbf{v}}\zeta)}_n^{\sim 0} + \frac{1}{4} \widehat{(\zeta^2)}_n^{\sim 0} \widehat{\underline{\mathbf{v}}}_n^{\sim 0} \right], \tag{3.155}$$

$$\Delta_z Q_{Z^L} = -\frac{\partial}{\partial z} \left[\frac{1}{4} \widehat{\zeta}_n^{\sim 0} \widehat{(w\zeta)}_n^{\sim 0} + \frac{1}{4} \widehat{(\zeta^2)}_n^{\sim 0} \widehat{w}_n^{\sim 0} \right], \tag{3.156}$$

then it is

$$\begin{aligned}
\mathbf{AD} \;=\;& \Delta_h Q_{Z^L} + \Delta_z Q_{Z^L} \\
&+ \left[-\Delta_h Q_{Z^L} - \widehat{\zeta}_n^{\sim 0} \nabla \cdot (\widehat{(\underline{\mathbf{v}}\zeta)}_n^{\sim 0}) + \frac{1}{4} \widehat{\zeta}_n^{\sim 0} \widehat{(\zeta \nabla \cdot \underline{\mathbf{v}})}_n^{\sim 0} + \frac{1}{4} \widehat{(\zeta^2)}_n^{\sim 0} \nabla \cdot \widehat{\underline{\mathbf{v}}}_n^{\sim 0} \right] \\
&+ \left[-\Delta_z Q_{Z^L} - \widehat{\zeta}_n^{\sim 0} \frac{\partial (\widehat{(w\zeta)}_n^{\sim 0})}{\partial z} + \frac{1}{4} \widehat{\zeta}_n^{\sim 0} \widehat{\left(\zeta \frac{\partial w}{\partial z} \right)}_n^{\sim 0} + \frac{1}{4} \widehat{(\zeta^2)}_n^{\sim 0} \frac{\partial \widehat{w}_n^{\sim 0}}{\partial z} \right] \\
\equiv\;& \Delta_h Q_{Z^L} + \Delta_z Q_{Z^L} + T_{Z^L, \partial_h \zeta} + T_{Z^L, \partial_z \zeta},
\end{aligned}$$

where $\Delta_h Q_{Z^L}$ and $\Delta_z Q_{Z^L}$ represent the horizontal and vertical transports, and $T_{Z^L,\partial_h \zeta}$, $T_{Z^L,\partial_z \zeta}$ the transfer rates for two distinct processes. It is easy to prove, with the aid of (3.109), that both of these processes belong to perfect transfer processes. Note the large-scale continuity equation (3.26) has been used in obtaining the above form of decomposition. If necessary, $\Delta_h Q_{Z^L}$ and $T_{Z^L,\partial_h \zeta}$ may be further decomposed into contributions from x and y directions, respectively.

The enstrophy equation now becomes, after some algebraic manipulation,

$$\begin{aligned}
\dot{Z}_n^L \;=\;& \Delta_h Q_{Z^L} + \Delta_z Q_{Z^L} \\
&+ \left[-\Delta_h Q_{Z^L} - \widehat{\zeta}_n^{\sim 0} \nabla \cdot (\widehat{(\underline{\mathbf{v}}\zeta)}_n^{\sim 0}) + \frac{1}{4} \widehat{\zeta}_n^{\sim 0} \widehat{(\zeta \nabla \cdot \underline{\mathbf{v}})}_n^{\sim 0} + \frac{1}{4} \widehat{(\zeta^2)}_n^{\sim 0} \nabla \cdot \widehat{\underline{\mathbf{v}}}_n^{\sim 0} \right] \\
&+ \left[-\Delta_z Q_{Z^L} - \widehat{\zeta}_n^{\sim 0} \frac{\partial (\widehat{(w\zeta)}_n^{\sim 0})}{\partial z} + \frac{1}{4} \widehat{\zeta}_n^{\sim 0} \widehat{\left(\zeta \frac{\partial w}{\partial z} \right)}_n^{\sim 0} + \frac{1}{4} \widehat{(\zeta^2)}_n^{\sim 0} \frac{\partial \widehat{w}_n^{\sim 0}}{\partial z} \right] \\
&- \beta \; \widehat{\zeta}_n^{\sim 0} \widehat{v}_n^{\sim 0} + f \; \widehat{\zeta}_n^{\sim 0} \frac{\partial \widehat{w}_n^{\sim 0}}{\partial z} + \widehat{\zeta}_n^{\sim 0} \; \widehat{\left(\zeta \frac{\partial w}{\partial z} \right)}_n^{\sim 0} \\
&+ \widehat{\zeta}_n^{\sim 0} \; \underline{\mathbf{k}} \cdot \widehat{\left(\frac{\partial \underline{\mathbf{v}}}{\partial z} \wedge \nabla w \right)}_n^{\sim 0} + F_{Z^L,z} + F_{Z^L,h}.
\end{aligned} \tag{3.157}$$

Or, symbolically,

$$\begin{aligned}
\dot{Z}_n^L \;=\;& \Delta_h Q_{Z^L} + \Delta_z Q_{Z^L} + T_{Z^L,\partial_h \zeta} + T_{Z^L,\partial_z \zeta} + \\
& S_{Z^L,\beta} + S_{Z^L,f\nabla\cdot\underline{\mathbf{v}}} + TS_{Z^L,\zeta\nabla\cdot\underline{\mathbf{v}}} + TS_{Z^L,\mathrm{tilt}} +
\end{aligned}$$

$$F_{Z^L,z} + F_{Z^L,h}. \tag{3.158}$$

Each term of Eq. (3.158) has a clear physical interpretation. We have known that $\Delta_h Q_{Z^L}$ and $\Delta_z Q_{Z^L}$ are horizontal and vertical transports of Z_n^L, respectively, and $T_{Z^L,\partial_h\zeta}$ and $T_{Z^L,\partial_z\zeta}$ transfer rates for two perfect transfer processes. If ζ is horizontally and vertically a constant, then $T_{Z^L,\partial_z\zeta}$ and $T_{Z^L,\partial_h\zeta}$ sum up to zero. We have also explained $F_{Z^L,z} + F_{Z^L,h}$ represents the diffusion process. Among the rest terms, $S_{Z^L,\beta}$ and $S_{Z^L,f\nabla\cdot\underline{\mathbf{v}}}$ stand for two sources/sinks of Z due to β-effect and vortex stretching. This is the reason why they earn symbols initialized with the letter S. The other two terms, $TS_{Z^L,\zeta\nabla\cdot\underline{\mathbf{v}}}$ and $TS_{Z^L,\mathrm{tilt}}$, transfer as well as generate/destroy enstrophy. Processes cannot be well separated for them. In a two-dimensional flow, both of these terms vanishes. As a result, the multiscale enstrophy equation will be more useful in a plane flow than a 3-D flow.

3.9.2 Meso-scale enstrophy equation

The meso-scale enstrophy is defined to be

$$Z_n^M = \frac{1}{2}\left(\widehat{\zeta}_n^{\sim 1}\right)^2. \tag{3.159}$$

A meso-scale transform of Eq. (3.151) followed by a multiplication by $\widehat{\zeta}_n^{\sim 1}$ gives an evolution equation for Z_n^L similar to Eq. (3.160),

$$\begin{aligned}
\dot{Z}_n^M \quad = \quad & \Delta_h Q_{Z^M} + \Delta_z Q_{Z^M} \\
& + \left[-\Delta_h Q_{Z^M} - \widehat{\zeta}_n^{\sim 1} \nabla \cdot \widehat{(\underline{\mathbf{v}}\zeta)}_n^{\sim 1}) + \frac{1}{4}\widehat{\zeta}_n^{\sim 1} \widehat{(\zeta\nabla\cdot\underline{\mathbf{v}})}_n^{\sim 1} + \frac{1}{4}\widehat{(\zeta^2)}_n^{\sim 1} \nabla\cdot\underline{\widehat{\mathbf{v}}}_n^{\sim 1} \right] \\
& + \left[-\Delta_z Q_{Z^M} - \widehat{\zeta}_n^{\sim 1} \frac{\partial \widehat{(w\zeta)}_n^{\sim 1})}{\partial z} + \frac{1}{4}\widehat{\zeta}_n^{\sim 1} \widehat{\left(\zeta\frac{\partial w}{\partial z}\right)}_n^{\sim 1} + \frac{1}{4}\widehat{(\zeta^2)}_n^{\sim 1} \frac{\partial \widehat{w}_n^{\sim 1}}{\partial z} \right] \\
& -\beta\ \widehat{\zeta}_n^{\sim 1}\widehat{v}_n^{\sim 1} + f\ \widehat{\zeta}_n^{\sim 1} \frac{\partial \widehat{w}_n^{\sim 1}}{\partial z} + \widehat{\zeta}_n^{\sim 1}\ \widehat{\left(\zeta\frac{\partial w}{\partial z}\right)}_n^{\sim 1} \\
& + \widehat{\zeta}_n^{\sim 1}\ \underline{\mathbf{k}} \cdot \widehat{\left(\frac{\partial \underline{\mathbf{v}}}{\partial z}\wedge\nabla w\right)}_n^{\sim 1} + F_{Z^M,z} + F_{Z^M,h},
\end{aligned} \tag{3.160}$$

where

$$\Delta_h Q_{Z^M} \quad = \quad -\nabla\cdot\left[\frac{1}{4}\widehat{\zeta}_n^{\sim 1}\widehat{(\underline{\mathbf{v}}\zeta)}_n^{\sim 1} + \frac{1}{4}\widehat{(\zeta^2)}_n^{\sim 1}\underline{\widehat{\mathbf{v}}}_n^{\sim 1}\right], \tag{3.161}$$

$$\Delta_z Q_{Z^M} \quad = \quad -\frac{\partial}{\partial z}\left[\frac{1}{4}\widehat{\zeta}_n^{\sim 1}\widehat{(w\zeta)}_n^{\sim 1} + \frac{1}{4}\widehat{(\zeta^2)}_n^{\sim 1}\widehat{w}_n^{\sim 1}\right]. \tag{3.162}$$

For mnemonic reasons, it is also written as

$$
\begin{aligned}
\dot{Z}_n^M &= \Delta_h Q_{Z^M} + \Delta_z Q_{Z^M} + T_{Z^M,\partial_h\zeta} + T_{Z^M,\partial_z\zeta} + \\
& S_{Z^M,\beta} + S_{Z^M,f\nabla\cdot\underline{\mathbf{v}}} + TS_{Z^M,\zeta\nabla\cdot\underline{\mathbf{v}}} + TS_{Z^M,\text{tilt}} + \\
& F_{Z^M,z} + F_{Z^M,h}.
\end{aligned} \tag{3.163}
$$

Interpretation of these terms is analogous to that of (3.158), except that now it is for meso-scale processes.

3.9.3 Sub-mesoscale enstrophy equation

Following the same procedure as above, an equation governing the sub-mesoscale enstrophy, $Z_n^S = \frac{1}{2}\left(\hat{\zeta}_n^{\sim 1}\right)^2$, can be obtained:

$$
\begin{aligned}
\dot{Z}_n^S \;=\;& \Delta_h Q_{Z^S} + \Delta_z Q_{Z^S} \\
& + \left[-\Delta_h Q_{Z^S} - \hat{\zeta}_n^{\sim 2} \nabla \cdot \widehat{(\underline{\mathbf{v}}\zeta)}_n^{\sim 2}) + \frac{1}{4}\hat{\zeta}_n^{\sim 2} \widehat{(\zeta\nabla\cdot\underline{\mathbf{v}})}_n^{\sim 2} + \frac{1}{4}\widehat{(\zeta^2)}_n^{\sim 2} \nabla\cdot \hat{\underline{\mathbf{v}}}_n^{\sim 2} \right] \\
& + \left[-\Delta_z Q_{Z^S} - \hat{\zeta}_n^{\sim 2} \frac{\partial \widehat{(w\zeta)}_n^{\sim 2})}{\partial z} + \frac{1}{4}\hat{\zeta}_n^{\sim 2} \widehat{\left(\zeta\frac{\partial w}{\partial z}\right)}_n^{\sim 2} + \frac{1}{4}\widehat{(\zeta^2)}_n^{\sim 2} \frac{\partial \hat{w}_n^{\sim 2}}{\partial z} \right] \\
& -\beta\; \hat{\zeta}_n^{\sim 2} \hat{v}_n^{\sim 2} + f\; \hat{\zeta}_n^{\sim 2} \frac{\partial \hat{w}_n^{\sim 2}}{\partial z} + \hat{\zeta}_n^{\sim 2} \; \widehat{\left(\zeta\frac{\partial w}{\partial z}\right)}_n^{\sim 2} \\
& + \hat{\zeta}_n^{\sim 2}\; \underline{\mathbf{k}} \cdot \widehat{\left(\frac{\partial \underline{\mathbf{v}}}{\partial z} \wedge \nabla w\right)}_n^{\sim 2} + F_{Z^S,z} + F_{Z^S,h},
\end{aligned} \tag{3.164}
$$

where

$$
\Delta_h Q_{Z^S} = -\nabla \cdot \left[\frac{1}{4}\hat{\zeta}_n^{\sim 2} \widehat{(\underline{\mathbf{v}}\zeta)}_n^{\sim 2} + \frac{1}{4}\widehat{(\zeta^2)}_n^{\sim 2} \hat{\underline{\mathbf{v}}}_n^{\sim 2} \right], \tag{3.165}
$$

$$
\Delta_z Q_{Z^S} = -\frac{\partial}{\partial z} \left[\frac{1}{4}\hat{\zeta}_n^{\sim 2} \widehat{(w\zeta)}_n^{\sim 2} + \frac{1}{4}\widehat{(\zeta^2)}_n^{\sim 2} \hat{w}_n^{\sim 2} \right]. \tag{3.166}
$$

It is also written in a pithy notation

$$
\begin{aligned}
\dot{Z}_n^S &= \Delta_h Q_{Z^S} + \Delta_z Q_{Z^S} + T_{Z^S,\partial_h\zeta} + T_{Z^S,\partial_z\zeta} + \\
& S_{Z^S,\beta} + S_{Z^S,f\nabla\cdot\underline{\mathbf{v}}} + TS_{Z^S,\zeta\nabla\cdot\underline{\mathbf{v}}} + TS_{Z^S,\text{tilt}} + \\
& F_{Z^S,z} + F_{Z^S,h}
\end{aligned} \tag{3.167}
$$

for later use.

3.10 Summary and discussion

Multiscale energy, vorticity, and enstrophy equations are derived from Eqs. (3.11)-(3.14). These equations form the basis of the multiscale energy and vorticity analysis (MS-EVA). Similar to its classical EVA counterpart (*i.e.*, the Pre-EVA developed by Spall, 1989), the MS-EVA admits processes like transport, buoyancy conversion, and diffusion, though perhaps appeared in different forms. In our formulation, transport of energy is in general much more complicated that it seems to be. A large-scale energy transport, for example, may involve fluxes contributed from all other scales. But the buoyancy conversion process is relatively simple. It acts as a protocol between kinetic and potential energies for the same scale window. No inter-scale contribution is present.

The distinction between the current and the classical versions of EVA lies in the emergence of those T-terms in our formulation. These terms represent energy transfer processes in the phase space. The transfer can be unambiguously separated from the transport via a careful classification of processes. When a flow is rotating with a nonstationary vorticity, kinetic energy begins to transfer and the transfer is perfect in the sense that rotation does no work to the whole system. Among other kinetic energy transfers is the one due to the vertical structure of horizontal velocity. It is actually also a rotation-caused transfer, considering that the vertical gradient of a horizontal velocity forms another component of vorticity in a three-dimensional flow. For the available potential energy, transfer (in fact, perfect transfer) occurs when the density gradient (either in x, y, or in z) is under variation. These processes, together with others, are summarized in Table 3.2.

Transfer terms can be further decomposed to unravel the complicated scale-scale interactions. This is the so-called interaction analysis. An interaction analysis of a transfer function T results in many interaction terms, which can be classified into the following four categories

$$\begin{array}{ll} T^{w_1 \to w}, & T^{w_2 \to w}, \\ T^{w_1 \oplus w_2 \to w}, & T^{w \to w}, \end{array}$$

each characteristic of an interaction process ($T^{w \to w}$ may be further decomposed if necessary). Here the superscript $w = 0, 1, 2$ stands for the large-, meso-, and sub-meso-scale windows,

Table 3.2: Symbols for the multiscale energy equations (Location n, Scale L., M., or S.).

Kinetic energy			Available potential energy	
$\dot{K}_n$	Time rate of change of kinetic energy at location n		$\dot{A}_n$	Time rate of change of available potential energy
$\Delta_h Q_K$	Horizontal kinetic energy advective working rate		$\Delta_h Q_A$	Horizontal APE advective working rate
$\Delta_z Q_K$	Vertical kinetic energy advective working rate		$\Delta_z Q_A$	Vertical APE advective working rate
$T_{K,\, w\partial_z}$	Rate of kinetic energy transfer due to the vertical advection		$T_{A,\, \partial_h \rho}$	Rate of APE transfer due to the horizontal gradient of density
$T_{K, \partial_z v}$	Rate of kinetic energy transfer due to the vertical shear of horizontal velocity		$T_{A,\, \partial_z \rho}$	Rate of APE transfer due to the vertical gradient of density
$T_{K,\, \zeta}$	Rate of kinetic energy transfer due to the time change of vorticity		b_n	Inverse buoyancy conversion
$\Delta_h Q_P$	Horizontal pressure working rate		TS_{A^M}	Rate of an imperfect APE transfer due to the stationary shear of the density profile
$\Delta_z Q_P$	Vertical pressure working rate		$F_{A,\, h}$	Horizontal diffusion
$-b_n$	Buoyancy conversion		$F_{A,\, z}$	Vertical diffusion
$F_{K,\, z}$	Vertical dissipation			
$F_{K,\, h}$	Horizontal dissipation			

respectively, and,

$$w_1 = (w+1) \bmod 3, \qquad w_2 = (w+2) \bmod 3.$$

Explicit expressions for these functions are given in Eqs. (3.131) - (3.134).

The MS-EVA developed is connected to the classical energetic analysis (e.g., that of Harrison and Robinson, 1978). When a two-scale window decomposition is performed, and if $j_0 = 0$ and a periodic extension is adopted, the meso-scale MS-EVA equations marginalize to the eddy energy equations in a classical framework, whereas the large-scale equations are essentially identical to their classical counterparts. This fact, together with the comparison made above, implies that our MS-EVA is a generalization of the classical energy analysis, either in terms of EVA, or in the sense of multiscale energetic studies.

Based on the same practice, multiscale enstrophy equations are also derived from the vorticity equation. These terms are summarized in Table 3.3. Interaction analyses can also be performed and the resulting processes are expressed in a way analogous to the above energetics.

To summarize, the multiscale energy analysis is schematized in Fig. 3.5. Arrows are employed to indicate energy flows, and box and discs to represent the KE and APE, respectively. From there, the APE and KE are related through buoyancy conversion within each window, whereas across the scales, the relations are established and maintained by different transfer processes.

A final remark on the energetic processes. It is well known that advection plays a role in redistributing energy in the physical space, and so does wave propagation. Now we know vorticity (to be precise, it should be the evolution of vorticity), which is a measure of rotation, acts to redistribute energy over a spectrum, i.e., to transfer energy across scales. The three mechanisms, advection, wave propagation, and rotation, can essentially make energy reach anywhere in either the physical or the phase space. It is not unreasonable to conjecture that, patterns formed in a geophysical fluid flow, complex as they may appear to be, could be a direct consequence of these energy redistributions. Examples will be seen in the next chapters, where we shall give the MS-EVA developed a validation and a real-ocean application.

Table 3.3: Symbols for the multiscale enstrophy equation (Location n, Scale L., M., or S.).

$\dot{Z}_n$: Time rate of change of the large-scale, mesoscale, or sub-mesoscale enstrophy				
$\Delta_h Q_Z$	Horizontal transport rate		$S_{Z,f\nabla\cdot\underline{v}}$	Source/sink of enstrophy due to horizontal divergence
$\Delta_z Q_Z$	Vertical transport rate		$TS_{Z,\zeta\nabla\cdot\underline{v}}$	Rate of Z transfer and generation due to the interaction between ς and horizontal divergence
$T_{Z,\partial_h\zeta}$	Rate of enstrophy transfer due to the horizontal variation of ς		$TS_{Z,\,\mathrm{titlt}}$	Rate of Z transfer and generation due to the vortex tube titlting
$T_{Z,\partial_z\zeta}$	Rate of enstrophy transfer due to the vertical variation of ς		$F_{Z,h}$	Horizontal diffusion rate
$S_{Z,\beta}$	β-effect-caused source or sink of enstrophy		$F_{Z,z}$	Vertical diffusion rate

Figure 3.5: A schematic of the multiscale energetics for a particular location. Arrows are used to indicate the directions of the energy flow, in both the physical space and the phase space, and labeled over these arrows are the processes associated with the flow. The symbols adopted are the same as those listed in Table 3.2, except that no mechanism (e.g., the evolution of vorticity and density gradient) is specified for the T-terms in their subscripts in order to have them represent transfer processes in a general sense. The superscripts of these T-terms stand for the corresponding interaction analyses whose interpretation is referred to the relevant parts of Section 3.7 (pp. 147-157). Note we have used bi-directional arrows in the superscripts to simplify the notation. Take $T_A^{0\leftrightarrow 1}$ for and example. It should be construed as $T_{A^M}^{0\to 1}$ and $T_{A^L}^{1\to 0}$. For clarity, transfers from the same window are not shown. From this diagram, we see that transports (ΔQ_{K^W}, ΔQ_{P^W}, ΔQ_{A^M}, for windows $W = L, M, S$) occur between different locations in the physical space, while transfers (the T-terms) mediate between scale windows in the phase space. The connection between the two forms of energy is established through the buoyancy conversion (positive if in the direction as indicated in the parenthesis), which invokes neither scale-scale interactions nor location-location energy exchange.

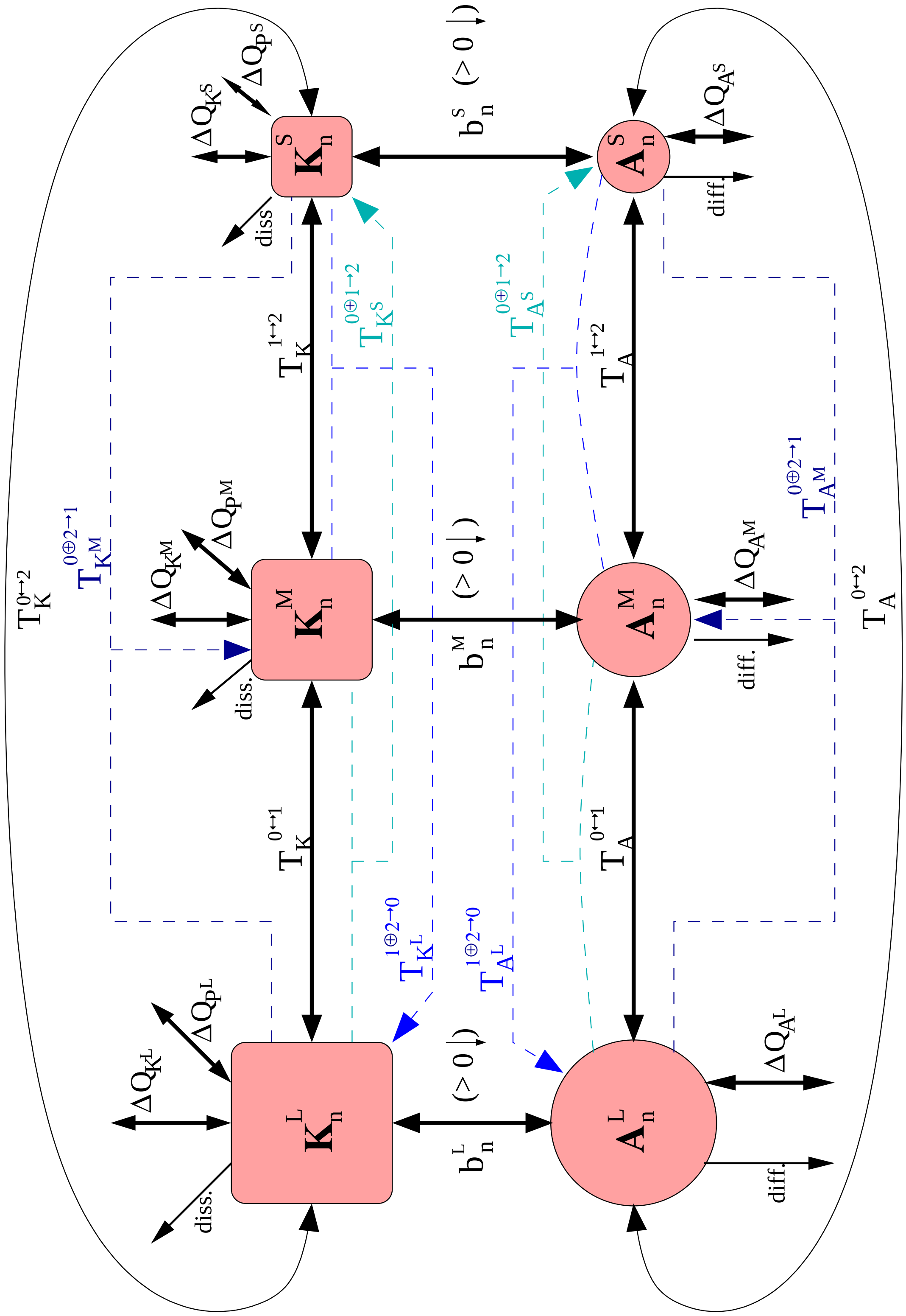

Figure 3.5: (Continued)

Appendix to Chapter 3

A3.1 Correction to the time derivative term

We have shown in §3.2.1 that there exists a correction term in the formulas with time derivatives. For a kinetic equation, this formula is

$$\underbrace{\hat{\delta}_n K_n}_{\text{(K)}} \quad - \quad \underbrace{(\Delta t)^2 \left(\hat{\delta}^2_{n^2} \hat{\underline{\mathbf{v}}}_n \cdot \hat{\delta}_n \hat{\underline{\mathbf{v}}}_n\right)}_{\text{(C)}}, \tag{A3.168}$$

where (C) is the correction term. Scale superscripts are omitted here since we don't want to limit the discussion to any particular scale window. Let's first do some nondimensional analysis so that a comparison is possible. Scale $\hat{\underline{\mathbf{v}}}_n$ with U, t with T, then

$$\text{Term (K)} \sim \frac{U^2}{T}, \qquad \text{Term (C)} \sim (\Delta t)^2 \frac{U}{T^2} \cdot \frac{U}{T} = (\Delta t)^2 \frac{U^2}{T^3}.$$

This enables us to evaluate the weight of (C) relative to (K):

$$\frac{\text{Term (C)}}{\text{Term (K)}} \sim \frac{(\Delta t)^2 U^2 / T^3}{U^2/T} = \left(\frac{\Delta t}{T}\right)^2.$$

Apparently, this ratio will become significant only when $T \sim \Delta t$, i.e., when the time scale is of the time step size. In our MS-EVA formulation, the correction term (C) is hence not significant for both large-scale and meso-scale equations. Fig. A3.6 confirms this conclusion. The correction (dashed line) is so small in either the left or middle plots that it is totally negligible. Only in the sub-mesoscale window can its effect be seen, which, as argued before, might be parameterized into the dissipation/diffusion.

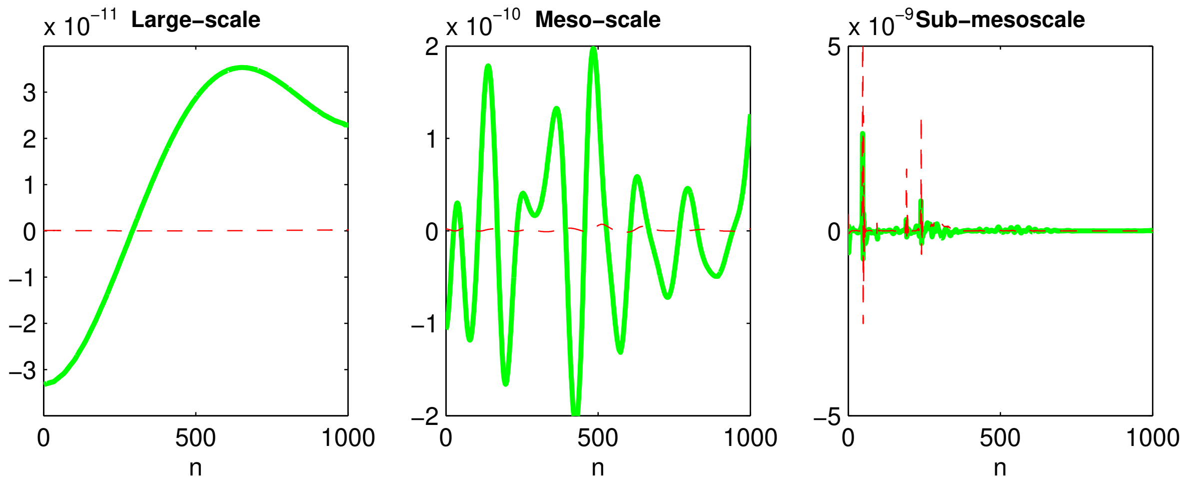

Figure A3.6: $\hat{\delta}_n K_n$ (thick solid) and its correction term (dashed) for the large-scale (left), meso-scale (middle), and sub-mesoscale (right) kinetic energy equations. Data source and parameter choice are the same as those of Fig. 3.2. The unit is $\mathrm{m}^2/\mathrm{s}^3$. Factor 2^{j_2} not multiplied.

A3.2 Some notes on the numerical computation

1. Discretization

The HOPS PE model adopts an Arakawa B-grid (Arakawa, 1977). The positioning of variables is schematized in Fig. A3.7, in which w_T and w_v are vertical velocities at tracer points and velocity points, respectively. MS-EVA equations are discretized in this grid, with uniform horizontal spacings Δx and Δy, and varying vertical step size Δz_k (Fig. A3.8). The discretization is realized with a central difference scheme. In the vertical direction, this requires some interpolation. Take the vertical transport term $\frac{\partial w\rho}{\partial z}$ as an example. It is

$$\left.\frac{\partial(w\rho)}{\partial z}\right|_k = \frac{1}{\Delta z_k}\left[w_{k-1}\cdot\frac{\rho_{k-1}\Delta z_k+\rho_k\Delta z_{k-1}}{\Delta z_k+\Delta z_{k-1}} - w_k\cdot\frac{\rho_{k+1}\Delta z_k+\rho_k\Delta z_{k+1}}{\Delta z_k+\Delta z_{k+1}}\right], \quad \text{(A3.169)}$$

where w is understood as w_T, the vertical velocity at tracer points.

2. Transform

Some fields, such as ρ, vary so little that the fluctuation is smaller than the mean value by several orders of magnitude. The huge energy concentrated in the mean makes a transform susceptible to big computational error because of the finite precision arithmetic. To overcome this difficulty, use decomposition (take ρ as an example)

$$\rho(t) = \bar{\rho} + \rho'(t)$$

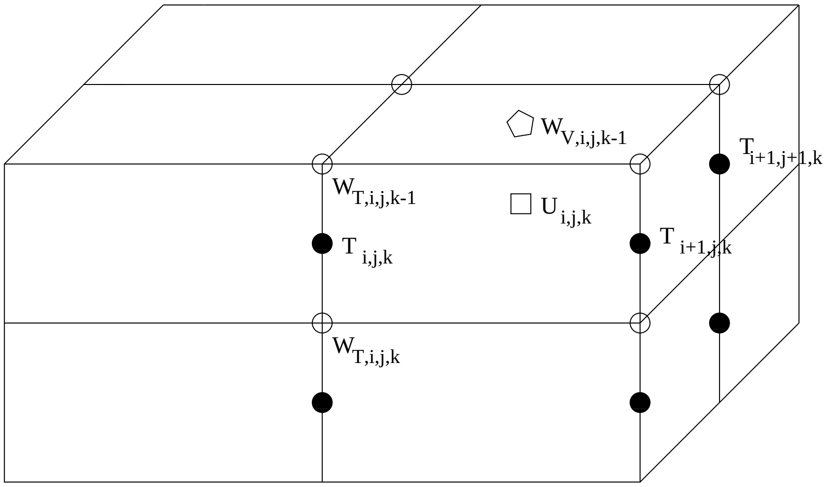

Figure A3.7: A diagram of the HOPS PE model discretization in an Arakawa B-grid. Indices i, j, k are used to position a location in the x, y, and z directions, respectively. T (filled circle) and U (square) represent the grid points for some tracer property T and horizontal velocity (u, v), whereas W_v (diamond) and W_T (circle) denotes respectively the vertical velocity w evaluated at the U and T points. The vertical dimension of the w array begins with level $1\frac{1}{2}$. At the ignored level $\frac{1}{2}$, w is always zero since a rigid-lid is imposed. For arrays of (u, v), the easternmost and northernmost values should be discarded.

to remove the mean part before any transform is performed. The final transform of ρ is obtained from that of ρ' through formulas

$$\begin{aligned}
\widehat{\rho}_n^{\sim 0} &= \widehat{(\bar{\rho})}_n^{\sim 0} + \widehat{(\rho')}_n^{\sim 0} \\
&= \int_0^{\varrho} \bar{\rho} \cdot \phi_n^{\varrho, j_2}(t)\ dt + \widehat{(\rho')}_n^{\sim 0} \\
&= \bar{\rho} \int_{\mathbb{R}} 2^{j_2/2} \phi(2^{j_2} t - n)\ dt + \widehat{(\rho')}_n^{\sim 0} \\
&= 2^{-j_2/2} \bar{\rho} + \widehat{(\rho')}_n^{\sim 0},
\end{aligned} \tag{A3.170}$$

and

$$\widehat{\rho}_n^{\sim 1} = \widehat{(\bar{\rho})}_n^{\sim 1} + \widehat{(\rho')}_n^{\sim 1} = \widehat{(\rho')}_n^{\sim 1} \tag{A3.171}$$

$$\widehat{\rho}_n^{\sim 2} = \widehat{(\bar{\rho})}_n^{\sim 2} + \widehat{(\rho')}_n^{\sim 2} = \widehat{(\rho')}_n^{\sim 2}. \tag{A3.172}$$

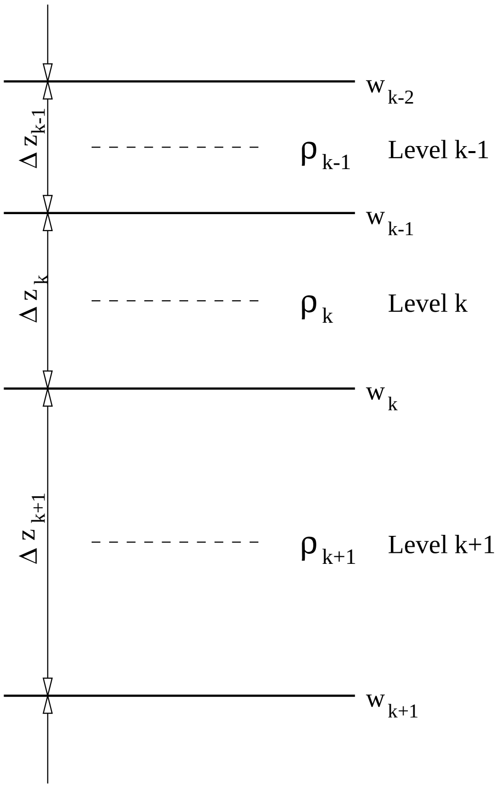

Figure A3.8: A diagram of the nonuniform vertical discretization used in the computation. Each dashed line marks the central level for a discretized vertical layer with some size Δz and bounded below and above by two solid horizontal lines. In each layer, the vertical velocity w is defined at the boundaries, while all other fields, denoted as ρ in the diagram, are evaluated at the center.

3. **Data communication**

More often than not, the dataset to be processed (produced by HOPS or other model systems) is huge, as the time series have to be long enough to yield an adequate multiscale decomposition. When programs are run on a connected clusters of computers within a local network, time spent on the data communication turns out significant in the overall expense. This problem becomes especially severe at a time when the traffic is heavy. But there is no way to alleviate it just by reducing program overheads. If possible, choose a physical transmission media (say, fiber) with better transmission characteristics (e.g., full-duplex), and adopt a digital subscriber line with high downstream (receiving) rate and upstream (sending) rate. In a word, do whatever possible to facilitate the communication between the server and clients. A through communication channel is expected to help improve the performance of MS-EVA applications significantly.

Bibliography

[1] Arakawa, A., and V. R. Lamb, 1977: Computational design of the basic dynamical processes of the UCLA general circulation model. *Methods in Comp. Phys.*, 174-265.

[2] Fournier, Aimé, 1999: Atmospheric energetics in the wavelet domain I: Governing equations and interpretation for idealized flows. *J. Atmos. Sci.*

[3] Goldstein, S., 1965: *Modern Developments in Fluid Dynamics*, Vol. II. Dover, New York, 371 pp.

[4] Hansen, A. R., 1981: A diagnostic study of the spectral energetics of blocking. Ph.D. dissertation, Iowa State University, 183pp.

[5] Harrison, D. E., 1979: Eddies and the general circulation of numerical model gyres: an energetic perspective. *Rev. Geophys. Space Phys.*, 17: 969-979.

[6] Harrison and Robinson, 1978: Energy analysis of open regions of turbulent flows-mean eddy energetics of a numerical ocean circulation experiment. *Dyn. Atmos. Oceans*, 2: 185-211.

[7] Hartman, P., 1982: *Ordinary Differential Equations* Birkhäusser Boston, 612pp.

[8] Xu, Qin, 1986: Generalized energetics for linear and nonlinear symmetric instabilities. *J. Atmos. Sci.*, 43, 972-984.

[9] Holton, James R., 1992: *An Introduction to Dynamic Meteorology.* Academic Press. 496pp.

[10] Iima, M., and S. Toh, 1995: Wavelet analysis of the energy transfer caused by convective terms: Application to the Burgers shock. *Phys. Rev.*, E, 52, 6189-6201.

[11] Jordan, D. W., and P. Smith, 1995: *Nonlinear Ordinary Differential Equation.* 2nd Ed., Claredon, Oxford University, 381 pp.

[12] Kanamitsu, M., T. Krishnamurti, and C. Depradine, 1972: On scale interactions in the tropics during northern summer. *J. Atmos. Sci.*, 29, 698-706.

[13] Khac, Minh Do, 1998: *Use of Orthogonal Wavelets in the Analysis of Two- and Three-Dimensional Inhomogeneous Turbulent Flows.* European Space Agency Technical Translation of ONERA NT 1997-04.

[14] Lesieur, M., 1990: *Turbulence in Fluids: Stochastic and Numerical Modeling.* 2d ed., Kluwer Academic Publishers, 400 pp.

[15] Lorenz, E. N., 1960: Energy and numerical weather prediction. *J. Atmos. Sci.*, 19, 39-51.

[16] McComb, W. D., 1991: *The Physics of Fluid Turbulence.* Oxford Science Publications. 572 pp.

[17] Monin, A. S., and A. M. Yaglom, 1971: *Statistical Fluid Mechanics*, Vol. 1. MIT Press, Cambridge, Mass., 769 pp.

[18] Ohkitani, Koji, and Shigeo Kida, 1992: Triad interactions in a forced turbulence. *Phys. Fluids* A 4(4). 794-802.

[19] Pedlosky, J., 1979: *Geophysical Fluid Dynamics.* Springer-Verlag.

[20] Pinardi, N., and A. R. Robinson, 1986: Quasigeostrophic energetics of open ocean regions. *Dyn. Atmos. Oceans* 10, 185-219.

[21] Rajaee, Mojtaba, and Sture Karlsson, 1992: On the Fourier space decomposition of free shear flow measurements and mode degeneration in the pairing process. *Phys. Fluids* A 4(2). 321-339.

[22] Salmon, Rick, 1998: *Lectures on Geophysical Fluid Dynamics.* Oxford University Press. 378pp.

[23] Saltzman, B., 1957: Equations governing the energetics of the larger scales of atmospheric turbulence in the domain of wave number. *J. Meteor.*, 14, 513-523.

[24] Spall, M., 1989: Regional primitive equation modeling and analysis of the POLYMODE data set. *Dyn. Atmos. Oceans*, 14, 125-174.

[25] Spall, M. A., and A. R. Robinson, 1990: Regional primitive equation studies of the Gulf Stream meander and ring formation region. *J. Phys. Oceanogr.*, 20, 985-1016.

[26] Waleffe, Fabian, 1992: The nature of triad interactions in homogeneous turbulence. *Phys. Fluids* A 4 (2). 350-363.

Chapter 4

MS-EVA validation and dynamic fingerprinting with the MS-EVA

4.1 Introduction

We have completed the formulation of the MS-EVA, the multiscale energy and vorticity analysis. This chapter serves to give it a validation with some highly idealized models, and in the meantime explore with this validated tool the associated energetic structures, which will help identify processes of the same kind in a real ocean or atmosphere.

The stability problem example, which was introduced in the beginning of this thesis, is chosen to achieve the above goals. Following this section is a brief review within the classical framework, where barotropic and baroclinic instabilities are introduced. In §4.3, these two fundamental processes are related to the MS-EVA, and one will see that the relationship is established through two combined MS-EVA transfer terms, which are denoted respectively by BC and BT. This section shows how the MS-EVA can be applied to a specific problem.

Processes represented by BC and BT are supposed to lead directly to the baroclinic and barotropic instabilities. This fact is the starting point to test MS-EVA diagnosis. In §4.4 and §4.5, we pick two well-known models, each admitting only one process, to validate what we have formulated before, and, along with the validation, to investigate the MS-EVA structures characteristic of the two processes. The last section is a brief summary and a discussion of how the procedure introduced in this chapter can be modified for a real regional ocean or atmosphere.

4.2 The classical linear stability problem

This section introduces two types of instabilities, barotropic instability and baroclinic instability, for a geophysical fluid flow with an idealized configuration. The treatment is standard and can be found in many textbooks (e.g., Pedlosky, 1979). The problem is re-derived here because the existing formulations are mostly with a quasi-geostrophic (QG) flow, which might not quite fit our primitive equation model. Indeed, as we will see later, an additional term will come into the final result if no QG approximation is invoked.

Consider an ideal (frictionless) zonal flow channeled between $y = \pm L$. The governing equations are (3.4)-(3.7),[1] with the dissipation and diffusion terms neglected. Written in a component form, they are

$$\frac{\partial u}{\partial t} = -u\frac{\partial u}{\partial x} - v\frac{\partial u}{\partial y} - w\frac{\partial u}{\partial z} + fv - \frac{1}{\rho_0}\frac{\partial P}{\partial x} \tag{4.1}$$

$$\frac{\partial v}{\partial t} = -u\frac{\partial v}{\partial x} - v\frac{\partial v}{\partial y} - w\frac{\partial v}{\partial z} - fu - \frac{1}{\rho_0}\frac{\partial P}{\partial y} \tag{4.2}$$

$$0 = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} \tag{4.3}$$

$$\frac{\partial P}{\partial z} = -\rho g \tag{4.4}$$

$$\frac{\partial \rho}{\partial t} = -u\frac{\partial \rho}{\partial x} - v\frac{\partial \rho}{\partial y} - w\frac{\partial \rho}{\partial z}. \tag{4.5}$$

In this system, field ρ (and hence P) is understood to be de-meaned. That is to say, the constant part ρ_0 has been removed (but the vertical background structure is not touched). Though in different variables, the equation are still of the same form as those of (3.11)-(3.14).

The boundary conditions are no-normal flux across $y = \pm L$, and rigid at the bottom $z = -H$ and surface $z = 0$. In the x direction, the boundary is assumed to be cyclic or periodic with period L.

[1] We do not adopt the set (3.11)-(3.14) because here the background density field, $\bar{\rho}$ (see later), has a dependence on y as well as z.

In order to linearize the problem, it is necessary to scale the variables so that a term-by-term comparison is possible. As usual, let

$$\begin{aligned}
(x,y) &\sim L, & z &\sim H, \\
(u,v) &\sim U, & w &\sim U\frac{H}{L}, \\
T &\sim \frac{L}{U}, & f &\sim f_0, \\
P \sim \Delta P &= \rho_0 f_0 LU, & \rho &\sim \Delta\rho = \frac{\rho_0 f_0 LU}{gH},
\end{aligned}$$

then the governing equations become,[2] after non-dimensionalization,

$$R_0 \frac{\partial u}{\partial t} = -R_0 \left(u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} + w\frac{\partial u}{\partial z} \right) + fv - \frac{\partial P}{\partial x}, \tag{4.6}$$

$$R_0 \frac{\partial v}{\partial t} = -R_0 \left(u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} + w\frac{\partial v}{\partial z} \right) - fu - \frac{\partial P}{\partial y}, \tag{4.7}$$

$$0 = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z}, \tag{4.8}$$

$$\rho = -\frac{\partial P}{\partial z}, \tag{4.9}$$

$$\frac{\partial \rho}{\partial t} = -u\frac{\partial \rho}{\partial x} - v\frac{\partial \rho}{\partial y} - w\frac{\partial \rho}{\partial z}, \tag{4.10}$$

where $R_0 = U/f_0 L$ is the Rossby number. In many problems of interest, R_0 is very small. But since what we consider here is a generic PE model, its size is not prescribed. With the scaled variables, the boundary conditions are

$$w = 0 \qquad \text{at} \quad z = 0, -1; \tag{4.11}$$

$$v = 0 \qquad \text{at} \quad y = \pm 1; \tag{4.12}$$

and cyclic in the x-direction.

Suppose we have a basic flow

$$(\bar{u}, \bar{v}, \bar{w}) = (\bar{u}(y,z), 0, 0), \tag{4.13}$$

which, with the dependence on both y and z, provides a background potential vorticity that admits both barotropic instability and baroclinic instability, the two fundamental instabilities for a geofluid flow. The basic pressure and density are determined through geostrophic and

[2] From now on to the end of this section, all variables are nondimensionalized.

thermal wind relations

$$\frac{\partial \bar{P}}{\partial y} = -\bar{u}, \tag{4.14}$$

$$\frac{\partial \bar{\rho}}{\partial y} = -\frac{\partial}{\partial z}\frac{\partial \bar{P}}{\partial y} = f\frac{\partial \bar{u}}{\partial z}. \tag{4.15}$$

Equations (4.13) - (4.15) form the background system for a linear instability problem. It has a general configuration which may be simplified to obtain many well-known formulations. For example, if the problem is considered on an f-plane ($f = 1$), and

$$\frac{\partial \bar{u}}{\partial z} = \Lambda = \text{const}, \qquad \frac{\partial \bar{u}}{\partial y} = 0,$$

it is nothing but the *Eady model*; if the geometry is a β-plane ($f = 1 + \beta y$ with β a constant), and

$$h^{-1} = -\frac{1}{\bar{\rho}}\frac{\partial \bar{\rho}}{\partial z}$$

is a constant, with other conditions the same as above, then it becomes the *Charney model.*

Now perturb the basic system with a small disturbance

$$(u', v', w', \rho', P')$$

such that

$$\max\left|\frac{u'}{\bar{u}}\right| = \epsilon \ll R_0. \tag{4.16}$$

The inequality seems to be a very stringent constraint on the perturbation, but, as we will see later, it is crucial to the linearization of those governing equations. Fortunately it can be justified if only disturbance growth at initial stage is considered.[3] One may argue that, based on the continuity of the field functions, the system can always be made satisfy the constraint so long as the evolution time is short enough.

With the perturbation superimposed on the background, the fields are now

$$\left.\begin{array}{lll} u = \bar{u} + u', & v = v', & w = w', \\ \rho = \bar{\rho} + \rho', & P = \bar{P} + P'. & \end{array}\right\} \tag{4.17}$$

[3]This is a limitation of linear stability theory. The disturbances may grow so fast that the theory soon becomes invalid.

Substitution of them into the governing equations yields, with terms of order higher than $O(\epsilon)$ dropped,

$$\begin{aligned}
R_0\frac{\partial u'}{\partial t} &= -R_0\left(\bar{u}\frac{\partial u'}{\partial x} + v'\frac{\partial \bar{u}}{\partial y} + w'\frac{\partial \bar{u}}{\partial z}\right) + fv' - \frac{\partial P'}{\partial x}, && (4.18)\\
R_0\frac{\partial v'}{\partial t} &= -R_0\bar{u}\frac{\partial v'}{\partial x} - fu' - \frac{\partial P'}{\partial y}, && (4.19)\\
0 &= \frac{\partial u'}{\partial x} + \frac{\partial v'}{\partial y} + \frac{\partial w'}{\partial z}, && (4.20)\\
\rho' &= -\frac{\partial P'}{\partial z}, && (4.21)\\
\frac{\partial \rho'}{\partial t} &= -\bar{u}\frac{\partial \rho'}{\partial x} - v'\frac{\partial \bar{\rho}}{\partial y} - w'\frac{\partial \bar{\rho}}{\partial z}. && (4.22)
\end{aligned}$$

The kinetic energy of the disturbance (eddy kinetic energy) is defined to be

$$K^M = \frac{1}{2}\left(\overline{u'^2} + \overline{v'^2}\right), \tag{4.23}$$

where the mean is performed over the whole sample space.[4] The evolution of K^M is obtained by a multiplication of (4.18) by u' plus a multiplication of (4.19) by v', followed by an averaging over the entire ensemble,

$$\begin{aligned}
R_0\frac{\partial K^M}{\partial t} &= -R_0\left(\frac{\partial (K^M\bar{u})}{\partial x} + \overline{u'w'}\frac{\partial \bar{u}}{\partial z} + \overline{u'v'}\frac{\partial \bar{u}}{\partial y}\right)\\
&\quad - \left(\frac{\partial \overline{u'P'}}{\partial x} + \frac{\partial \overline{v'P'}}{\partial y} + \frac{\partial \overline{w'p'}}{\partial z}\right) - \overline{w'\rho'}. && (4.24)
\end{aligned}$$

Let

$$\left\langle R_0K^M\right\rangle^\Omega = \frac{1}{V_\Omega}\iiint_\Omega R_0K^M dV,$$

namely the perturbation energy averaged over the model domain

$$\Omega = \{(x,y,z) \ \mid x\in[0,1] \text{ and } \text{ cyclic}, \ y\in[-1,1], \ z\in[-1,0]\},$$

which has a volume V_Ω. Take this spatial average on both sides of the above equation for K^M to get

$$\frac{\partial\left\langle R_0K^M\right\rangle^\Omega}{\partial t} = -R_0\left\langle\overline{u'w'}\frac{\partial \bar{u}}{\partial z} + \overline{u'v'}\frac{\partial \bar{u}}{\partial y}\right\rangle^\Omega - \left\langle\overline{w'\rho'}\right\rangle^\Omega. \tag{4.25}$$

[4]Here the sample space is a statistical concept, and the mean over it is independent of time and the physical space.

Likewise, define the available potential energy of the perturbation field:

$$A^M = \frac{1}{-\partial\bar{\rho}/\partial z} \frac{\overline{\rho'^2}}{2}. \tag{4.26}$$

Note $-\frac{\partial\bar{\rho}}{\partial z} > 0$ for a stably stratified ocean. The positive-definiteness of A^M is then guaranteed. Multiply (4.22) by $\frac{\rho'}{-\partial\bar{\rho}/\partial z}$ and then perform an ensemble average. The evolution of A^M is obtained thereby:

$$\frac{\partial A^M}{\partial t} = -\frac{\partial \bar{u} A^M}{\partial x} - \frac{1}{-\partial\bar{\rho}/\partial z}\overline{v'\rho'}\frac{\partial\bar{\rho}}{\partial y} + \overline{w'\rho'}. \tag{4.27}$$

Integration of this equation over Ω, followed by an application of the specified boundary conditions and a division by V_Ω, gives

$$\frac{\partial \left\langle A^M \right\rangle^\Omega}{\partial t} = -\left\langle \frac{1}{-\partial\bar{\rho}/\partial z}\overline{v'\rho'}\frac{\partial\bar{\rho}}{\partial y} \right\rangle^\Omega + \left\langle \overline{w'\rho'} \right\rangle^\Omega, \tag{4.28}$$

where, as before,

$$\left\langle A^M \right\rangle^\Omega = \frac{1}{V_\Omega} \iiint_\Omega A^M dV,$$

is the spatially averaged perturbation APE. The sum of $\left\langle A^M \right\rangle^\Omega$ and $\left\langle R_0 K^M \right\rangle^\Omega$ measures the capacity of perturbation. Write it as $\left\langle E^M \right\rangle^\Omega$. Its evolution is obtained by adding Eq.(4.28) to Eq.(4.25),

$$\frac{\partial \left\langle E^M \right\rangle^\Omega}{\partial t} = -\left\langle \frac{f}{-\partial\bar{\rho}/\partial z}\overline{v'\rho'}\frac{\partial\bar{u}}{\partial z} \right\rangle^\Omega - R_0 \left\langle \overline{u'v'}\frac{\partial\bar{u}}{\partial y} + \overline{u'w'}\frac{\partial\bar{u}}{\partial z} \right\rangle^\Omega. \tag{4.29}$$

In arriving at this equation, the thermal wind relation (4.15) has been used.

Equation (4.29) is virtually the same as the equation (7.3.11) of Pedlosky (1979), except for the introduction of a second term in the second angle bracket on the right hand side. The reason why Pedlosky's formulation lacks this term is that he begins with the quasi-geostrophic assumption, which basically excludes vertical advection in the momentum equations at order $O(R_0)$.[5] In the primitive equation model, this does not necessarily hold, however. The representation in our problem is the retaining of the extra term $-\left\langle R_0 \overline{u'w'}\frac{\partial\bar{u}}{\partial z} \right\rangle^\Omega$. In fact,

[5]This follows from the fact that, to the lowest order, the flow is geostrophic, and geostrophic flow is horizontally nondivergent (i.e., $\frac{\partial w}{\partial z} = 0$ to the lowest order). With the rigid lid boundaries specified above and below, this yields a zero w (to the lowest order) throughout the water column, which eliminates the vertical advection mechanism in the first order momentum equations.

it is possible to estimate, in terms of the known scaling numbers, the relative importance of this term (call it Term (II)) and its counterpart, Term (I)= $-\left\langle \frac{f}{-\partial\bar{\rho}/\partial z}\overline{v'\rho'}\right\rangle^{\Omega}$, if the vertical structure of $\bar{\rho}$ is rewritten as a factor times the buoyancy frequency N^2. Since

$$-\left(\frac{g}{\rho_0}\frac{\Delta\rho}{H}\right)\frac{\partial\bar{\rho}}{\partial z} = \left(N_0^2\right)N^2,$$

(where those quantities in the parentheses are all dimensional and N^2 is the scaled buoyancy frequency), and according to the scaling introduced before, $\Delta\rho = \rho_0 f_0 LU/gH$, we have

$$-\frac{\partial\bar{\rho}}{\partial z} = \frac{N_0^2 H^2}{f_0 LU}\cdot N^2 = \frac{N_0^2 H^2}{f_0^2 L^2}\cdot\frac{1}{R_0}\cdot N^2.$$

Now let the buoyancy frequency be correctly scaled (i.e., $N^2 \sim O(1)$), the ratio of Term (II) to Term (I) is then

$$\begin{aligned}\frac{\text{Term }(II)}{\text{Term }(I)} &\sim R_0\times\frac{1}{R_0}\frac{N_0^2 H^2}{f_0^2 L^2} \\ &= \frac{N_0^2 H^2}{f_0^2 L^2}.\end{aligned} \tag{4.30}$$

This is exactly the reciprocal of the Γ used in the formulation of Pinardi and Robinson (1986), or the stratification number S in Pedlosky (1979). Note

$$\Gamma = L^2/\frac{N_0^2 H^2}{f_0^2} \equiv \frac{L^2}{L_D^2},$$

where L_D is the *Rossby deformation radius.* As argued by Pedlosky (1979) (pp. 454 - 455), length scales of the order L_D are most preferred for fluctuations to be fed with the mean potential energy. In other words, $\Gamma \sim O(1)$ for most of the meso-scale atmospheric or oceanic motions, if these motions are generated through baroclinic instabilities. For this reason, Term (II) could be of the same importance as Term (I) in a PE model. To make the scaling more clear, Eq. (4.29) is recast as

$$\frac{\partial\left\langle E^M\right\rangle^{\Omega}}{\partial t} = -R_0\left\langle \Gamma\frac{f}{N^2}\overline{v'\rho'}\frac{\partial\bar{u}}{\partial z}\right\rangle^{\Omega} - R_0\left\langle\overline{u'v'}\frac{\partial\bar{u}}{\partial y} + \overline{u'w'}\frac{\partial\bar{u}}{\partial z}\right\rangle^{\Omega}. \tag{4.31}$$

A system is unstable if $\left\langle E^M\right\rangle^{\Omega}$ increases. From Eq. (4.31), there exist two mechanisms that may cause an instability. They are, when properly grouped, represented by the two spatial averages on the right hand side (see, for example, Pedlosky, 1979; Robinson, 1996). The first average comes from the equation for A^M. The perturbation growth due to this mechanism is called a *baroclinic instability.* The second average, which includes two parts here, represents the rate of the energy extracted by the Reynolds stress against the background velocity profile. It has a kinetic energy origin, and the instability brought about by it is said to be *barotropic.* From Eq. (4.31), a barotropic instability is possible only when the perturbation streamlines are aligned against the basic flow profile (cf. Pedlosky, 1979. p. 436).

It is a common practice to define barotropic instability and baroclinic instability according to the structure of the basic flow $\bar{u}$. People usually refer to the former as perturbation growth on a background field without vertical dependence, while defining the latter to be instability with $\bar{u}$ as function of z only. In the quasi-geostrophic formalism, the term involving $\overline{u'w'}$ in (4.31) diappears (e.g., Pedlosky, 1979; Robinson, 1996), and the dependence of $\bar{u}$ corresponds precisely to the energy source for the perturbation growth. This definition is thence equivalent to the concept of instability introduced above. In a PE model, however, things are more complicated. With the introduction of $\overline{u'w'}$, the one-to-one correspondence between instability type and mean flow dependence is no longer valid. In addition to $\left\langle A^M \right\rangle^\Omega$, barotropic growth may also be fed with energy extracted against a vertical profile of $\bar{u}$. In general, the instability type of a PE model flow cannot be told simply by the structure of its mean flow alone.

4.3 MS-EVA in wave instability identification

Our MS-EVA is developed on basis of a dimensional primitive equation model. It will be more convenient to have the energy equation, Eq. (4.31), also expressed in its dimensional form for comparison:

$$\frac{\partial \left\langle E^M \right\rangle^\Omega}{\partial t} = -\left\langle \frac{fg}{\rho_0 N^2} \overline{v'\rho'} \frac{\partial \bar{u}}{\partial z} \right\rangle^\Omega - \left\langle \overline{u'v'} \frac{\partial \bar{u}}{\partial y} + \overline{u'w'} \frac{\partial \bar{u}}{\partial z} \right\rangle^\Omega . \tag{4.32}$$

Mnemonically this is written

$$\partial_t \left\langle E^M \right\rangle^\Omega = \langle BC^* \rangle^\Omega + \langle BT_H^* \rangle^\Omega + \langle BT_V^* \rangle^\Omega , \tag{4.33}$$

where

$$BC^* = -\frac{fg}{\rho_0 N^2} \overline{v'\rho'} \frac{\partial \bar{u}}{\partial z}, \tag{4.34}$$

$$BT_H^* = -\overline{u'v'} \frac{\partial \bar{u}}{\partial y}, \tag{4.35}$$

$$BT_V^* \quad = \quad -\overline{u'w'}\frac{\partial \bar{u}}{\partial z}, \tag{4.36}$$

with BT^* and BC^* symbolizing the mechanisms that lead to a barotropic instability and baroclinic instability, respectively. The subscripts V and H of BT^* stand for the vertical and horizontal profiles of the basic flow against which the Reynolds stress extracts energy to fuel the barotropic perturbation growth. When the basic flow lacks a vertical structure, both $\langle BC^* \rangle^\Omega$ and $\langle BT_V^* \rangle^\Omega$ vanish, and the system will be unstable if $\langle BT_H^* \rangle^\Omega > 0$. But if what is lacking in the basic flow structure is a horizontal gradient, disturbances could grow with either positive or negative $\langle BC^* \rangle^\Omega$ or $\langle BT_H^* \rangle^\Omega$, provided their sum, $\langle BC^* + BT_H^* \rangle^\Omega$, is greater than zero.

It should be pointed out that, in practice, a statistical ensemble is usually not feasible. We have to use another form of average in place of the ensemble mean during the derivation of Eq. (4.33). If replaced by a spatial average, all the space derivatives on the right hand side will be influenced and no such simple form as (4.33) would result; if replaced by a temporal average, only one term, i.e., the $\partial_t \left\langle E^M \right\rangle^\Omega$ on the left hand side, needs some modification. The modified $\partial_t \left\langle E^M \right\rangle^\Omega$ is just as that in Harrison and Robinson (1978), which still has an interpretation of eddy energy change with respect to time if manipulated in the same way as we did in the preceding chapter.

We now proceed to establish a connection of Eq. (4.33) to the cross-scale parts of the MS-EVA terms in (3.63) and (3.72). We first examine the interaction analysis with linearizations applied, then study how the MS-EVA kinetic and potential energetics can be related to (4.33). In the last subsection, some criteria are summarized for the identification of barotropic and baroclinic instabilities.

4.3.1 Linearized interaction analysis

Recall for a linear instability problem, only the meso-scale part of a two-scale decomposition is considered. For a meso-scale transfer term [cf., (3.124) and (3.125) in Chapter 3]

$$T(1,n) = \widehat{\mathcal{Z}}_n^{\sim 1} \cdot \widehat{pq}_n^{\sim 1}, \qquad \forall\ \mathcal{Z}, p, q \in V_{\varrho, j_2},$$

the interaction analysis will be

$$T(1,n) = T_n^{0\to 1} + T_n^{1\to 1} = T_n^{0\to 1} + \widehat{\mathcal{Z}}_n^{\sim 1} \cdot \widehat{(p^{\sim 1} q^{\sim 1})}_n^{\sim 1}. \tag{4.37}$$

Connection with the classically formulated energetics is possible when $j_0 = 0$ and a periodic extension is used. In this case,

$$q^{\sim 0} = \bar{q}, \qquad q^{\sim 1} = q' \qquad \forall\, q \in V_{\varrho, j_2},$$

so the marginalization property (2.70) becomes

$$\mathcal{M}_n \widehat{p}_n^{\sim 1} \widehat{q}_n^{\sim 1} = \overline{p'q'}, \qquad \forall\, p, q \in V_{\varrho, j_2}, \tag{4.38}$$

where $\mathcal{M}_n$ is over $\{n \mid n \in \mathbb{Z}, 2^{-j_2} n \in [0,1)\}$.

Marginalization of (4.37) yields

$$\begin{aligned}
\mathcal{M}_n T(1,n) &= \mathcal{M}_n T_n^{0\to 1} + \overline{\mathcal{Z}' \cdot (p'q')'} \\
&= \mathcal{M}_n T_n^{0\to 1} + \overline{\mathcal{Z}' \cdot \left(p'q' - \overline{p'q'}\right)} \\
&= \mathcal{M}_n T_n^{0\to 1} + \overline{\mathcal{Z}' p'q'} \\
&\approx \mathcal{M}_n T_n^{0\to 1} \qquad \text{by linearization.}
\end{aligned}$$

So, for a meso-scale transfer, only the cross-scale part will be left in the linearized energetics when marginalized. That is,

$$\mathcal{M}_n T_n^{0\to 1} = \mathcal{M}_n T(1,n), \tag{4.39}$$

where "=" is understood to be an equivalence relation up to linearization.

4.3.2 Kinetic energetics

In the meso-scale kinetic equation (3.63), $K_n^M = \frac{1}{2} \widehat{\underline{\mathbf{v}}}_n^{\sim 1} \cdot \widehat{\underline{\mathbf{v}}}_n^{\sim 1}$. Clearly $\mathcal{M}_n K_n^M = K^M = \frac{1}{2} \overline{\underline{\mathbf{v}}' \cdot \underline{\mathbf{v}}'}$, a direct result of the marginalization property. So

$$\mathcal{M}_n \left\langle K_n^M \right\rangle^{\Omega} = \left\langle K^M \right\rangle^{\Omega} \tag{4.40}$$

which forms the kinetic part of the $\left\langle E^M \right\rangle^{\Omega}$ in Eq. (4.33).

Among the other terms of Eq. (3.63), those with the symbol T ($T_{K^M,\zeta}$, $T_{K^M, w\partial_z}$, and $T_{K^M, \partial_z v}$) have been argued to characterize two transfer processes. By what we have defined, it is anticipated that some instabilities of interest will be represented by these terms. To see if it is true, first marginalize $T_{K^M,\zeta}$ (location dependence suppressed for simplicity):

$$\begin{aligned}
\mathcal{M}_n T^{0\to1}_{K^M,\zeta} &= \mathcal{M}_n T_{K^M,\zeta} = -\mathcal{M}_n \widehat{\underline{\mathbf{v}}}_n^{\sim1} \cdot \underline{\mathbf{k}} \wedge \widehat{(\zeta\underline{\mathbf{v}})}_n^{\sim1} \\
&= -\overline{\underline{\mathbf{v}}' \cdot \underline{\mathbf{k}} \wedge (\zeta\underline{\mathbf{v}})'} \\
&= -\overline{\underline{\mathbf{v}}' \cdot \underline{\mathbf{k}} \wedge (\zeta\underline{\mathbf{v}} - \overline{\zeta\underline{\mathbf{v}}})} \\
&= -\overline{\underline{\mathbf{v}}' \cdot \underline{\mathbf{k}} \wedge (\zeta'\bar{\underline{\mathbf{v}}} + \bar{\zeta}\underline{\mathbf{v}}' + \overline{\zeta'\underline{\mathbf{v}}'})} \\
&= -\overline{\underline{\mathbf{v}}' \cdot \underline{\mathbf{k}} \wedge (\zeta'\bar{\underline{\mathbf{v}}})} \\
&= -\overline{v'\left(\frac{\partial v'}{\partial x} - \frac{\partial u'}{\partial y}\right)}\,\bar{u}.
\end{aligned}$$

So,

$$\begin{aligned}
\iiint_\Omega \mathcal{M}_n T^{0\to1}_{K^M,\zeta}\, dV &= -\iiint_\Omega \overline{v'\left(\frac{\partial v'}{\partial x} - \frac{\partial u'}{\partial y}\right)}\,\bar{u}\, dV \\
&= \iiint_\Omega \overline{v'\frac{\partial u'}{\partial y}}\,\bar{u}\, dV - \iiint_\Omega \frac{\partial(\bar{u}\ \overline{v'^2}/2)}{\partial x}\, dV.
\end{aligned}$$

After integrating by parts followed by an application of the boundary conditions (no-slip at northern and southern flanks; rigid-lid on the top and bottom; cyclic in the x-direction), this integral is reduced to

$$\begin{aligned}
\iiint_\Omega \mathcal{M}_n T^{0\to1}_{K^M,\zeta}\, dV &= -\iiint_\Omega \overline{u'\frac{\partial(\bar{u}v')}{\partial y}}\, dV \\
&= -\iiint_\Omega \overline{u'v'}\frac{\partial \bar{u}}{\partial y}\, dV - \iiint_\Omega \overline{u'\frac{\partial v'}{\partial y}}\bar{u}\, dV \\
&= -\iiint_\Omega \overline{u'v'}\frac{\partial \bar{u}}{\partial y}\, dV + \iiint_\Omega \overline{u'\left(\frac{\partial u'}{\partial x} + \frac{\partial w'}{\partial z}\right)}\,\bar{u}\, dV \qquad \text{(by continuity)} \\
&= -\iiint_\Omega \overline{u'v'}\frac{\partial \bar{u}}{\partial y}\, dV + \iiint_\Omega \overline{u'\frac{\partial w'}{\partial z}}\,\bar{u}\, dV. \qquad (4.41)
\end{aligned}$$

Likewise, the marginalization of $T^{0\to1}_{K^M,w\partial_z}$ is

$$\begin{aligned}
\mathcal{M}_n T^{0\to1}_{K^M,w\partial_z} &= \mathcal{M}_n T_{K^M,w\partial_z} \\
&= \mathcal{M}_n \widehat{w}_n^{\sim1} \frac{\partial \widehat{K}_n^{\sim1}}{\partial z} \\
&= \overline{w'\frac{\partial K'}{\partial z}}.
\end{aligned}$$

Notice that here

$$K' = \frac{1}{2}\left[(\bar{u}+u')^2 + v'^2 - \bar{u}^2\right],$$

different from the K_n^M in the meso-scale energy equations. After linearization (terms dropped at order higher than 2), the above equation is integrated over Ω to get

$$\begin{aligned}
\iiint_\Omega \mathcal{M}_n T^{0\to1}_{K^M,w\partial_z} \, dV &= \iiint_\Omega \overline{w' \frac{\partial}{\partial z}(\bar{u}u')} \, dV \\
&= \iint \overline{\bar{u}w'u'}\Big|_{z=-H}^{z=0} dxdy - \iiint \overline{u'\frac{\partial w'}{\partial z}} \, \bar{u} \, dV \\
&= -\iiint_\Omega \overline{u'\frac{\partial w'}{\partial z}} \, \bar{u} \, dV. \qquad (4.42)
\end{aligned}$$

Put (4.42) and (4.41) together,

$$\iiint_\Omega \mathcal{M}_n (T^{0\to1}_{K^M,\zeta} + T^{0\to1}_{K^M,w\partial_z}) \, dV = -\iiint_\Omega \overline{u'v'}\frac{\partial \bar{u}}{\partial y} \, dV, \qquad (4.43)$$

or

$$\mathcal{M}_n \left\langle T^{0\to1}_{K^M,\zeta} + T^{0\to1}_{K^M,w\partial_z} \right\rangle^\Omega = -\left\langle \overline{u'v'}\frac{\partial \bar{u}}{\partial y} \right\rangle^\Omega \qquad (4.44)$$

which is exactly the term $\langle BT_H^* \rangle^\Omega$ in Eq. (4.33). That is to say, $T^{0\to1}_{K^M,\zeta}$ and $T^{0\to1}_{K^M,w\partial_z}$ together marginalize to the energy needed for a barotropic instability due to BT_H^*.

The other T-term in the kinetic equation is $T_{K^M,\partial_z v}$. We anticipate that its cross-scale part, $T^{0\to1}_{K^M,\partial_z v}$, will be marginalized to BT_V^* in Eq.(4.33), and this is indeed true. For convenience, the expression for $T_{K^M,\partial_z v}$ is rewritten as follows:

$$T_{K^M,\partial_z v} = \hat{\underline{\mathbf{v}}}_n^{\sim1} \cdot \underline{\mathbf{k}} \wedge \widehat{(\underline{\zeta}^* w)}_n^{\sim1}$$

where

$$\underline{\zeta}^* = \zeta_1 \underline{\mathbf{i}} + \zeta_2 \underline{\mathbf{j}}$$

and

$$\zeta_1 = -\frac{\partial v}{\partial z}, \qquad \zeta_2 = \frac{\partial u}{\partial z}.$$

By the marginalization property, we have

$$\begin{aligned}
\mathcal{M}_n T^{0\to1}_{K^M,\partial_z v} &= \mathcal{M}_n T_{K^M,\partial_z v} \\
&= \overline{\underline{\mathbf{v}}' \cdot \underline{\mathbf{k}} \wedge (\underline{\zeta}^* w)'} \\
&= \overline{\underline{\mathbf{v}}' \cdot \left[(\zeta_1 w)'\underline{\mathbf{j}} - (\zeta_2 w)'\underline{\mathbf{i}}\right]} \\
&= \overline{-(\bar{u}+u')(\zeta_2 w)' + v'(\zeta_1 w)'} \\
&= -\overline{u'w'}\frac{\partial \bar{u}}{\partial z}
\end{aligned}$$

after linearization. So,

$$\mathcal{M}_n \left\langle T^{0\to1}_{K^M,\partial_z v} \right\rangle^\Omega = -\left\langle \overline{u'w'} \frac{\partial \bar{u}}{\partial z} \right\rangle^\Omega . \tag{4.45}$$

Again, this is exactly BT_V^*, one of the terms in Eq.(4.33).

4.3.3 Potential energetics

The treatment of the available potential energy equation runs a little differently from that of its kinetic peer. Before making any comparison between (3.72) and its classical counterpart, one thing about the density field ρ has to be clarified. Recall in arriving at (3.18), the ρ actually has been removed by a stationary vertical profile, and this removed structure is embedded in the parameter N^2, c, or whatever derives. For this reason, if that density is averaged again, the so-obtained mean should be z-independent. That is to say, $\frac{\partial \bar{\rho}}{\partial z} = 0$. This is somewhat confusing since we insist that there exists a sizable buoyancy frequency N^2. In deriving the classical energy equations, ρ is not removed by that vertical profile, $\bar{\rho}$ is a function of both y and z. The two different starting points from which we obtain the available potential energy equations could cause a little confusion, but fortunately everything, at least in notation, should be the same (even $\partial\bar{\rho}/\partial y$ is *quantitatively* the same, albeit $\bar{\rho}$ itself has a different meaning), except that in the following derivations the term $\frac{\partial \bar{\rho}}{\partial z}$, if any, should be set nil.

As the kinetic case, the meso-scale available potential energy A_n^M marginalizes to A^M, and hence $\mathcal{M}_n \left\langle A_n^M \right\rangle^\Omega = \left\langle A^M \right\rangle^\Omega$. For the T-terms on the right hand side of Eq.(3.72), we expect they together contribute to the $\langle BC^* \rangle^\Omega$ of Eq.(4.33). Since Eq.(3.72) involves a complicated separation of the field ρ, we will not follow how it is formulated. Instead, we just consider the sum of all the right hand side terms in that equation except the buoyancy conversion. Because the transports integrate to zero over the whole domain Ω, the final result should just be the same. In other words, what we need to consider is the summation of

$$-c\widehat{\rho}_n^{\sim1} \nabla \cdot \widehat{(\rho \underline{\mathbf{v}})}_n^{\sim1} - c\widehat{\rho}_n^{\sim1} \frac{\partial}{\partial z} \widehat{(w\rho)}_n^{\sim1}$$

over all possible n's. By the marginalization equality, this sum is

$$T = -\overline{c\rho' \nabla \cdot (\rho \underline{\mathbf{v}})'} - \overline{c\rho' \frac{\partial}{\partial z} (w\rho)'}.$$

Integrate T over Ω. The first part is, with terms of higher order dropped,

$$
\begin{aligned}
&\iiint_\Omega -\overline{c\rho'\nabla\cdot(\rho\underline{\mathbf{v}})'}\ dV = \iiint -\overline{c\rho'\nabla\cdot(\bar\rho\underline{\mathbf{v}}' + \rho'\bar{\underline{\mathbf{v}}} + \rho'\underline{\mathbf{v}}')}\ dV \\
&= \iiint -\overline{c\rho'\nabla\cdot(\bar\rho\underline{\mathbf{v}}' + \rho'\bar{\underline{\mathbf{v}}})}\ dV \\
&= \iiint -\overline{c\left[\bar\rho\rho'\left(-\frac{\partial w'}{\partial z}\right) + \left(\rho' u'\frac{\partial\bar\rho}{\partial x} + \rho' v'\frac{\bar\rho}{\partial y}\right)\right]}\ dV \\
&= \iiint_\Omega c\bar\rho\overline{\rho'\frac{\partial w'}{\partial z}}\ dV - \iiint_\Omega c\frac{\partial\bar\rho}{\partial y}\overline{\rho' v'}\ dV.
\end{aligned}
$$

The second part involves the z-derivative of the mean of ρ. As explained before, it should be set zero. This part is thus integrated to

$$
\begin{aligned}
\iiint_\Omega -c\overline{\rho'\frac{\partial}{\partial z}(w\rho)'}\ dV &= -\iiint c\overline{\rho'\frac{\partial}{\partial z}(w'\bar\rho)}\ dV \\
&= -\iiint_\Omega c\bar\rho\overline{\rho'\frac{\partial w'}{\partial z}}\ dV.
\end{aligned}
$$

Therefore, the integral of T over Ω, which is the sum of the above two parts, is equal to

$$
-\iiint_\Omega c\frac{\partial\bar\rho}{\partial y}\overline{\rho' v'}\ dV. \tag{4.46}
$$

Recall that

$$
c = \frac{g^2}{\rho_0^2 N^2}.
$$

Substituting it for the c in (4.46) followed by an application of the thermal wind relation, Eq. (4.46) turns out to be none other than the $\langle BC^*\rangle^\Omega$ in Eq. (4.33):

$$
\begin{aligned}
&\iiint_\Omega \mathcal{M}_n(T_{A^M,\partial_h\rho} + T_{A^M,\partial_z\rho} + TS_{A^M})\ dV \\
&= -\iiint_\Omega \frac{fg}{\rho_0 N^2}\overline{v'\rho'}\frac{\partial\bar u}{\partial z}\ dV \\
&= \langle BC^*\rangle^\Omega.
\end{aligned}
$$

This is to say,

$$
\mathcal{M}_n\left\langle T^{0\to1}_{A^M,\partial_h\rho} + T^{0\to1}_{A^M,\partial_z\rho} + TS^{0\to1}_{A^M}\right\rangle^\Omega = \langle BC^*\rangle^\Omega, \tag{4.47}
$$

by Eq. (4.39).

4.3.4 Instability identification with the MS-EVA

According to the foregoing proofs, marginalization of MS-EVA cross-scale transfers yields exactly the corresponding energetic terms for linear stabilities, so long as the large-scale window is set with the gravest index (j_0=0) and a periodic extension is adopted. Denoting

$$\begin{aligned} BT_H &= T^{0\to1}_{K^M,\zeta} + T^{0\to1}_{K^M,w\partial_z}, \\ BT_V &= T^{0\to1}_{K^M,\partial_z v}, \\ BC &= T^{0\to1}_{A^M,\partial_h\rho} + T^{0\to1}_{A^M,\partial_z\rho} + TS^{0\to1}_{A^M}, \end{aligned}$$

(n-dependence suppressed for simplicity) with $BT = BT_H + BT_V$ and BC corresponding respectively to the barotropic and baroclinic instabilities as with their starred counterparts, Eq. (4.33) then can be written

$$\mathcal{M}_n \hat{\delta}_n \left\langle A^M_n + K^M_n \right\rangle^\Omega = \mathcal{M}_n \langle BC \rangle^\Omega + \mathcal{M}_n \langle BT_H + BT_V \rangle^\Omega. \tag{4.48}$$

Here the time derivative has been modified as we did in deriving the multiscale energetic equations in Chapter 3.

Equation (4.48) is a marginalized statement of energetics for linear instability. It reflects the accumulated effect of all the dynamic phenomena on the meso-scale window. A localized version of it is

$$\hat{\delta}_n \left\langle A^M_n + K^M_n \right\rangle^\Omega = \langle BC \rangle^\Omega + \langle BT_H + BT_V \rangle^\Omega + \langle R \rangle^\Omega, \tag{4.49}$$

where the extra term $\langle R \rangle^\Omega$ (again location dependence suppressed) is such that $\mathcal{M}_n R = 0$. R is apparently the total transfer from the same scale window (meso-scale), and because of its existence, $\hat{\delta}_n \left\langle A^M_n + K^M_n \right\rangle^\Omega$ is no longer the instability indicator as $\mathcal{M}_n \hat{\delta}_n \left\langle A^M_n + K^M_n \right\rangle^\Omega$ is in (4.33) (unless R is negligible). Given a location, one has to calculate the sum of the cross-scale transfers $\langle BC \rangle^\Omega + \langle BT_H + BT_V \rangle^\Omega$ to see whether a system is stable or not. This, together with the notion of barotropic and baroclinic instabilities, gives the following criteria for process identification:

(1) A system is unstable on the meso-scale window if $\langle BT_H + BT_V + BC \rangle^\Omega > 0$ and vice versa (in a quasi-geostrophic model, BT_V vanishes).

(2) For an unstable system, if $\langle BT_H + BT_V \rangle^\Omega > 0$ and $\langle BC \rangle^\Omega \leq 0$, the instability the system undergoes is barotropic.

(3) For an unstable system, if $\langle BC \rangle^\Omega$ is positive but $\langle BT_H + BT_V \rangle^\Omega$ is not, then the instability is baroclinic.

(4) If both $\langle BT_H + BT_V \rangle^\Omega$ and $\langle BC \rangle^\Omega$ are positive, the system must be undergoing a mixed instability.

It should be pointed out that, if the system is not a closed one, or if it is a spatially local feature that needs studying, the average cannot be made over the whole spatial domain. Ω must be chosen to encompass only the feature of concern with dynamical integrity. In a model with an idealized configuration, this is not a difficult business; in real problems, however, an unambiguous choice for Ω is usually not feasible. We will get back to this issue after the following two sections, the validation of MS-EVA with two idealized models, the Eady model and the Kuo model.

4.4 MS-EVA validation with a baroclinic instability model

The Eady model is among one of the most extensively studied examples of linear wave stability. Introduced by Eady (1949) in an elegantly simple form, it has since become a benchmark for baroclinic instability studies. In this section, we will use this model to validate the MS-EVA, and at the same moment explore using MS-EVA processes how a baroclinic instability is formed.

4.4.1 An overview of the Eady model

Originally, Eady built his model from a frictionless primitive equation set [cf., (4.1)-(4.5)], on an f-plane with Bousinesq and hydrostatic approximations. The model domain extends infinitely in either horizontal direction, and is bounded above and below by two rigid-lids, i.e.

$$w = 0 \qquad \text{at } z = 0, z = -H \ (H = \text{const}). \tag{4.50}$$

The basic flow has only a z-dependence:

$$\bar{u} = \bar{u}(z), \quad \frac{\partial \bar{u}}{\partial z} = \Lambda = \text{const}, \tag{4.51}$$

permitting no gradient in the mean potential vorticity. Nevertheless, this background configuration still meets the requirements of the Rayleigh theorem for baroclinic instability (see, for example, Holton, 1992; Robinson, 1996). As $\nabla \bar{u} = 0$, barotropic instabilities are essentially excluded. Simplify the model further by assuming a basic density profile:

$$\bar{\rho} = B_y y + B_z z, \tag{4.52}$$

where B_y and B_z are two constants. The linear dependence on y is a result of (4.51) by the thermal wind relation (4.15). A straightforward calculation gives $B_y = \rho_0 f_0 \Lambda / g$. The linear dependence on z is not essential, though it does help simplify the derivation that follows. Introduce a perturbation small enough so that (4.16) is satisfied. A linearized equation set is obtained thereby, in the dimensional forms of (4.18) - (4.22):

$$\frac{\partial u'}{\partial t} = -\bar{u}\frac{\partial u'}{\partial x} - v'\frac{\partial \bar{u}}{\partial y} - w'\frac{\partial \bar{u}}{\partial z} + fv' - \frac{1}{\rho_0}\frac{\partial P'}{\partial x}, \tag{4.53}$$

$$\frac{\partial v'}{\partial t} = -\bar{u}\frac{\partial v'}{\partial x} - fu' - \frac{1}{\rho_0}\frac{\partial P'}{\partial y}, \tag{4.54}$$

$$0 = \frac{\partial u'}{\partial x} + \frac{\partial v'}{\partial y} + \frac{\partial w'}{\partial z}, \tag{4.55}$$

$$\rho' g = -\frac{\partial P'}{\partial z}, \tag{4.56}$$

$$\frac{\partial \rho'}{\partial t} = -\bar{u}\partial \rho' - v'\frac{\partial \bar{\rho}}{\partial y} - w'\frac{\partial \bar{\rho}}{\partial z}. \tag{4.57}$$

If a solution of the form

$$\begin{bmatrix} u' \\ v' \\ w' \\ \rho' \\ P' \end{bmatrix} = \text{Re}(\begin{bmatrix} \tilde{u}(z) \\ \tilde{v}(z) \\ \tilde{w}(z) \\ \tilde{\rho}(z) \\ \tilde{P}(z) \end{bmatrix} \cdot e^{i(kx+ly-\omega t)}), \tag{4.58}$$

is to be sought (k and l non-negative, and $\omega = \omega_0 + i\omega_1$ complex), Eady shows that the problem is reduced to solving for $\tilde{w}$ an ordinary differential equation

$$\xi(\xi^2 - 1)\frac{d^2\tilde{w}}{d\xi^2} + 2(1 - i\sigma\xi)\frac{d\tilde{w}}{d\xi} + [R_\sigma \xi + 2i\sigma]\tilde{w} = 0 \tag{4.59}$$

subject to boundary conditions

$$\tilde{w} = 0, \qquad \text{at } \xi = \xi_{top},\ \xi_{bot} \tag{4.60}$$

where

$$\xi = \xi(z) = \frac{\bar{u}k - \omega}{f}, \qquad \xi_{top} = \xi(0), \quad \xi_{bot} = \xi(-H), \tag{4.61}$$

$$\sigma = \frac{l}{k}, \qquad R_\sigma = (1+\sigma^2)R_i, \tag{4.62}$$

and the Richardson number

$$R_i = -\frac{g/\rho_0 \cdot \partial\bar{\rho}/\partial z}{(\partial u/\partial z)^2} = -\frac{gB_z}{\rho_0 \Lambda^2}. \tag{4.63}$$

He also showed, when $R_i \gg 1$, and $|\xi^2| \ll 1$, this equation can be well approximated by

$$\frac{d^2W}{d\xi^2} - \frac{2}{\xi}\frac{dW}{d\xi} - R_\sigma W = 0, \tag{4.64}$$

with the new variable

$$W(z) \quad = \quad \left(\frac{1-\xi}{1+\xi}\right)^{-\frac{i\sigma}{2}} \tilde{w}(z), \tag{4.65}$$

Eq. (4.64) and boundary conditions

$$W = 0, \qquad \text{at } \xi = \xi_{top}, \text{ and } \xi = \xi_{bot}, \tag{4.66}$$

form the new eigenvalue problem. Let $\bar{u}_{top} = \bar{u}(0)$, $\bar{u}_{bot} = \bar{u}(-H)$. It has nontrivial solutions if and only if

$$\begin{cases} \omega_0 = \frac{k}{2}(\bar{u}_{top} + \bar{u}_{bot}) \\ \omega_1 = \gamma \frac{f}{\sqrt{R_\sigma}}. \end{cases} \tag{4.67}$$

Here γ is defined such that

$$\gamma^2 \quad = \quad (\alpha - \tanh\alpha)(\coth\alpha - \alpha) \tag{4.68}$$

and

$$\alpha \quad = \quad \frac{\sqrt{R_\sigma}}{2f}(-\Lambda)kH = \frac{\sqrt{R_\sigma}}{2f}(\bar{u}_{bot} - \bar{u}_{top})k. \tag{4.69}$$

is the scaled wavenumber. Plotted in Fig. 4.1 is the variation of γ^2 as a function of α. Unstable solutions exist if and only if γ is real, i.e., $\gamma^2 > 0$. Unfolded on the axis of α, this is possible only if

$$|\alpha| < \alpha_{crt} \approx 1.1997, \qquad \text{or} \tag{4.70}$$

$$|k| < \frac{2f}{\sqrt{R_\sigma}}\Delta\bar{u} \cdot \alpha_{crt} \approx 2.389\frac{f}{\sqrt{R_\sigma}}\Delta\bar{u}, \tag{4.71}$$

where $\Delta\bar{u} = \bar{u}_{top} - \bar{u}_{bot}$.

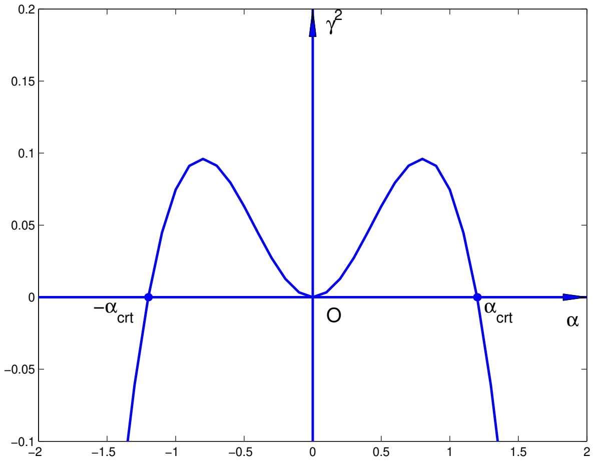

Figure 4.1: γ^2 as a function of α.

Eq. (4.67) gives the eigenvalue ω for problem (4.64) plus (4.66). Corresponding to it is the eigenfunction of $\tilde{w}(z)$:

$$\tilde{w}(z) = \left(\frac{1-\xi}{1+\xi}\right)^{i\frac{\sigma}{2}} \cdot W(z), \tag{4.72}$$

where

$$W(z) = C_1 e^{\sqrt{R_\sigma}\xi}(1 - \sqrt{R_\sigma}\xi) + C_2 2^{-\sqrt{R_\sigma}\xi}(1 + \sqrt{R_\sigma}\xi)$$

and the constants C_1 and C_2 are related through

$$\frac{C_2}{C_1} = -e^{-2\sqrt{R_\sigma}\xi_{top}}\left(\frac{1 - \sqrt{R_\sigma}\xi_{top}}{1 + \sqrt{R_\sigma}\xi_{top}}\right).$$

Once $\tilde{w}$ is found, other eigenfunctions are easily determined. Let $a = -\Lambda/f$. They are,

$$\tilde{u}(z) = \frac{a}{1+\sigma^2}\left[-\frac{\sigma}{\xi}\frac{d\tilde{w}}{d\xi} - i(\sigma^2\frac{\tilde{w}}{\xi} + \frac{d\tilde{w}}{d\xi})\right], \tag{4.73}$$

$$
\begin{aligned}
\tilde{v}(z) &= \frac{a}{1+\sigma^2}\left[\frac{1}{\xi}\frac{d\tilde{w}}{d\xi} + i\sigma(\frac{\tilde{w}}{\xi} - \frac{d\tilde{w}}{d\xi})\right], && (4.74)\\
\tilde{P}(z) &= \frac{\rho_0}{il}(i\omega\tilde{v} - ik\bar{u}\tilde{v} - f\tilde{u}) \\
&= \frac{a}{1+\sigma^2}\cdot\frac{f\rho_0}{ik}\cdot\left[(\frac{1}{\xi}\frac{d\tilde{w}}{d\xi} + i\sigma\frac{\tilde{w}}{\xi}) + (\tilde{w} - \xi\frac{d\tilde{w}}{d\xi})\right], && (4.75)\\
\tilde{\rho}(z) &= -\frac{1}{g}\frac{d\tilde{p}}{dz} = -\frac{a}{1+\sigma^2}\cdot\frac{fa\rho_0}{g}\cdot\frac{i}{\xi}\cdot\left[(\frac{1}{\xi}\frac{d\tilde{w}}{d\xi} + R_\sigma\tilde{w}) + i\sigma(\frac{\tilde{w}}{\xi} - \frac{d\tilde{w}}{d\xi})\right], && (4.76)
\end{aligned}
$$

which, when substituted into (4.58), yield the perturbation fields $(u', v', w', \rho', P')^T$. Notice again, here all variables are dimensional.

4.4.2 Dataset generating

To form a MS-EVA-ready dataset, choose a configuration as follows:

ρ_0	=	$1025\ \text{kg/m}^3$	H	=	$1000\ \text{m}$
$\bar{u}_{top}$	=	$20\ \text{cm/s}$	$\bar{u}_{bot}$	=	$10\ \text{cm/s}$
B_z	=	$-1\times10^{-2}\ \text{kg/m}^4$	f	=	$1\times10^{-4}\ \text{Hz}$

which is to some extent typical of the mid-ocean setting. Let the domain extend 1000 km in both x and y directions (not essential), and make the following calculations:

$$
\begin{aligned}
\Lambda &= (\bar{u}_{top} - \bar{u}_{bot})/H = 1.000\times10^{-4}\text{s}^{-1}\\
B_y &= \rho_0 f\Lambda/g = 1.046\times10^{-6}\text{kg/m}^4\\
R_i &= -\frac{gB_z}{\rho_0\Lambda^2} = 9.561\times10^3\\
a &= -\Lambda/f = -1.000
\end{aligned}
$$

Obviously, the Richardson number $R_i \gg 1$. This is one of the two conditions under which the Eady solution is valid. To see whether another condition $\xi^2 \ll 1$ holds, notice that the instability requirement $|\alpha| < 1.1997$ implies

$$k < 2.389\frac{f}{\sqrt{R_\sigma}}\Delta\bar{u} \sim O(10^{-5})\ \ \text{m}^{-1}.$$

The wavenumbers of interest are thus of order $O(10^{-5})$ m^{-1} or smaller. By (4.67) a wavenumber of this order will give an $|\omega|$ smaller than order $O(10^{-6})$ s^{-1}. Therefore

$$\xi = \frac{\bar{u}k - \omega}{f} < O(10^{-2}),$$

which, together with the fact $R_i \gg 1$, validates Eady's approximations and thus the above solution indeed holds so long as waves of concern are long enough such that $k \leq O(10^{-5})$ m^{-1}.

Next pick an eigen mode to examine the energetics. Without affecting the representativeness, choose a zero y wavenumber,[6] i.e., let $l = 0$. In this case, $\sigma = 0$, $R_\sigma = R_i$. The solution (4.72) - (4.76) is hence very much simplified, and the shortwave cut-off (corresponding to α_{crt}) is $k_{crt} = 2.454 \times 10^{-5}$ m^{-1}. Given a k smaller than this k_{crt}, an unstable solution is obtained, and correspondingly a dataset can be generated. What we need consider next is the consistency of this set.

For a dataset to be consistent in the sense of MS-EVA application, its time sequences must be

- Long enough to span some integer cycle(s), otherwise aliasing effects will come in and give a spurious basic flow for the decomposition;
- Short enough that the amplitude of perturbation is within the tolerance of a valid linear solution.

These seemingly contradictory requirements can be easily met if we choose an unstable system with perturbation large in frequency while small in growth rate, i.e., $|\omega_r| \gg |\omega_i|$. By (4.67), if we have k distinctly larger than zero, and at the same time let $\gamma \to 0$, then this condition is satisfied. Such k's do exist. They are near the critical wavenumber k_{crt}, as shown in Fig. 4.1 (notice $k \propto \alpha$ when $l = 0$). In the present configuration, R_i is very large, so this issue is not very serious. We simply choose $k = 1 \times 10^{-5}$ m^{-1}, which gives an $\omega = 1.50 \times 10^{-6} + 2.53 \times 10^{-7} i$ rad/s, meeting the above requirements.

With the eigenvalue and eigenvector, we compute the solutions of all the fields and truncate them at exactly the end of the second cycle to form a series for each data point. The series length is thence

$$2 \times \frac{2\pi}{\omega_r} \approx 8.38 \times 10^6 \text{ s} = 97 \text{ days}.$$

Within this interval, disturbances grow at most by $\exp(8.38 \times 10^6 \omega_i) \approx 8.31$ times. So, if initially the perturbation is set smaller than the basic field by $\frac{1}{8.31}\varepsilon$ (ε the permissible relative

[6] A nonzero l results in a similar solution to the case $l = 0$ except that on the horizontal levels, the crest/valley-alignment now has an angle relative to the meridional lines.

magnitude for a linear disturbance), then the solution will be valid throughout the 97-day time duration. This can be done by properly manipulating the constant C_1 of $W(z)$ (p. 195). Choose $C_1 = 2 \times 10^{-7}$. It yields a maximal $|u'|/\bar{u} < 10^{-3}$, which is much smaller than the Rossby number $Ro = \frac{U}{fL} \sim 10^{-2}$.

Apart from the unstable solution, we also need to examine the energetics for a stable mode and a neutrally stable mode. Here growth rate is not an issue any more. The solution will always be valid as long as it is so in the beginning. We choose for the two modes $k = 3 \times 10^{-5}\mathrm{m}^{-1} > k_{crt}$ and $k = k_{crt} = 2.454 \times 10^{-5}$ m^{-1}, which yield $\omega_r = 4.04 \times 10^{-6}$ rad/s and $\omega_r = 3.68 \times 10^{-6}$ rad/s, respectively. Again, the solution series are truncated at the end point of the second cycle (36 days and 39 days in length, respectively). The three chosen solutions are sampled at $2^{10} = 1024$ time instants, and mapped onto an Arakawa B-grid (see Fig. A3.7) with a mesh of 20 km $\times$ 50 km $\times$ 100 m (50 points in x, 20 points in y, and 10 levels in z). The datasets obtained are now ready for the MS-EVA application.

4.4.3 MS-EVA validation with the Eady model

With the datasets thus generated, a two-scale decomposition is performed on each field. Let $j_0 = 0$ ($j_2 = 10$ known) and adopt a periodic extension. The perturbation is expected to be reconstructed precisely on the meso-scale window, i.e., $p^{\sim 1}$ should be equal to p' in its discretized form, for any p of concern. We therefore need only consider the meso-scale energetics, and in fact, we need only consider the meso-scale energetics for a particular instant. This is because, as a result of linearization, solutions of the Eady model are similar at all time points. Any snapshot of a field is typical of the evolution pattern of that field throughout the duration.

Contoured in Figs. 4.2-4.4 are the day-10 meso-scale syntheses of the solution vectors on section j=10 (y=500 km) for the three chosen modes. Their distributions in y are not displayed as there is no variation in that direction, a result of the choice of a zero l. Different modes bear different section structures. For the unstable mode, the reconstructed pattern is shown in Figure 4.2. It was documented in Eady's original paper, and is also consistent with the outcome of a quasi-geostrophic model (e.g., Holton, 1992. p. 261). A conspicuous feature is the tilting of the phase lines on all the section plots. On the same section with the other two datasets, this phenomenon is not seen. In the neutral case where $k = k_{crt}$ (Fig. 4.3), crests and valleys are aligned parallel to the z coordinate and distributed symmetrically about the mid-depth. When k exceeds the short wave cut-off (Fig. 4.4), the phase line alignment is still as that of Fig. 4.3, though the mid-depth symmetry disappears and the disturbances are generally bottom-trapped (except $w^{\sim 1}$). In the Eady model, therefore, the tilting phenomenon is peculiar to the unstable mode.

The above is a general picture of the section distribution of the unstable mode versus its stable counterparts. If observed closely, the tilting pattern actually varies from field to field. Phase lines of different fields generally have different slopes. Of particular interest is that of $\rho^{\sim 1}$, which tilts in a way distinctly different from its peers. Recall that the basic flow is eastward. The counter-tilting of $\rho^{\sim 1}$ thence means a phase advance of the upper-level perturbation relative to the lower-level perturbation (or a phase lag for other fields). This fact, which is believed to characterize baroclinic instabilities, has been identified in many baroclinically unstable events in both atmosphere and oceans. (e.g., Cummins and Mysak, 1988; Holton, 1992).

The phase line tilting is not a characteristic belonging to the meso-scale syntheses only. It is also seen on the energetic distributions for the same scale window. Plotted in Figs. 4.5 and 4.6 are the MS-EVA terms computed respectively for the corresponding potential and kinetic equations, each with axes of trough/ridge sloping with height. In Fig. 4.5, the phase line titling of $\hat{\delta}_n A_n^M$ (a), $T_{A^M,\partial_x \rho}$ (e), and $BC = T^{0\to 1}_{A^M,\partial_h \rho} + T^{0\to 1}_{A^M,\partial_z \rho}$ (h) is toward the east with height. Their peers, b_n^M (d) and $T_{A^M,\partial_y \rho}$ (f), and all the kinetic terms shown in Fig. 4.6, have an opposite tilting direction, however. Particularly, the two interaction analytical terms, $BT_H = T^{0\to 1}_{K^M,\zeta} + T^{0\to 1}_{K^M,w\partial_z}$ and BC, tilt in a way at odds with each other.

While the energetics vary with a tilting pattern similar to the perturbation, its scale of variation is quite different. Roughly speaking, the zonal scale has been reduced by a half compared to that of Fig. 4.2 (as energetic terms are all quadratic). Recall that energetics for a scale window are related to the energy variation, rather the energy itself, on that window. The difference in scale is thus not surprising. For a similar example we have already mentioned the scaling discrepancy between the streamfunction and vorticity in a barotropic turbulence. One should not, therefore, attempt to tell from maps of energetics the scale for the corresponding

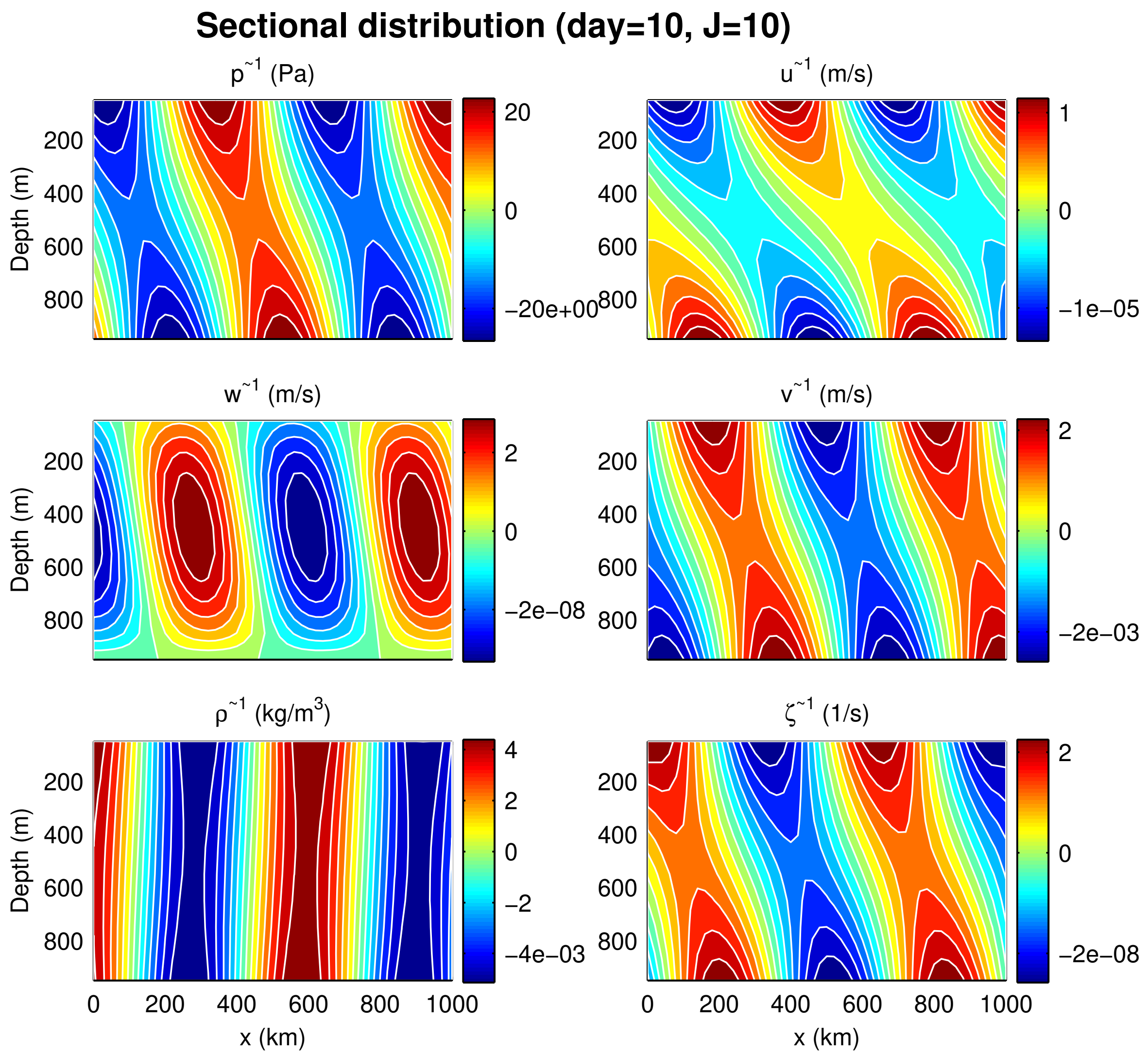

Figure 4.2: Meso-scale reconstruction of the unstable Eady mode for day 10 on section $y = 500$ km. $l = 0$, $k = 1 \times 10^{-5}$ m^{-1} (corresponding to $|\alpha| = 0.489 < \alpha_{crt}$), and $C_1 = 2 \times 10^{-7}$. Note $\zeta^{\sim 1}$ is not an element of the solution vector.

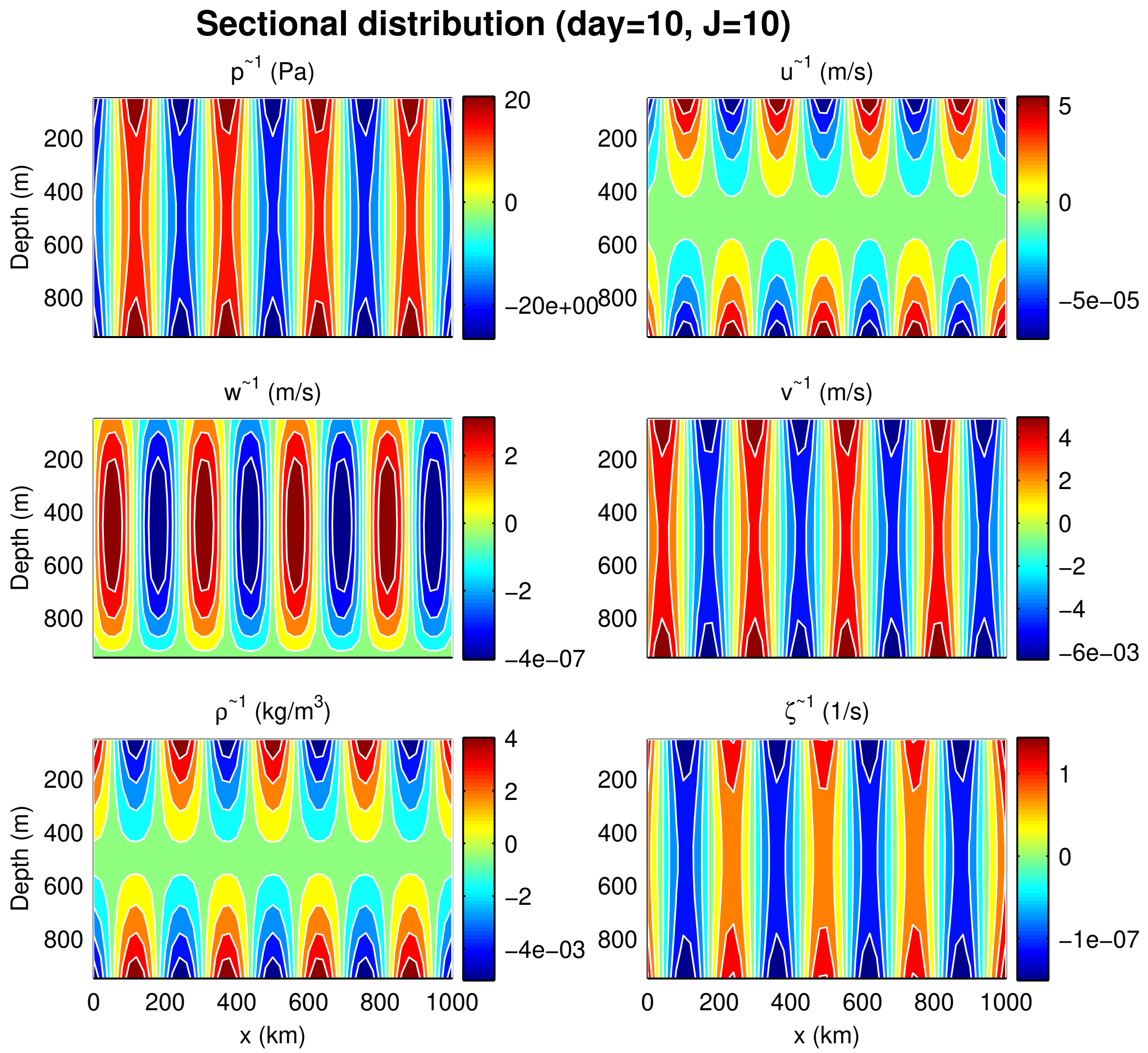

Figure 4.3: Meso-scale reconstruction of the neutrally stable Eady solution. Parameters are set as Fig. 4.2, except $k = 2.454 \times 10^{-5}$ m^{-1} here (corresponding to $|\alpha| = \alpha_{crt}$).

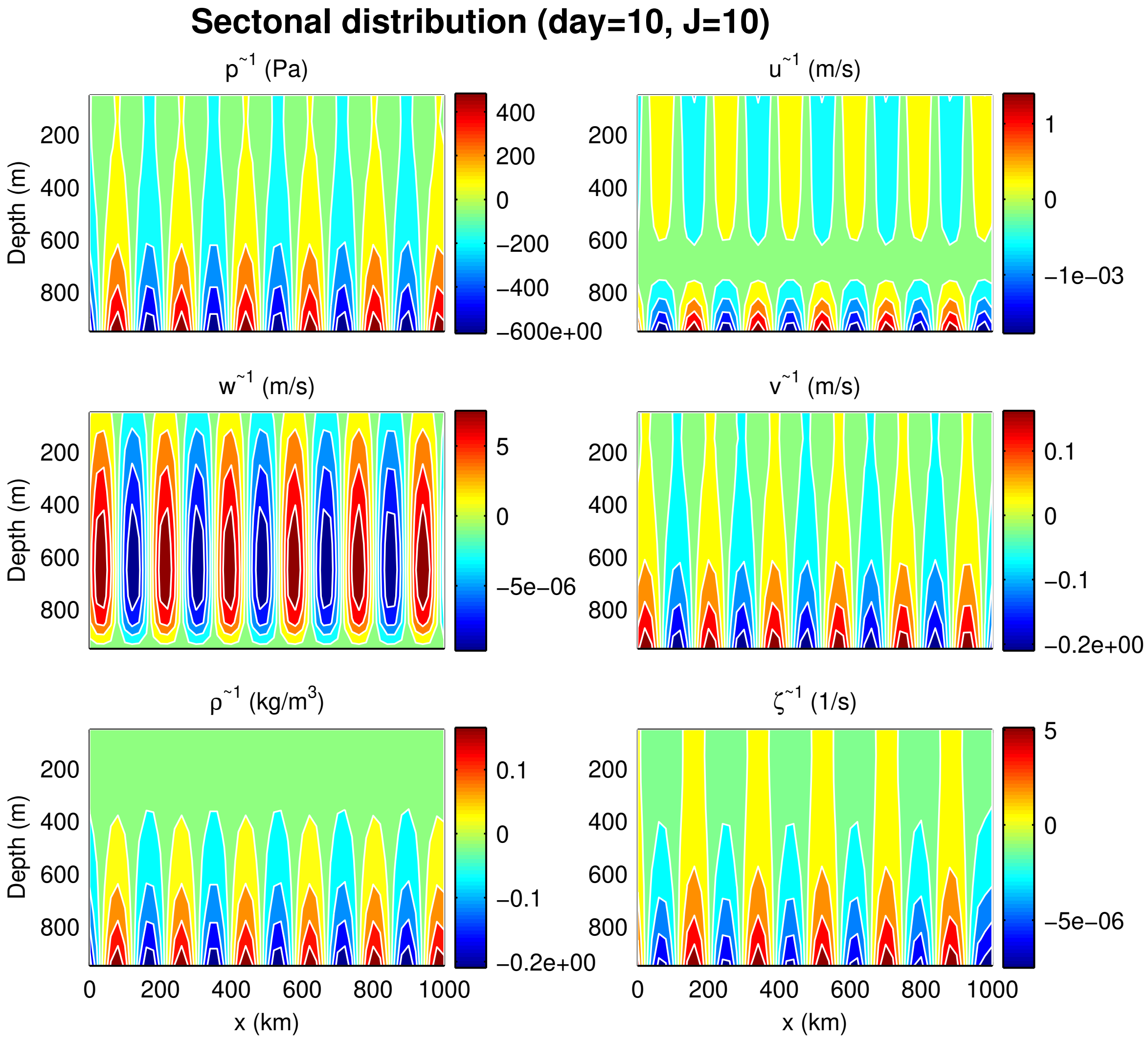

Figure 4.4: Meso-scale reconstruction of the stable Eady solution. Parameters are set as Fig. 4.2, except $k = 3 \times 10^{-5}$ m^{-1} here (corresponding to $|\alpha| = 1.467 > \alpha_{crt}$).

perturbation.

The computation of MS-EVA terms allows an easy identification of the energetic balances. From Fig. 4.5, the dominant sources for the meso-scale APE evolution are from the two transfer processes, $T_{A^M,\partial_x\rho}$ and $T_{A^M,\partial_y\rho}$. The buoyancy conversion b_n^M is one order smaller, and others are negligible (at least two orders smaller). Notice that in Pinardi and Robinson (1986)'s QG-EVA and Spall (1989)'s Pre-EVA, no transfer process is present, and buoyancy conversion is hence dominantly important. When transfer processes come in, the balance changes. The T terms are obviously of much more substance. While they marginalize and sum cross-scalewise to zero (that's the reason they do not exist in the classical EVA formulation), they account for the largest part of the change of A_n^M. The balance of the meso-scale APE equation is thus mainly between $\hat{\delta}_n A_n^M$, $T_{A^M,\partial_x\rho}$, $T_{A^M,\partial_y\rho}$, and only when processes of one order smaller are included, does b_n^M weigh on the balance.

But the buoyancy conversion is definitely of importance. It is the only connection between the two types of energy, APE and KE (cf. Fig. 3.5). From the Fig. 4.5d, it is negatively defined at all vertical levels, with a maximum achieved in the middle. The negative b_n^M indicates a net conversion of meso-scale APE to meso-scale KE throughout the water column, which has been identified as a signature of baroclinic instability by previous EVA studies (e.g., Pinardi and Robinson, 1986). If compared to Fig. 4.2, this negative definiteness of b_n^M actually has been implied by the relative distributions of $w^{\sim 1}$, $\rho^{\sim 1}$, and $v^{\sim 1}$. Look at the right corner of the maps of these fields in Fig. 4.2: $v^{\sim 1}$ and $w^{\sim 1}$ are positive (ridge), while $\rho^{\sim 1}$ is negative (trough). This relative positioning of the ridge and trough yields an advection of negative perturbation density toward the top and the north. The direction of this advection has a slope of the order $2 \times 10^{-8}/2 \times 10^{-3} \sim \mathrm{O}(10^{-5})$. Remember that the background isopycnals incline northward with height at a slope $|B_y/B_z| \sim O(10^{-4})$, which is greater than $O(10^{-5})$. The milder slope of the advection lines relative to the isopycnals implies that there will be inevitably an injection of buoyancy into the heavy isopycnal layers, leading to a negative conversion and hence a release of available potential energy.

So far we have described the meso-scale potential energetics. For the kinetic energetics on the same window (Fig. 4.6), things are quite different. Transfer processes no longer dominate. Although $T_{K^M,\zeta}$ is still the biggest over most of the domain, the buoyancy conversion and pressure working rates are also of significance. The balance is thence between $\hat{\delta}_n K_n^M$, $T_{K^M,\zeta}$, $\Delta_h Q_{PM}$, $\Delta_z Q_{PM}$, and $-b_n^M$. Note the negative b_n^M represents the buoyancy working rate converted from the APE. Its section distribution pattern bears a strong resemblance to $\Delta_z Q_{PM}$, the vertical pressure working rate, except for an opposite sign. This implies, by and large, that the incoming energy from the APE is redistributed away by $\Delta_z Q_{PM}$ upon arriving. And, by the loss-at-mid-depth and gain-at-boundaries pattern, the redistributing is mainly toward the surface and the bottom. This is the very scenario which has already been depicted by Pinardi and Robinson (1986) in their baroclinic instability study with the QG EVA.

Our MS-EVA therefore yields a result in agreement with the classical scenario. We proceed to show that the process is indeed a baroclinic instability by the MS-EVA diagnosis. This is the key step in the validation.

First look at $\hat{\delta}_n A_n^M$ and $\hat{\delta}_n K_n^M$. Their section distributions are plotted in Figs. 4.5a and 4.6a. Both of them oscillate in x, contour lines of $\hat{\delta}_n A_n^M$ tilting with height toward the east, while those of $\hat{\delta}_n K_n^M$ toward the west. Their horizontal maps are also plotted, at the left-top corners of Figs. 4.7 and 4.8. Compared to Figs. 4.5a and 4.6a, these maps are rather plain. No variation in y is in existence. Simple as they may look, important information does exist in the relative amplitudes of the ridges and troughs. On either of them there is an apparent gain of energy, and this is true for each level throughout the water column (maps other than level 5 not shown). This can be seen clearly from plots of $\hat{\delta}_n A_n^M$ and $\hat{\delta}_n K_n^M$ as functions of x only. Shown in Fig. 4.9 are such plots at some y for the surface, mid-depth, and bottom levels. The signals are sinusoidal, and if averaged over a wavelength, there will be a positive residue for each level (see Fig. 4.11c). Recall the classical definition of instability: A system is unstable if the marginalization of the total meso-scale energy increases when averaged over a closed spatial domain. Here since transfer from the same scale window is not significant (not shown), these positive residues then imply that an instability is indeed going on there. The next question is: Is this instability really a baroclinic one?

As proved in the preceding section, the two types of instabilities can be distinguished by performing a proper interaction analysis on the transfer terms. For a baroclinic instability, it is these transfer processes represented by $T_{A^M,\partial_h\rho}$ and $T_{A^M,\partial_z\rho}$ that matter, and $BC = T^{0\to1}_{A^M,\partial_h\rho} + T^{0\to1}_{A^M,\partial_z\rho}$ integrated over the definition domain must be positive; for a barotropic instability, what we should pay attention to is $BT_H = T^{0\to1}_{K^M,\zeta} + T^{0\to1}_{K^M,w\partial_z}$. These two terms are contoured, respectively, in Figs. 4.5h and 4.6h, and Figs. 4.7i and 4.8i.

From Figs. 4.5 and 4.6, the variations of BC and BT_H generally follow the same trend as those of $\hat{\delta}_n A^M_n$ and $\hat{\delta}_n K^M_n$, respectively. Their axes of troughs and ridges tilt against each other. To make a quantitative evaluation possible, we plot, as we did for $\hat{\delta}_n A^M_n$ and $\hat{\delta}_n K^M_n$, these transfer rates as functions of x for each vertical level. Shown in Fig. 4.10 are such plots for levels 1, 5 and 10 (corresponding to 50 m, 450 m, and 950 m). Plots for other levels are similar. It is obvious that there is a net gain in the transfer BC at each level, while for BT_H, no gain, nor loss. This fact is true for all the horizontal levels throughout the water column (cf. the averaged BC in Fig. 4.11d), implying from the MS-EVA point of view that the instability is indeed baroclinic, and moreover, purely baroclinic. This result is exactly what is expected for an unstable Eady mode. Our previous assertion with the MS-EVA on baroclinic instability is thence validated here.

To continue the validation, we extend the MS-EVA diagnosis to the neutrally stable dataset. The energetics computed are graphed in Figs. 4.12 and 4.13. From these figures the patterns are fairly symmetric, and the unstable scenario described before totally disappears here. As $k = k_{crt}$, there should be no transfer and nor conversion for the system as a whole. This turns out to be true with our calculation. Although b^M_n, BC, and BT_H oscillate from negative to positive or vice versa, they average to zero on every horizontal level of the model. To see this, a selected group of MS-EVA terms are plotted versus x for level 10. (All other levels have shown the same trend. We choose level 10 simply because the perturbation is enhanced there.) All these energetic functions are nearly sinusoidal, which, if averaged over an integer number of wavelengths, result in values at least two order smaller than their respective amplitudes. It is fairly reasonable to say that within the computational error tolerance, their horizontal averages are nil. These vanishing transfers and conversions from another aspect validate our MS-EVA diagnosis for baroclinic instabilities.

We close this section with a pictorial representation of the scenario of baroclinic instabilities. Sketched in Fig. 4.15 is a cartoon of how several MS-EVA-represented processes (roughly) make such an instability. We examine the horizontally averaged properties only (the bracket signifies an averaging over the x-y plane). Originally, there is no perturbation energy in

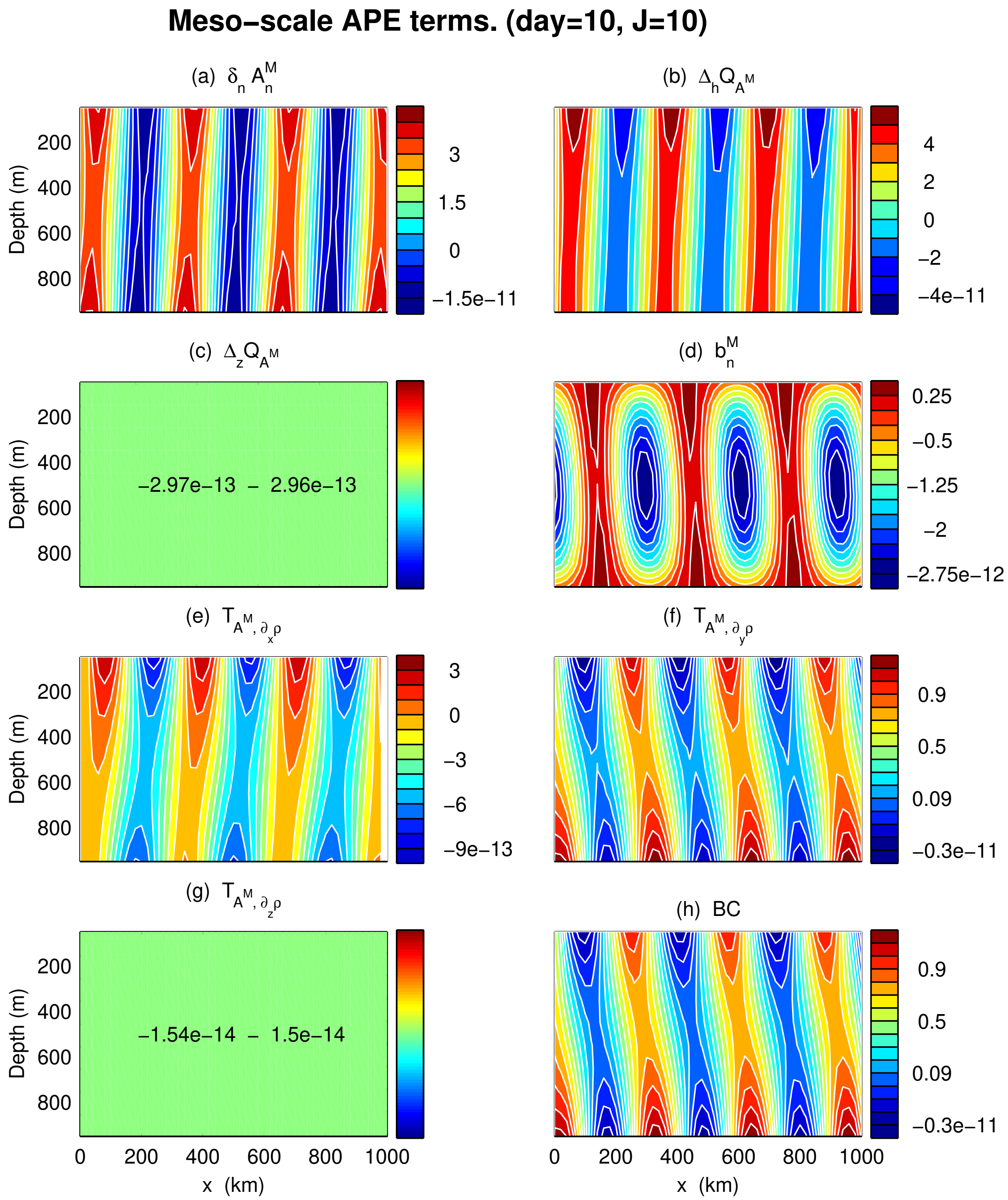

Figure 4.5: (a)-(g) Meso-scale potential energetics (in m^2/s^3) of the unstable Eady mode for day 10 on section j=10 (y = 500 km). Those insignificant terms are masked for clarity. (h) $BC = T^{0\to1}_{A^M,\partial_h\rho} + T^{0\to1}_{A^M,\partial_z\rho} + TS^{0\to1}_{A^M}$ ($TS_{A^M} = 0$ with this Eady model setting).

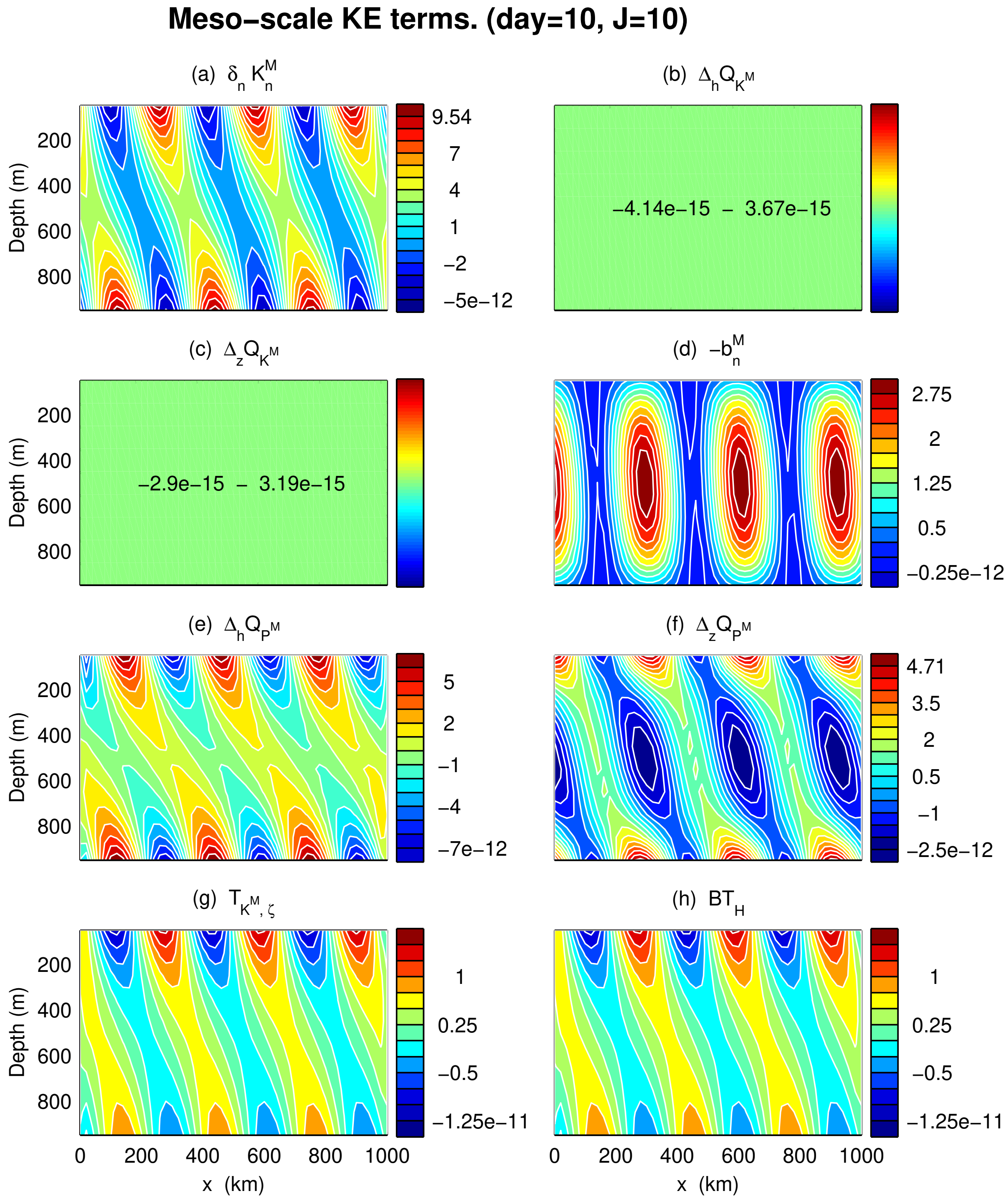

Figure 4.6: As Fig. 4.5, except for kinetic energetics. Note the two transfer terms $T_{K^M,w\partial_z}$ and $T_{K^M,\partial_z v}$ are not shown here. They are less than $O(10^{-15})$ and are negligible compared to $T_{K^M,\zeta}$.

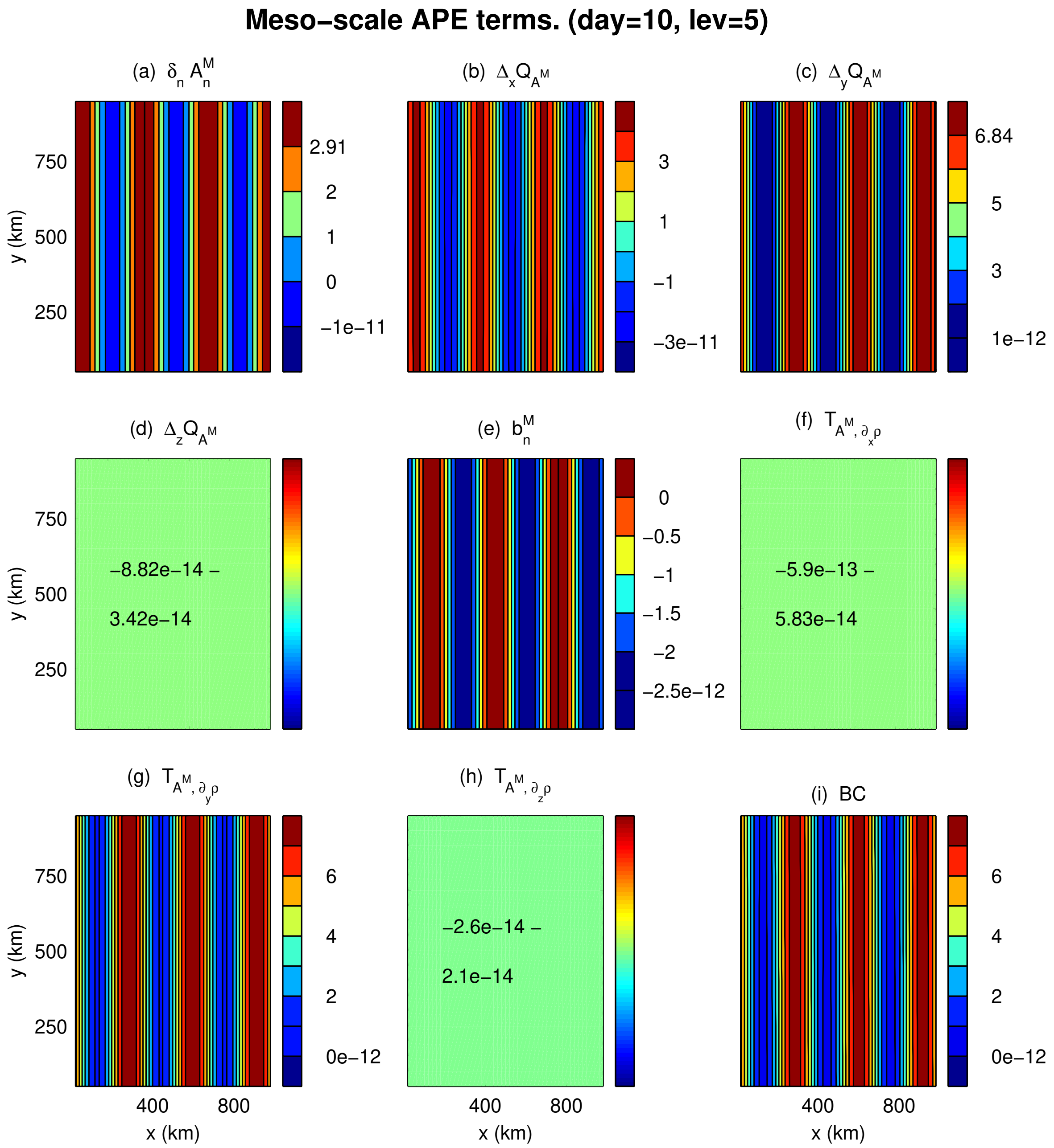

Figure 4.7: Meso-scale APE terms (in m^2/s^3) of the unstable Eady mode for day 10 on the mid-depth level (500 m).

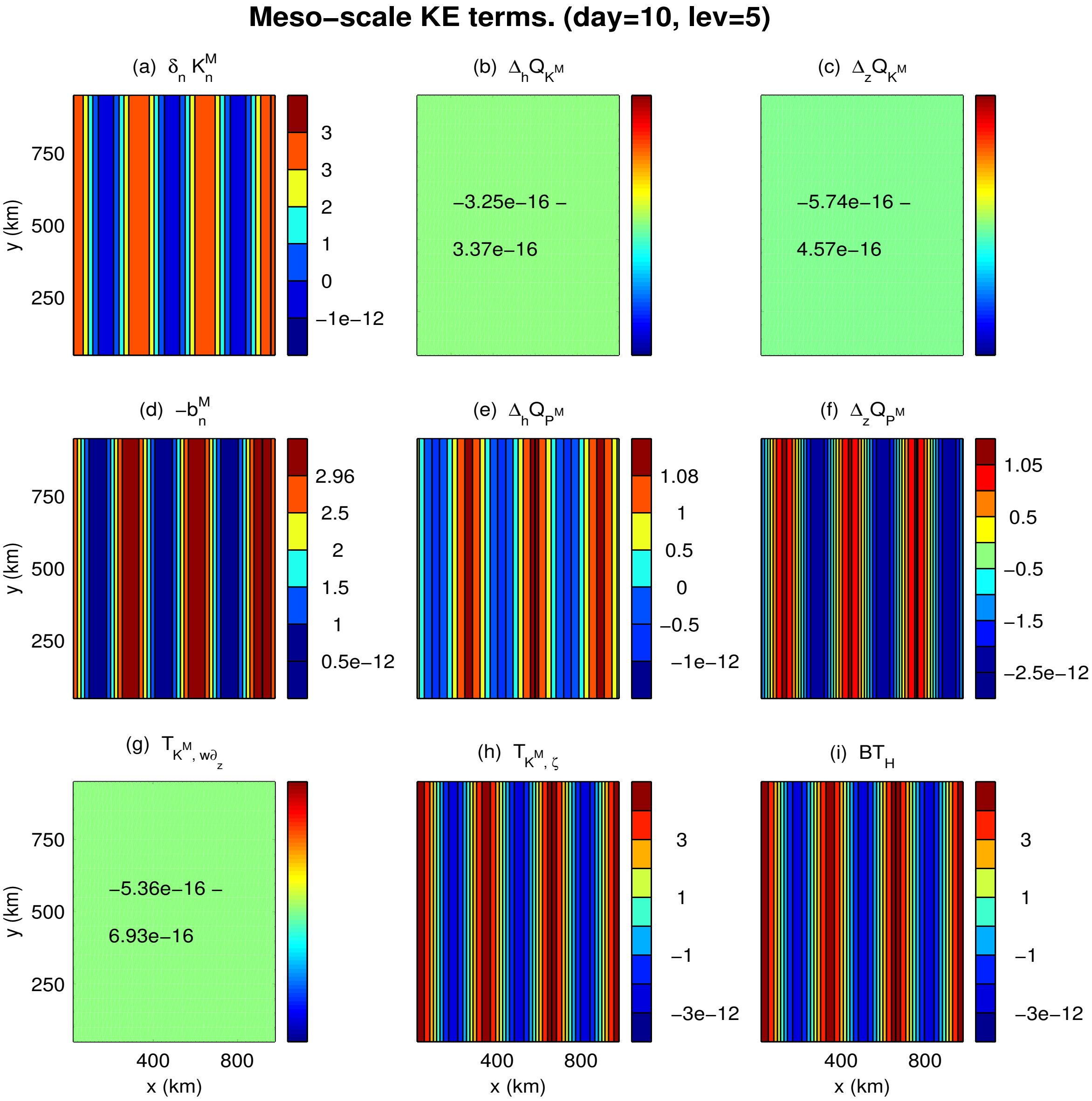

Figure 4.8: As Fig. 4.7, except for KE. The term $T_{K^M,\partial_z v}$ is not shown here. It is of order $O(10^{-16})$

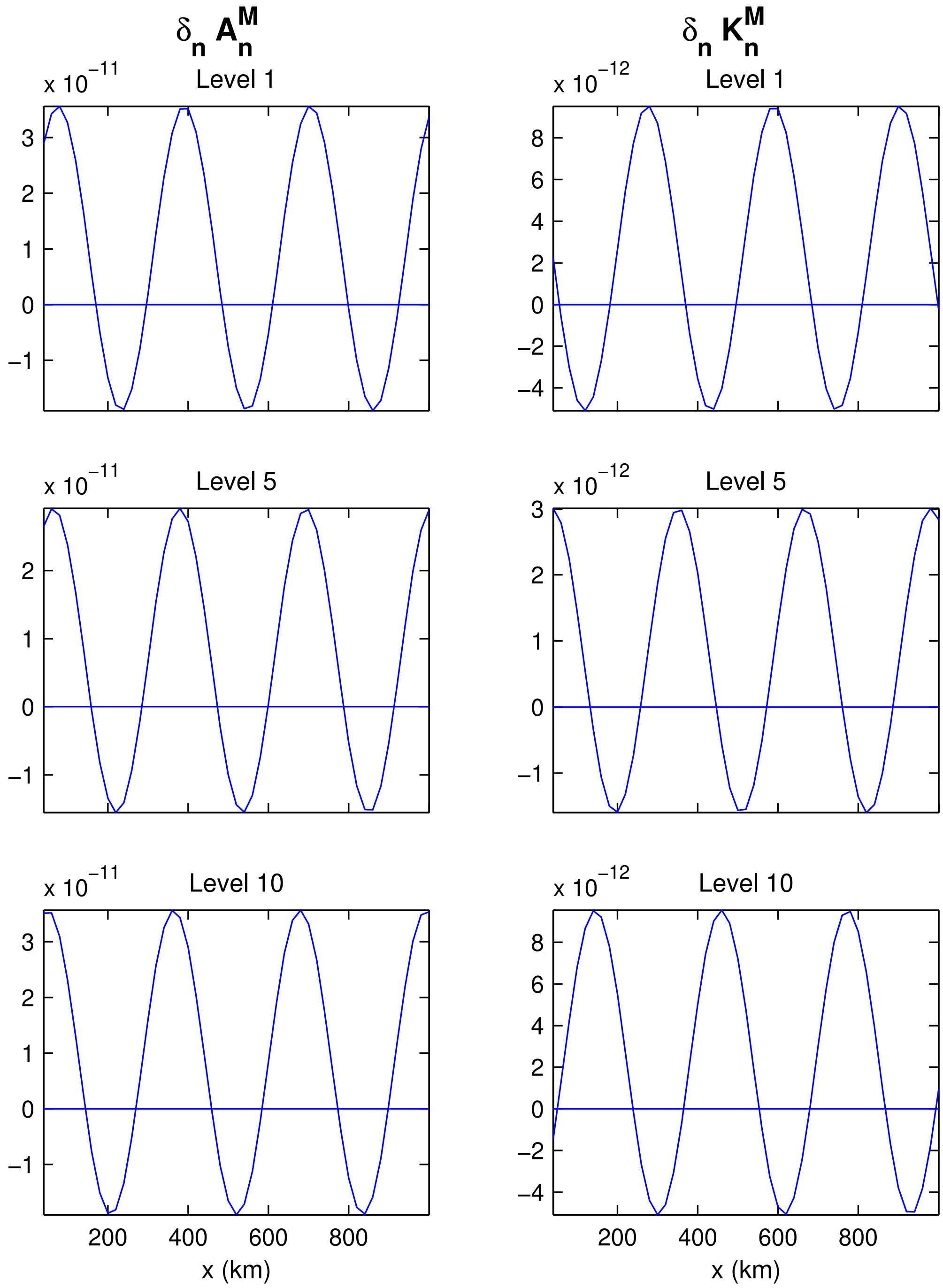

Figure 4.9: Time rates of change of meso-scale kinetic energy (right panel) and available potential energy (left panel) as functions of x for the top (level 1), mid-depth (level 5), and bottom levels (level 10) of the unstable Eady mode.

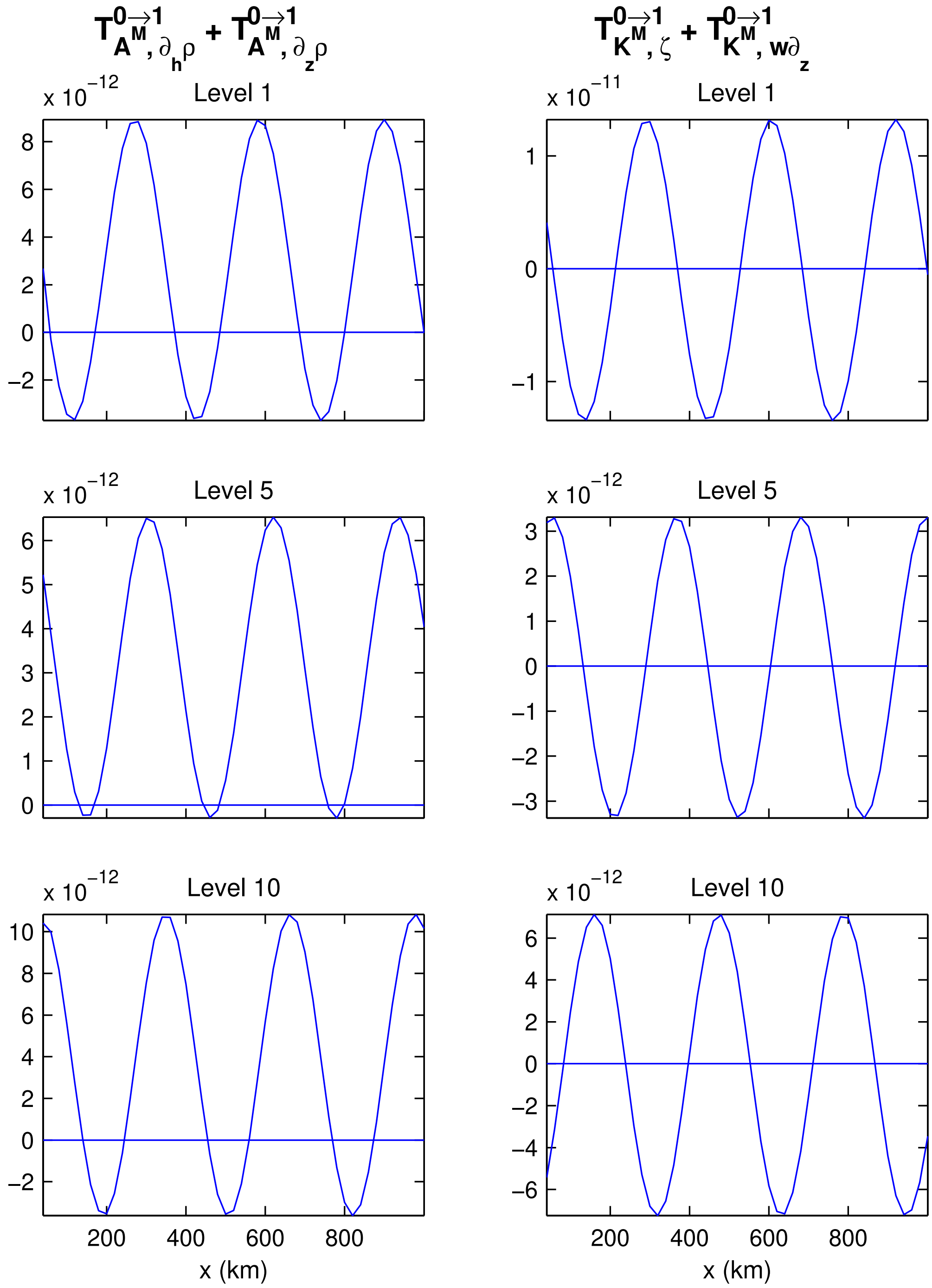

Figure 4.10: As Fig. 4.9, but for $BC = T^{0\to1}_{A^M,\partial_h\rho} + T^{0\to1}_{A^M,\partial_z\rho} + TS^{0\to1}_{A^M}$ (left panel) and $BT_H = T^{0\to}_{K^M,\zeta} + T^{0\to}_{K^M,w\partial_z}$ (right panel).

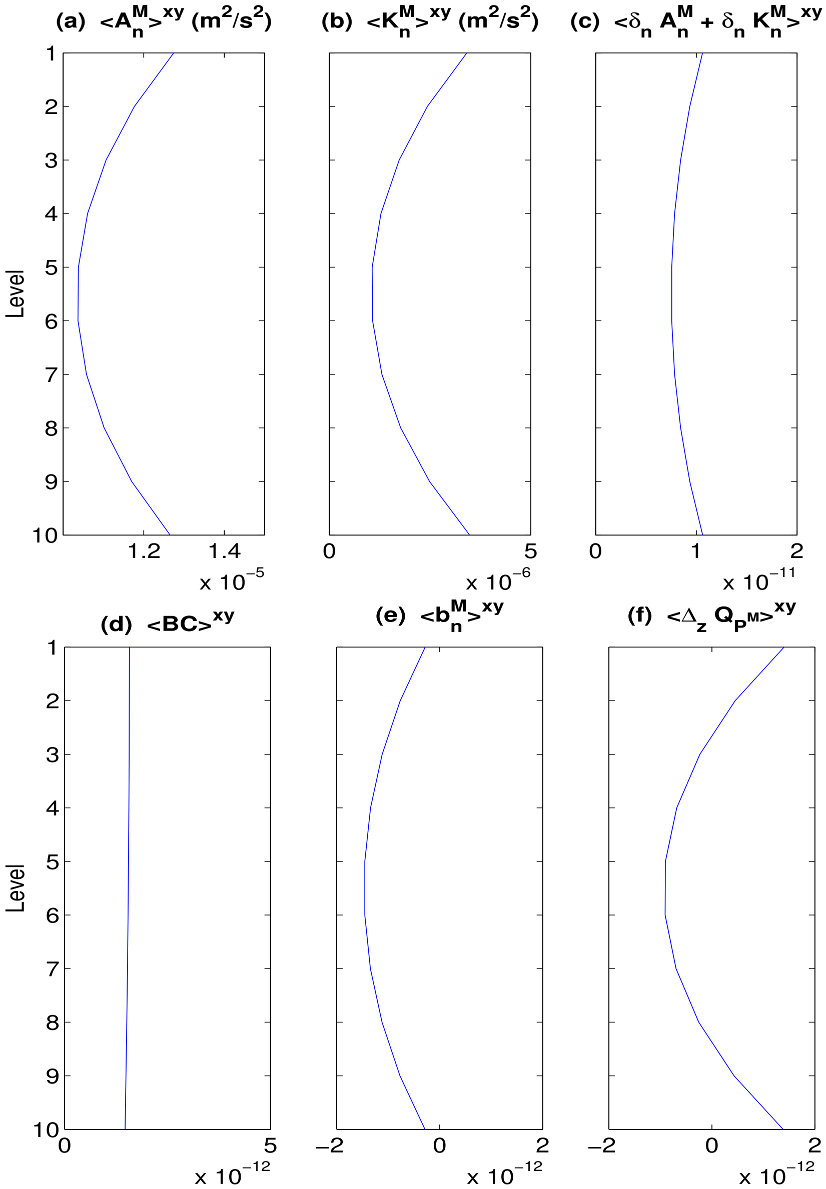

Figure 4.11: Horizontally averaged energies $\langle A_n^M \rangle^{xy}$ (a) and $\langle K_n^M \rangle^{xy}$ (b) (in m^2/s^2), and energetic terms $\left\langle \hat{\delta}_n A_n^M \right\rangle^{xy} + \left\langle \hat{\delta}_n K_n^M \right\rangle^{xy}$ (c), $\langle BC \rangle^{xy}$ (d), $\langle b_n^M \rangle^{xy}$ (e), and $\langle \Delta_z Q_P \rangle^{xy}$ (f) (in m^2/s^3).

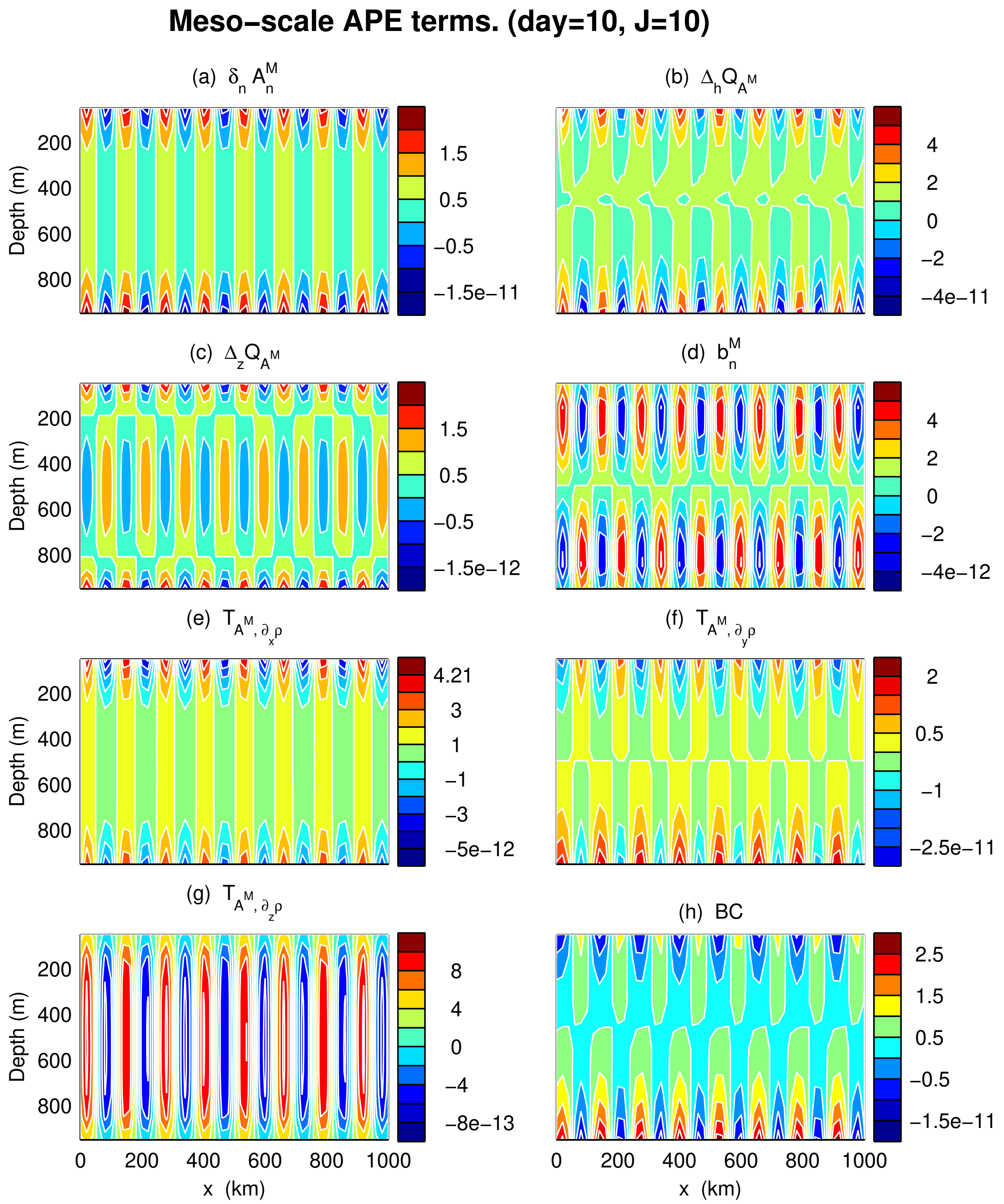

Figure 4.12: Meso-scale potential energetics (in m^2/s^3) of the neutrally stable Eady mode for day 10 on section y=500 km.

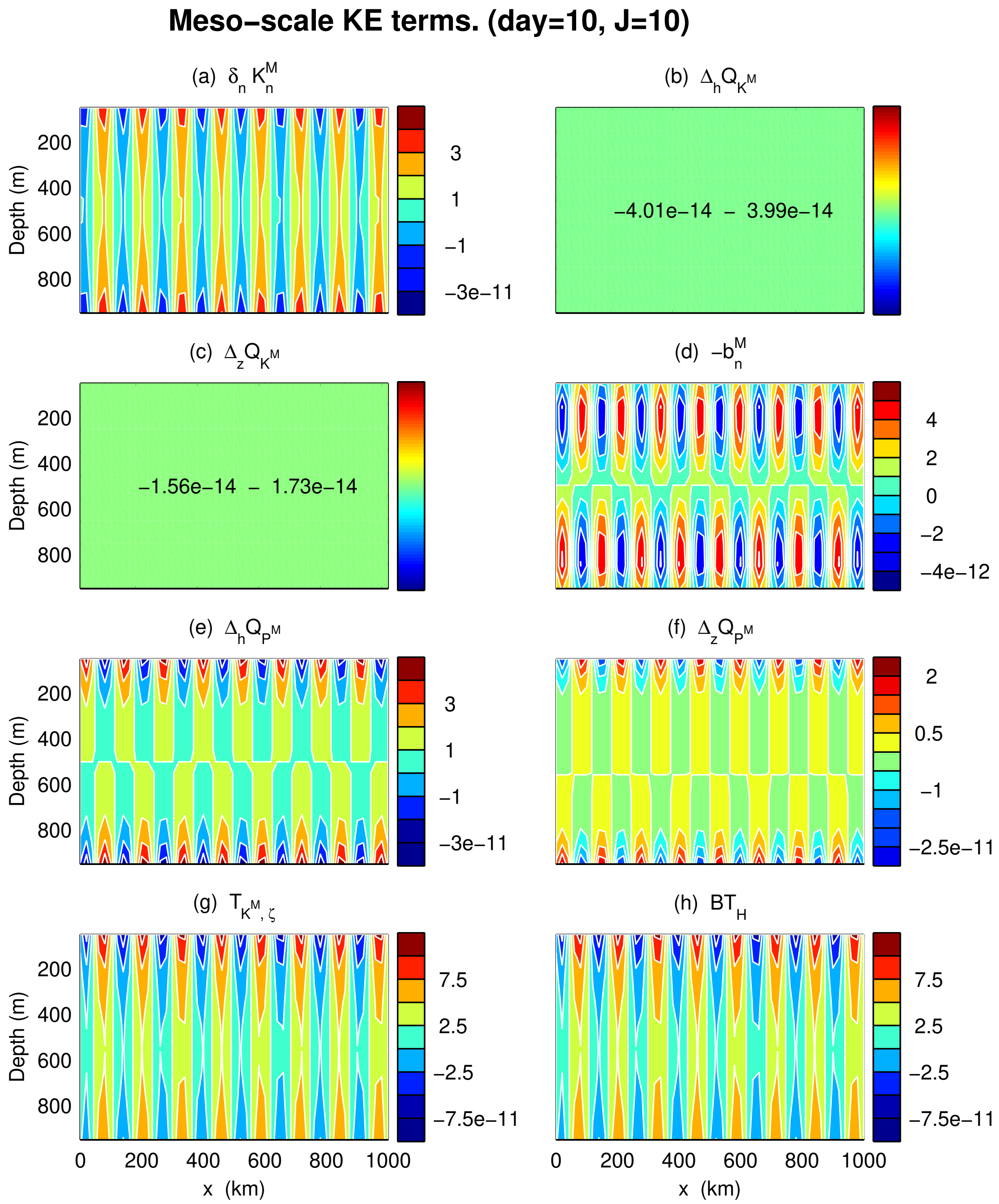

Figure 4.13: As Fig. 4.12, but for kinetic energetics.

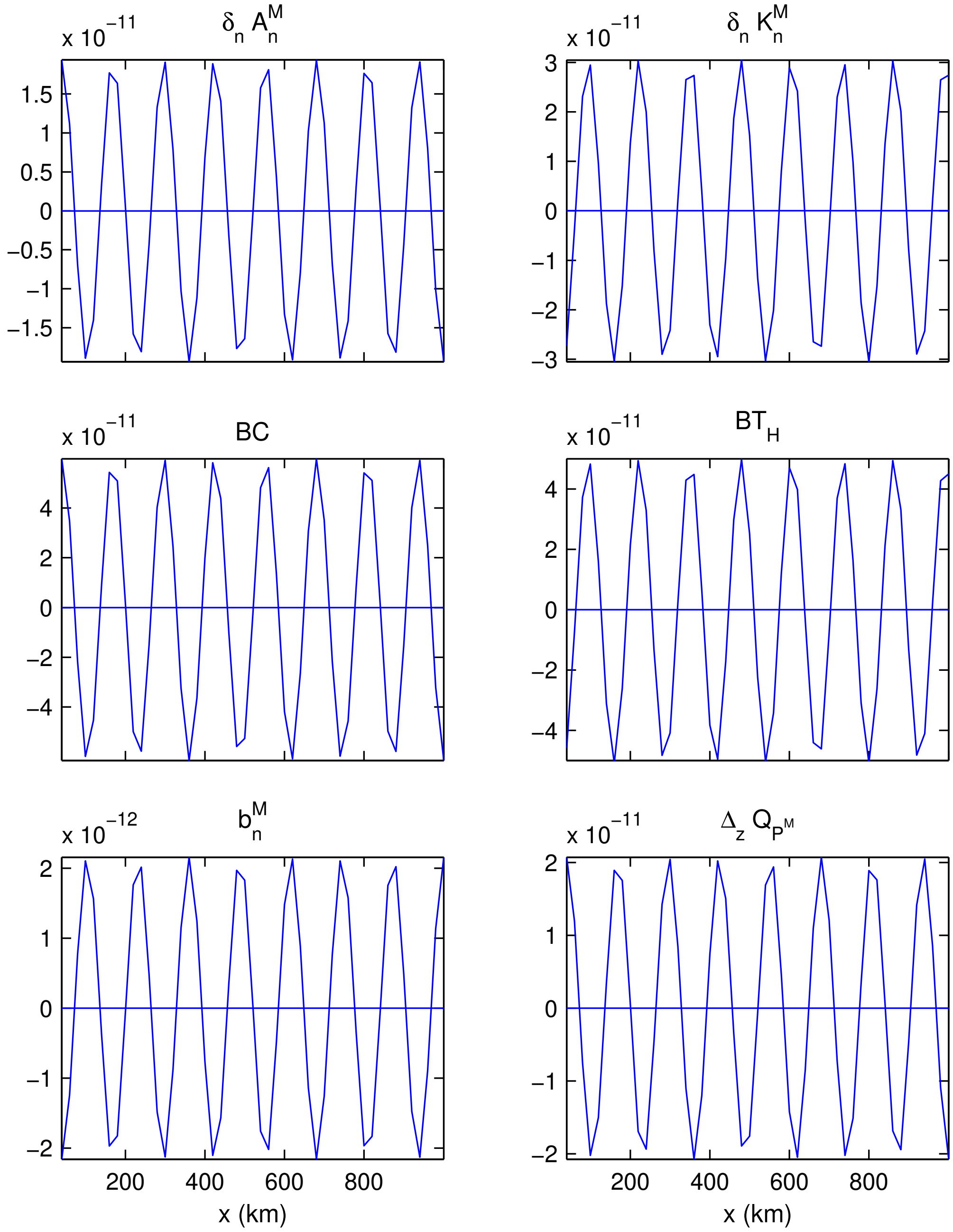

Figure 4.14: A selected set of MS-EVA terms as functions of x for a neutrally stable Eady mode. These plots are for level 10 (bottom level), but all other levels are similar (except for the amplitude). Notice that all these terms are nearly sinusoidal. Averaging of them over an integer number of wavelengths results in values at least two order smaller than their respective amplitudes.

presence. Both $\left\langle A_n^M \right\rangle^{xy}$ and $\left\langle K_n^M \right\rangle^{xy}$ are zero. When a baroclinic instability occurs, a part of potential energy is released from the large-scale reservoir, directing toward the meso-scale window of APE. This release is achieved through the transfer process represented by $\langle BC \rangle^{xy}$. It is more or less uniform through the water column, though from Fig. 4.11d we do see an increasing trend with height. Due to this transfer, the meso-scale APE is increased accordingly. But the increase is different level by level, as there exists a sink for $\left\langle A_n^M \right\rangle^{xy}$, the buoyancy conversion, which is minimal at the two boundaries and maximal in the middle. The buoyancy conversion carries the perturbation potential energy over the APE-KE bridge to activate the meso-scale motion. The converted energy would be piled up at the mid-depths, if there were no other process to help evacuate it. The fact is that, there does exist such a process. It is the vertical pressure work, which is represented by $\langle \Delta_z Q_{PM} \rangle^{xy}$. The energetic scenario is now clear: When a baroclinic instability occurs, the large-scale APE is released to feed the growth of $\left\langle A_n^M \right\rangle^{xy}$, while $\left\langle A_n^M \right\rangle^{xy}$ is converted simultaneously into $\left\langle K_n^M \right\rangle^{xy}$ through the buoyancy conversion. The pressure work greets at the mid-depths (roughly) the arrivals from the APE side, and ushers them immediately upstairs or downstairs. The outcome of the whole process is simple. It is two dumbbell-shape vertical profiles, for the meso-scale available potential energy and kinetic energy (the bottom panel of Fig. 4.15. Also see Figs. 4.11a,b). This kind of vertical structure for either $\left\langle A_n^M \right\rangle^{xy}$ or $\left\langle K_n^M \right\rangle^{xy}$ implies that, when the instability ceases, all the disturbances, either in horizontal velocity or in density, will be trapped near the two vertical boundaries, a result in precise agreement with what we have observed in Fig. 4.3. The MS-EVA is thence validated again with success.

4.5 MS-EVA validation with a barotropic stability model

We have verified with the Eady model that, a system is baroclinically unstable if the spatially averaged BC is positive; and vice versa. We have also verified with the same model that the spatially averaged BT vanishes (BT_V is insignificant) when no barotropic instability occurs. The remaining question is then whether a positive so-averaged BT implies a barotropically unstable system. In this section we will give this question an answer with Kuo's model, a model admitting only barotropic instabilities.

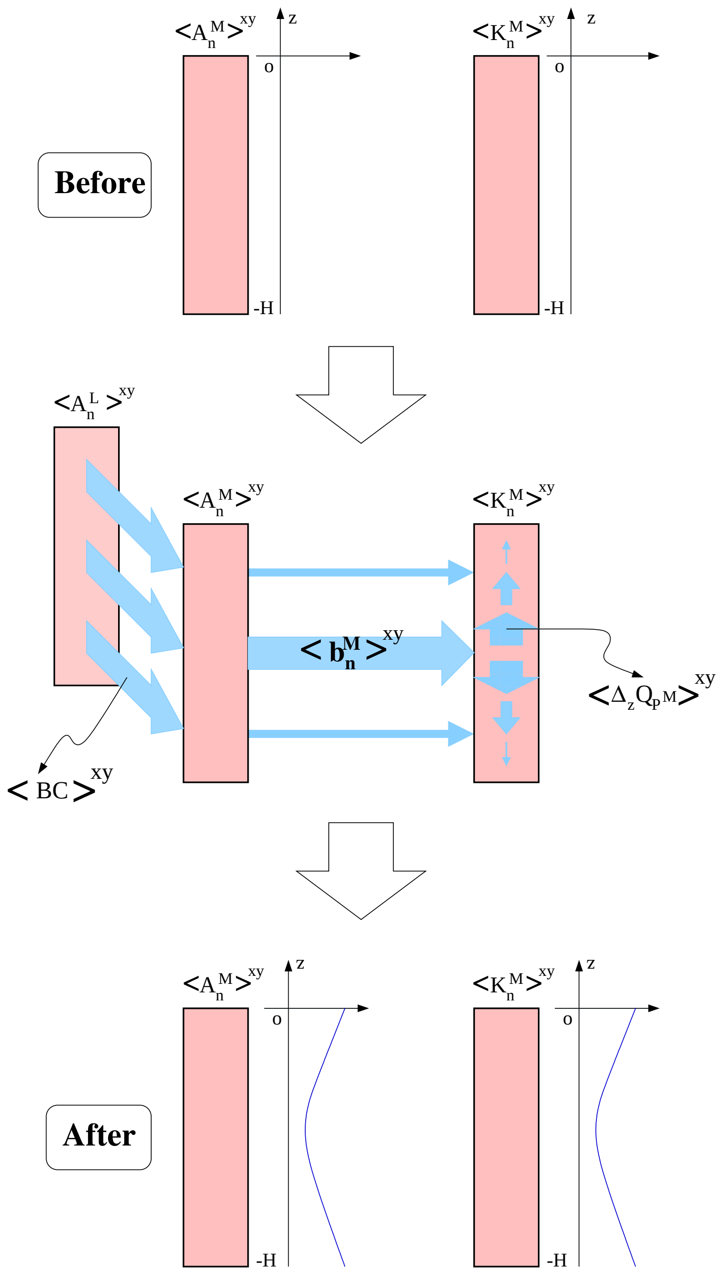

Figure 4.15: A cartoon of the energetic processes for an Eady-like baroclinic instability. Before disturbances grow, the averaged perturbation APE and KE (resp. $\langle A_n^M \rangle^{xy}$ and $\langle K_n^M \rangle^{xy}$) are zero (top row). (The angle bracket represents an averaging over the x-y plane.) When the instability happens, potential energy is transferred from the large-scale window toward the meso-scale window, and the measure of this transfer, $\langle BC \rangle^{xy}$ is uniformly distributed in the vertical direction (middle row). At the same time, the meso-scale APE is converted to the meso-scale KE. The conversion is maximized at the middle depth, where the converted energy is brought upward and downward by the vertical pressure work ($\langle \Delta_z Q_{PM} \rangle^{xy}$). The results of these processes are two dumbbell-shape vertical distributions of $\langle A_n^M \rangle^{xy}$ and $\langle K_n^M \rangle^{xy}$ (bottom row).

4.5.1 An overview of Kuo's barotropic instability model

Barotropic instabilities can be completely described by a barotropic model (Pedlosky, 1979). In this section, we examine the stability of a zonal barotropic jet $\bar{u} = \bar{u}(y)$. On an f-plane, the two-dimensional version of the equation set (4.18) - (4.22) is, after translated to its dimensional form,

$$\begin{aligned}
\frac{\partial u'}{\partial t} &= -\bar{u}\frac{\partial u'}{\partial x} - v'\frac{\partial \bar{u}}{\partial y} + fv' - \frac{1}{\rho_0}\frac{\partial P'}{\partial x}, && (4.77)\\
\frac{\partial v'}{\partial t} &= -\bar{u}\frac{\partial v'}{\partial x} - fu' - \frac{1}{\rho_0}\frac{\partial P'}{\partial y}, && (4.78)\\
0 &= \frac{\partial u'}{\partial x} + \frac{\partial v'}{\partial y}. && (4.79)
\end{aligned}$$

The basic flow $\bar{u}(y)$, which depends on the meridional coordinate y only, requires a background pressure field

$$\bar{P}(y) = \bar{P}(y) = -\rho_0 f \int_0^y \bar{u}(y)\ dy$$

by geostrophy.

Let

$$\begin{bmatrix} u' \\ v' \\ P' \end{bmatrix} = \mathrm{Re}(\begin{bmatrix} \tilde{u}(y) \\ \tilde{v}(y) \\ \tilde{P}(y) \end{bmatrix} e^{i(kx-\omega t)}). \tag{4.80}$$

Eqs. (4.77) - (4.79) are thereby reduced to

$$\frac{d^2\tilde{v}}{dy^2} + \left(\frac{\ddot{u}}{c-\bar{u}} - k^2\right)\tilde{v} = 0, \tag{4.81}$$

where $c = \omega/k$ is the phase speed, and $\ddot{u} = \frac{d^2\bar{u}}{dy^2}$ the gradient of the background potential vorticity. By Rayleigh's theorem, $\ddot{u}$ must change sign over its definition domain in order for any instability to occur (see Pedlosky, 1979).

Consider a channel bounded to the north and south respectively by $y = L$ and $y = -L$. The boundary condition for this system is then

$$\tilde{v} = 0, \qquad \text{at } y = \pm L. \tag{4.82}$$

Eqs. (4.81) and (4.82) form the eigenvalue problem of barotropic instability with a basic flow $\bar{u}$. It is the same as the one obtained by Kuo (1949) with a quasi-geostrophic model ($\beta = 0$ here), except for the dependent variable and hence the boundary condition. For convenience, we still call it a Kuo model. With $\tilde{v}$ solved, the other two fields are obtained easily:

$$\tilde{u} = \frac{i}{k}\frac{d\tilde{v}}{dy}, \tag{4.83}$$

$$\tilde{P} = \rho_0 \frac{i}{k}\left[(\dot{\bar{u}} - f)\tilde{v} - (\bar{u} - c)\frac{d\tilde{v}}{dy}\right]. \tag{4.84}$$

Consider now a cosine background jet:

$$\bar{u}(y) = \frac{1 + \cos(\pi\frac{y}{L})}{2}\bar{u}_{\max} = \bar{u}_{\max}\cos^2\left(\frac{\pi}{2}\frac{y}{L}\right) \tag{4.85}$$

(see Fig. 4.19a). This configuration was originally studied by Kuo (1949) (also seen in Kuo, 1973). With it the ambient potential vorticity gradient is

$$\bar{q}_y = -\ddot{\bar{u}} = -\frac{\pi^2}{2L^2}\bar{u}_{\max}\cos\frac{\pi y}{L}, \tag{4.86}$$

which vanishes at $y = \pm\frac{L}{2}$, meeting the necessary condition for a barotropic instability (cf. Fig. 4.19b).

For convenience, scale the variables as follows

$$(\bar{u},\ c,\ \tilde{v}) \sim \bar{u}_{\max}, \qquad y \sim L, \qquad k \sim \frac{1}{L}. \tag{4.87}$$

Using the same notation as before, the eigenvalue problem is re-cast as

$$\frac{d^2\tilde{v}}{dy^2} + \left(\frac{\ddot{\bar{u}}}{c-\bar{u}} - k^2\right)\tilde{v} = 0, \tag{4.88}$$

$$\tilde{v} = 0 \qquad \text{at } y = \pm 1, \tag{4.89}$$

with

$$\bar{u} = \cos^2(\frac{\pi}{2}y),$$

and

$$\bar{P} = -\int_0^y \bar{u}(y)dy = -\frac{1}{2\pi}(\pi y + \sin\pi y).$$

Note all variables are now understood to be nondimensional, and this convention will continue until the end of next subsection, where dimensions are recovered to generate the MS-EVA dataset.

4.5.2 Particular solutions and dataset formation

The eigenvalue problem (4.88)-(4.89) is solved with the approach used by Kuo (1949) (see also, Kuo, 1973; Pedlosky, 1979). We need a neutrally stable solution and an unstable solution. For the former, there exists an eigenvalue

$$c = \bar{u}(y_c) = \frac{1}{2}$$

which gives

$$\frac{\ddot{u}}{c - \bar{u}(y)} = -\pi^2 = \text{const},$$

and hence a simplified version of (4.88):

$$\frac{d^2\tilde{v}}{dy^2} + (\pi^2 - k^2)\tilde{v} = 0. \tag{4.90}$$

This equation together with (4.89) has a nontrivial solution (Pedlosky, 1979. p. 508)

$$k = \frac{\sqrt{3}}{2}\pi, \qquad \tilde{v} = \cos\frac{\pi y}{2}. \tag{4.91}$$

The perturbation flow pattern obtained therefrom is shown in Fig. 4.17.

The choose of the unstable mode is not arbitrary. For a valid dataset, as in the Eady model case, the eigenvalue $c = c_r + ic_i$ must be such that $|c_i| \ll |c_r|$ while c_i is significantly greater than zero. Again, a k near its critical value $\frac{\sqrt{3}}{2}\pi$ (on the unstable side, of course) will yield such a solution. We pick $k = 0.75\pi$. Use the shooting method (e.g., Press et al., 1992) to solve (4.88). The convergence is usually pretty fast, and the resulting eigenvalue thus obtained is $c = c_r + ic_i = 0.4504 + 0.0476i$. Clearly $|c_i|$ is much smaller than $|c_r|$, allowing an extraction of several cycles from the solution to form the dataset. Plotted in Fig. 4.16 are the growth of perturbation [subplot (a)] and the corresponding complex eigenfunction $\tilde{v} = \tilde{v}_r(y) + i\ \tilde{v}_i(y)$ [subplots (b) and (c)]. Notice the inflection of $\tilde{v}_r$ near $y = \pm\frac{1}{2}$. Its implication will be clear soon. In obtaining the eigenfunction, the magnitude of $|\tilde{v}_r|$ has been made small so as to have the solution valid throughout. (Specifically its magnitude is controlled such that $\left.\frac{d\tilde{v}_r}{dy}\right|_{y=-1} = 0.1$.) It is thus not normalized as usual. For an instantaneous pattern of the solution, refer to Figs. 4.18a, b, and c.

The solutions are now re-dimensionalized to generate the datasets. Choose $L = 100$ km, $\bar{u}_{\max} = 1$ m/s, $f = 10^{-4}$ 1/s, which are typical of western ocean jets, and let x run over

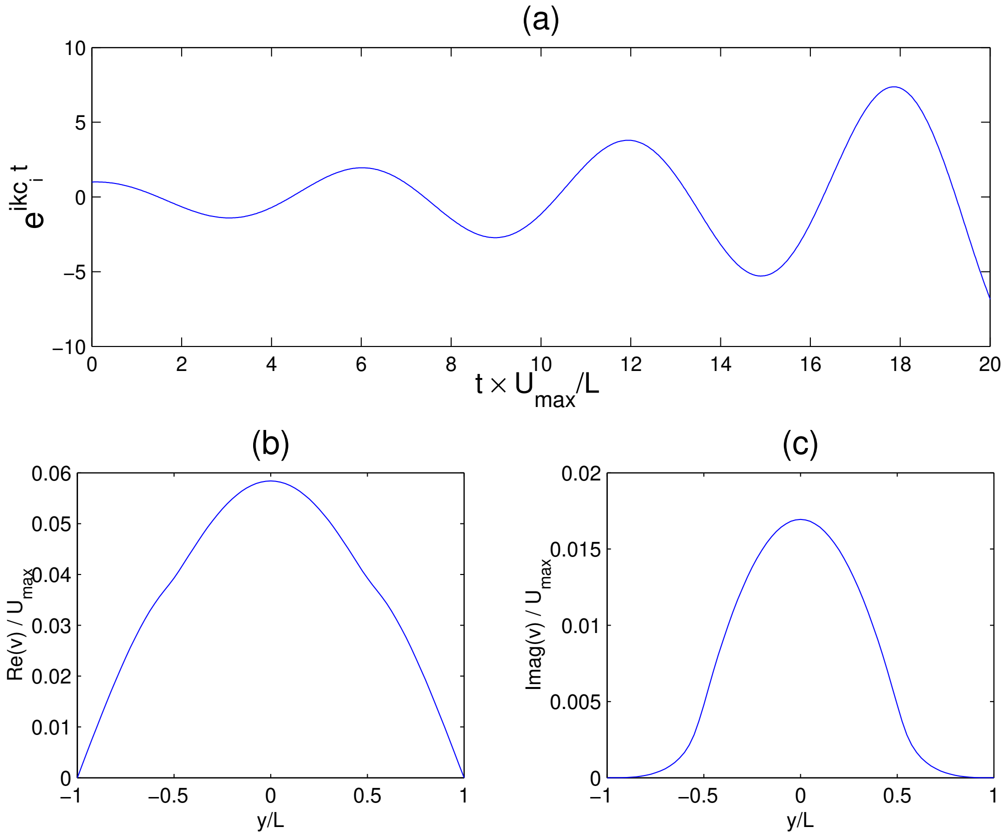

Figure 4.16: The growth of perturbation (a) and eigenvector $\tilde{v} = \tilde{v}_r + i\tilde{v}_i$ [(b) and (c)] of the unstable mode chosen to generate the barotropic instability dataset. The eigenvector has been multiplied by a factor such that $\left.\frac{d\tilde{v}_r}{dy}\right|_{y=-L} = 0.1\bar{u}_{max}$. (All variables are dimensional here.)

[-2L, 2L], and t span exactly two cycles (approximately 21 days for the neutrally stable mode and 14 days for the unstable mode). With the k's picked, the x extends at least one wavelength for either of the datasets, allowing an application of the zonal average whenever necessary. The solutions are mapped on an Arakawa B-grid with 40×80 grid points ($\Delta x = 10$ km, $\Delta y = 0.5$ km), and sampled at 2^{10}=1024 time moments. These gridded solution sequences are now ready for MS-EVA.

4.5.3 MS-EVA validation with the Kuo model

To start the MS-EVA study, we first need to specify options for the multiscale analysis and synthesis. As for the Eady model case, a two-scale window decomposition is performed with j_0=0 and a periodic extension. This specification is supposed to reconstruct the perturbation field on the meso-scale window, i.e., $p^{\sim 1}$ should be equal to p' for any field p. The results confirm this claim. Shown in Figs. 4.17a, b, and 4.18a, b are instantaneous maps of the meso-scale syntheses for the velocity components of the two modes under study. In either the neutrally stable or the unstable case, they are exactly equal to u' and v', respectively. By (4.79), the meso-scale continuity equation, these u' and v' can be used to define a perturbation streamfunction ψ' such that

$$u' = -\frac{\partial \psi'}{\partial y}, \qquad v' = \frac{\partial \psi'}{\partial x}. \tag{4.92}$$

This derived field, which is respectively contoured in Figs. 4.17d and 4.18d for the two modes of concern, describes more clearly the perturbation flow. Observe that for the unstable mode, all the phase lines bulge in the middle toward the west. Compared to Fig. 4.19a, this pattern shows a tilting trend of phase lines against the distribution of the background velocity. That is to say, wherever $\frac{d\bar{u}}{dy} > 0$, the phase lines slope to the west with latitude (slope being negative), and vice versa. This distribution has been proved to be necessary for a barotropic jet to lose stability, as only through this kind of tilting can the Reynolds stress extract energy from the basic flow (Pedlosky, 1979, p. 511).

Another conspicuous feature of the unstable modal solution is the kinky variation of u' with y (Fig. 4.18a). The kinks occur at the y positions where the two inflections on the eigenvector $\tilde{v}(y)$ appear (Fig. 4.16b). It seems that there are two discontinuities near $y = \pm\frac{L}{2}$ which separates the interior flow from its side peers. To understand why there should be such a phenomenon, recall that for a barotropic instability to exist, it requires not only that a basic potential vorticity gradient $\bar{q}_y$ change sign through $y \in [-L, L]$, but also that $(\bar{u} - c_r)$ and $\bar{q}_y$ be positively correlated over the same domain (Kuo's theorem). (See Kuo, 1973, pp. 277-279.) Although this positive correlation requirement is stated in an integral form, locally (in y) it may also apply wherever the system is more or less meridionally isolated. This is to say, the sign of the product $(\bar{u} - c_r) \times \bar{q}_y$ may be pivotal to the instability, and hence the zero points of $\bar{u} - c_r$ and $\bar{q}_y$ are critical to the dynamics. Recall here $c_r = 0.4504\bar{u}_{\max}$, and

$$\bar{u}(y) = \frac{1}{2}\left(1 + \cos\frac{\pi y}{L}\right)\bar{u}_{\max}.$$

Thus

$$\bar{u}(y) - c_r = 0 \implies y = \pm 0.53L \tag{4.93}$$

$$\bar{q}_y = -\ddot{\bar{u}} = 0 \implies y = \pm 0.50L. \tag{4.94}$$

These critical y's divide the flow into five regimes, as sketched in Fig. 4.19c. In the middle regime (hatched), both $\bar{q}_y$ and $\bar{u} - c_r$ are positive, so their product is also positive. The regimes at the two ends, with both $\bar{q}_y$ and $\bar{u} - c_r$ being negative, also have a positive product $\bar{q}_y(\bar{u} - c_r)$. These polar regimes and the interior regime are separated near $y = \pm 0.5L$ by two narrow strips (with a width of only $0.03L$) where $\bar{q}_y(\bar{u} - c_r) < 0$, and just because of the opposite sign of $\bar{q}_y(\bar{u} - c_r)$, the dynamics in the strips could be completely different from the interior regime. The existence of the kinks in the pattern of u' near $y = \pm 0.5L$ is therefore not surprising. (Kuo's theorem tells only the possibility of such an existence, as it is the integral with respect to y over [-L, L] that matters.) We will soon see that the energetics in these two narrow strips do turn out to be completely different from their surrounding regimes.

The dynamic features above are now diagnosed with the MS-EVA. The computed energetic terms are contoured for the neutrally stable mode in Figs. 4.20a-f, and for the unstable mode in Figs. 4.21a-f. First look at $\hat{\delta}_n K_n^M$ [subplot (a)], the rate of change of the meso-scale kinetic energy. In either case, the map of $\hat{\delta}_n K_n^M$ shows strong variation in both x and y. The difference lies in the pattern of variation. For the neutrally stable mode, the variation is precisely sinusoidal and thus averages[7] to zero. That is to say, this mode does not admit

[7] In this section, if not specified, the average (or sum) is understood to be over a wavelength in x and the whole domain of y.

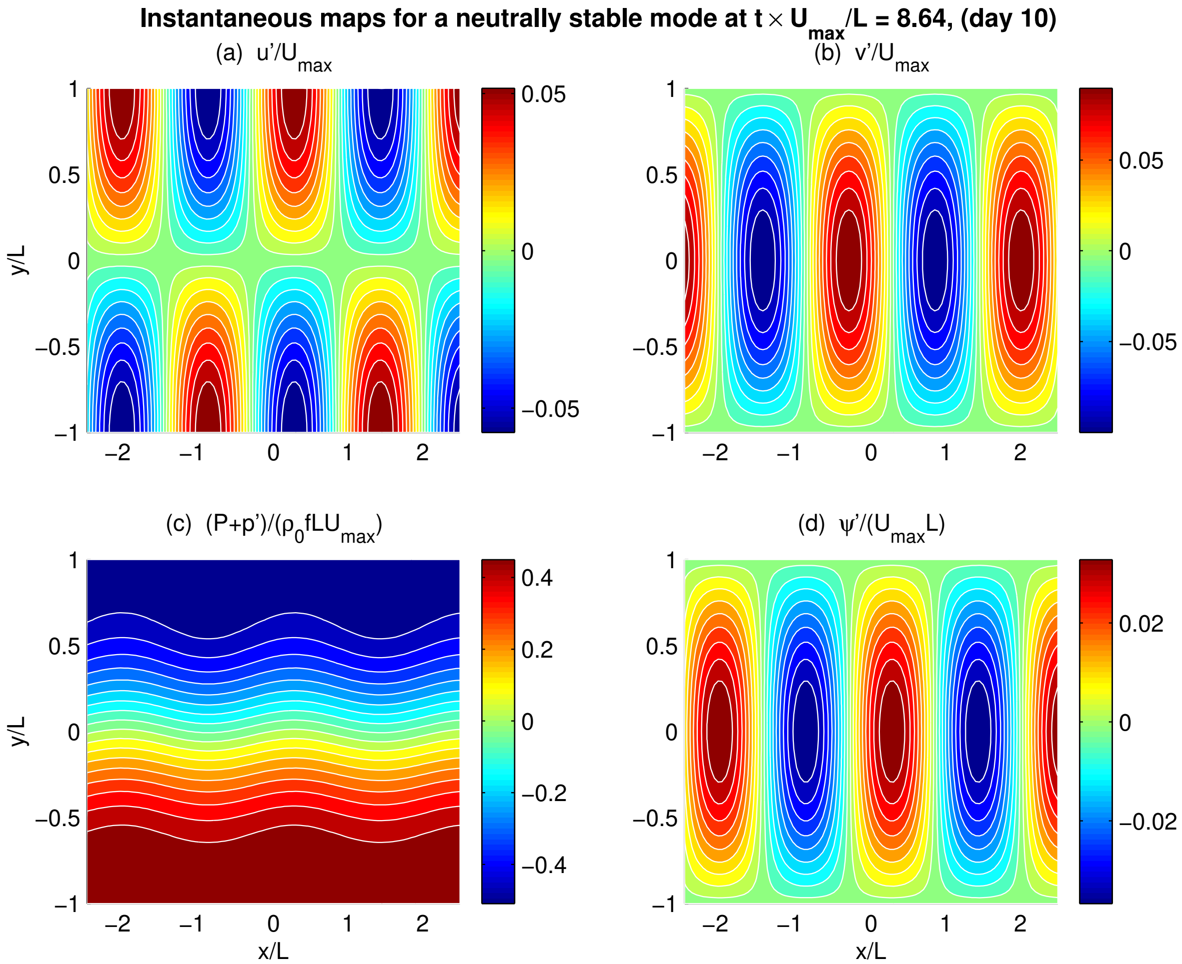

Figure 4.17: The day-10 instantaneous maps of u' (a), v' (b), total pressure $(\bar{P} + P')$ (c), and perturbation streamfunction ψ' (d) for the neutrally stable mode of the barotropic instability model $(kL = \frac{\sqrt{3}}{2}\pi)$.

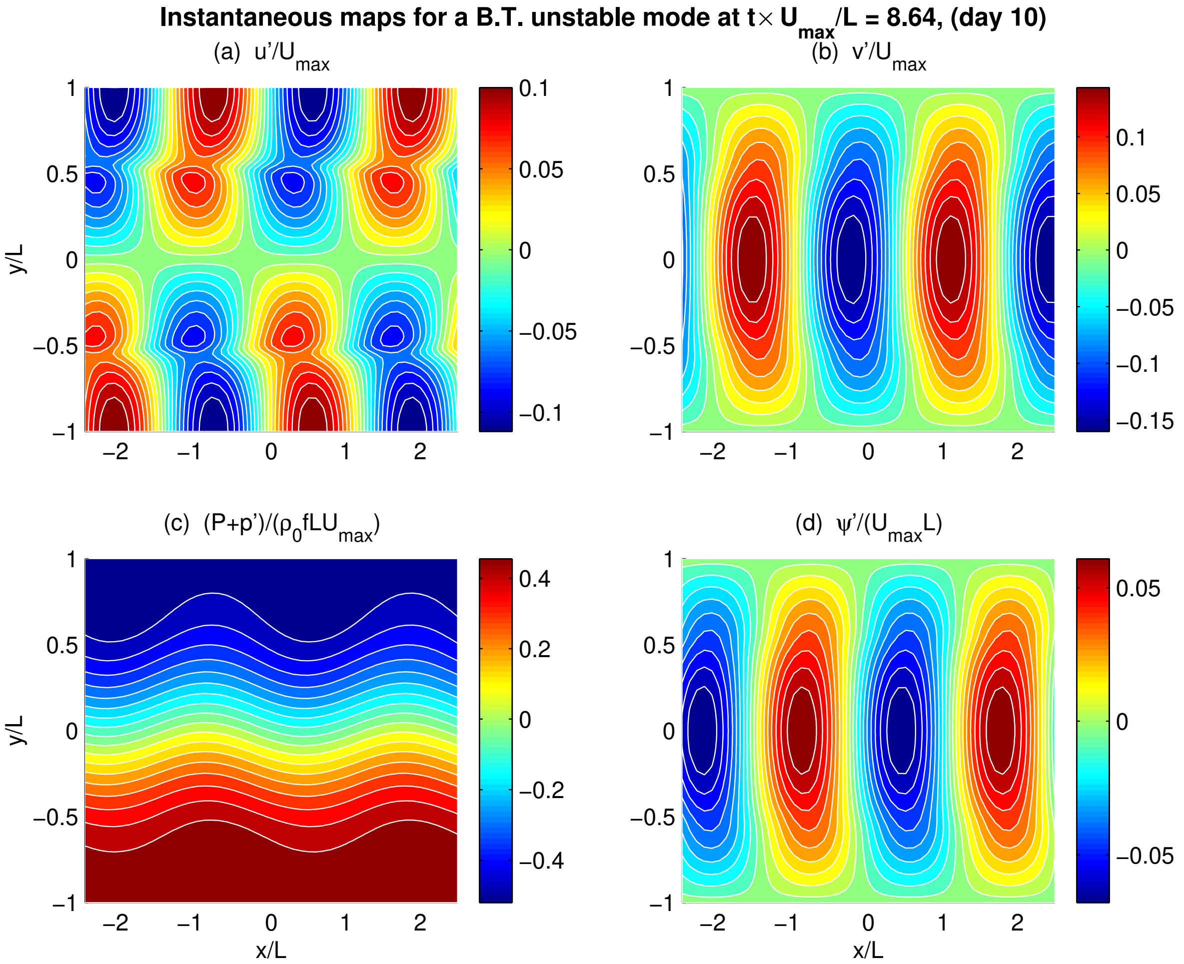

Figure 4.18: The day-10 instantaneous maps of u' (a), v' (b), total pressure $(\bar{P}+P')$ (c), and perturbation streamfunction ψ' (d) for the barotropically unstable mode ($kL = \frac{3}{4}\pi$).

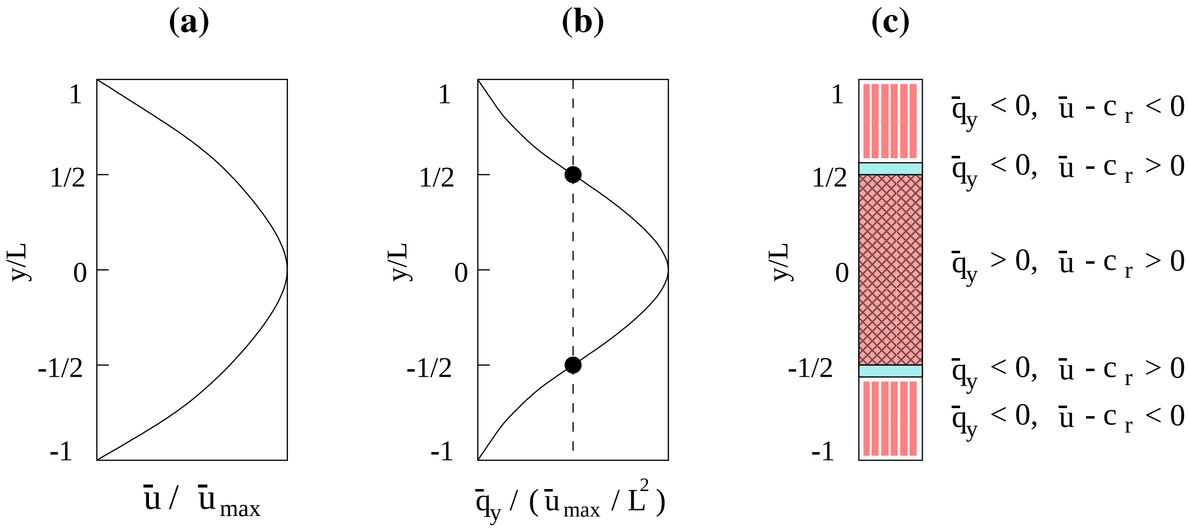

Figure 4.19: A sketch of the background velocity $\bar{u}(y)$ (a), and vorticity gradient $\bar{q}_y = -\frac{\partial^2 \bar{u}}{\partial y^2}$ (b) for the Kuo model. The five regimes resulting from this basic structure and the chosen phase speed c_r are indicated in (c).

any perturbation growth by our calculation. For the unstable case, however, things are quite different. The zonally-averaged $\hat{\delta}_n K_n^M$ (over one wavelength), as is shown in Fig. 4.22b, displays a positive residue over the whole domain of y, indicating that the mode is indeed unstable by the MS-EVA diagnosis.[8]

While $\hat{\delta}_n K_n^M$ implies whether the system is unstable, $BT = BT_H = T^{0\to 1}_{K^M,\zeta}$ (for a barotropic model, BT_V vanishes) tells whether the instability is barotropic. As expected, the BT for the neutrally stable mode vanishes, if summed horizontally (Fig. 4.20). The unstable mode, however, has an averaged BT (over both x and y) distinctly greater than zero. This fact is easily seen from Fig. 4.22d, which is the zonally-averaged profile of BT as shown in Fig. 4.21. Recall that a positive averaged BT means a barotropic instability. Our MS-EVA calculation is clearly in agreement with the analytical result.

The MS-EVA is therefore validated with the Kuo model. We proceed to investigate the energetic processes of the barotropic instability. Look at Fig. 4.21g, the map of perturbation energy, and its zonally averaged version Fig. 4.22a. A certain structure is displayed in both of these figures. We want to know how such a structure is generated through these processes from an originally unperturbed system.

[8]This claim is true only if the same-scale transfer is negligible, as is the case here (cf. Figs. 4.20h and 4.21h).

We have described earlier in this section that, for the unstable mode, there exist in the flow pattern two remarkable features, the westward bulging and the kinks near $y = \pm\frac{L}{2}$ (± 50 km). On the perturbation energy map, such features are also noticeable. From (Fig. 4.21g), the jet is energetic at the core while much more sluggish on the two wings. The contour lines tilt from the center toward the right while extending poleward. Along the two narrow critical bands, they are almost disconnected, separating the interior core from the northern and southern regimes. The energetics that account for the formation of this pattern are those shown in Figs 4.21a-f. The rate of change of K_n^M, $\hat{\delta}_n K_n^M$, has a spatial distribution similar to K_n^M itself. This is not a surprise, for K_n^M emerges from a horizontally uniform zero pattern. The remaining terms appear on the right hand side of the energetic equation. They stand for the mechanisms that account for the growth of disturbances. From the scaling, it seems that the pressure work [subplots (e) and (f)] dominates the balance, and this does account at least in part for the tilting trend of K_n^M. But, if one observes closely, the variations of $\Delta_x Q_{PM}$, $\Delta_y Q_{PM}$ are almost out of phase. A huge part of pressure working rate thence cancels out when the two summed together. Moreover, if a zonal averaging (over some integer number of wavelengths) is performed, neither of these two terms turns out significant compared to their peers on the energetic balance. Thus the pressure work does not function as substantially as it appears to, considering processes like instabilities are concepts based on average properties. In the present context, the advection and transfer and pressure work are of comparable importance, to say the least.

With the energetic balance clear, it is now possible to analyze the the processes that lead to the perturbation structure. Although the contour lines of $\hat{\delta}_n K_n^M$ bulge westward conspicuously, the map of $T_{K^M,\zeta}$ (Fig. 4.21d) shows no indication of this phenomenon. Among the other processes, the y advection (Fig. 4.21c) does not evidence much relevance, either. The only mechanisms causing the bulging are then the zonal advection and the horizontal pressure work. For the contour breaking of $\hat{\delta}_n K_n^M$, which corresponds to the kinky distribution of u' in Fig. 4.18a, things are quite different. Recall what we have analyzed with Kuo's theorem: the narrow strips where the kinks lie are a direct result of the model configuration and depend only on latitudes. On the map of an energetic term essential to these strips, their signature(s) should remain there in spite of any kind of zonal averaging. However, all the three terms, $\Delta_x Q_{K^M}$, $\Delta_x Q_{PM}$, and $\Delta_y Q_{PM}$, will be virtually gone if averaged in x over some integer wavelengths, despite those obvious "necks" on their maps at the critical latitudes. Therefore, in this case we need only focus on $T_{K^M,\zeta}$ and $\Delta_y Q_{K^M}$ for processes that lead to the instability structure.

To some extent, $T_{K^M,\zeta}$ and $\Delta_y Q_{K^M}$ have a very similar distributive pattern, except for the opposite sign. Both of their horizontal maps, Figs. 4.21c and d, show an active interior and two inert wings, and between these dynamically different regions there wedge many isolated small positive/negative "tear-drops". This pattern is also seen on the map of the interaction analysis $BT = T^{0\to1}_{K^M,\zeta}$, a result of the insignificant $T^{1\to1}_{K^M,\zeta}$ (see Figs. 4.21h, i) and vanished BT_V.[9] The extended "negative similarity" between $\Delta_y Q_{K^M}$ and BT is particularly clear in their zonally-averaged y-profiles (Figs. 4.22c and d). Let $\langle . \rangle^x$ stand for the x average over a wavelength. It is very interesting to see that the positive values of $\langle BT \rangle^x$ and negative values of $\langle \Delta_y Q_{K^M} \rangle^x$ appear only in a well, with two peaks/negative peaks near $y = \pm 50$ km as the "walls". The implication is that unstable events occur only in this energetic well, and the y-advection serves to transport the transferred energy toward the two poles, which, if not removed by the inverse transfer, would be accumulated at the walls. The inverse transfer near the critical bands is another feature of interest. We have conjectured from Kuo's theorem that it's hard for free disturbances to exist in the critical narrow bands if these regions are relatively isolated. The distribution of $\langle BT \rangle^x$ in these peaky regions verifies this conjecture. There energy transfers are not directed from the large-scale window to the meso-scale window, but in the opposite direction (the stable-like direction), although the system as a whole is unstable. This inverse transfer is important, as it is the only feed-back admissible in this model. If computed with a higher resolution, these peaks will not be this sharp. They do have some structure, as shown in Figs. 4.22e-f, the close-ups for $\langle \Delta_y Q_{K^M} \rangle^x$ and $\langle BT \rangle^x$ near the southern critical band. These two figures are obtained with a resolution six times higher than the original one ($\Delta y = 500$ m). From their distribution, we see that the two peaks do not really span only the critical bands ($0.5L$ to $0.53L$ and $-0.53L$ to $0.5L$). For either of $\langle \Delta_y Q_{K^M} \rangle^x$ and $\langle BT \rangle^x$, there is a tail extending beyond the $\pm 0.53L$ limits. This is in agreement with our early conjecture, as what Kuo's theorem states is in the form of an integral with respect to y from $-L$ to L, rather than a local one. The scenario could be made more clear if a dataset is generated with a c_r

[9] For a two-scale decomposition, $T_{K^M,\zeta} = T^{0\to1}_{K^M,\zeta} + T^{1\to1}_{K^M,\zeta} = BT_H + T^{1\to1}_{K^M,\zeta}$, and $BT_H = BT - BT_V = BT$ for 2-D flows.

and $\bar{u}(\pm\frac{L}{2})$ better distinguished.[10]

The general picture for a barotropic instability with Kuo's model is now clear. Sketched in Fig. 4.23 is a cartoon of the processes that constitute such an instability. The two narrow hatched regimes, where $(\bar{u}-c_r)\bar{q}_y < 0$, form the two "walls" of a well which limits the energetic activities. Originally the system is not perturbed, K_n^M (and hence $\left\langle K_n^M \right\rangle^x$) is uniformly zero. When an instability begins, a part of energy is transferred from the large-scale window to feed the growth. The transfer takes place only in the well, and the transferred energy due to this parietal process is transported from the center toward its two wings until it hits the walls where the poleward transportation is almost halted (a very small part of transport still exists and that makes the K_n^M increase, as shown in Fig. 4.22b). An inverse transfer then brings the transported energy back to the large-scale window, effectively barring the perturbation from reaching the southern and northern boundaries. An equilibrium result of this series of energetic processes is a distribution of meso-scale energy as shown in Fig. 4.22a.

It is worthwhile to investigate the fate of that part of energy inversely transferred near the critical strips (the hatched regions in Fig. 4.23). Theoretically this is not a question answerable here as no feedback is permissible to the background field in a linearized model. But one thing is for certain, while the basic jet loses energy at its core, it gains energy at its wings. This implies a broadened jet after a barotropic instability, a result in precise agreement with what has been predicted by Pedlosky (1979. p. 511) with analytical arguments.

4.6 Summary and discussion

We have shown in this chapter how the MS-EVA can be applied to a stability problem. The result of this application has given the MS-EVA a successful validation, and revealed energetic structures for the two fundamental processes, barotropic instability and baroclinic instability.

The MS-EVA validation is made essentially with two combined interaction analytical terms, $BC = (T^{0\to1}_{A^M,\partial_h\rho} + T^{0\to1}_{A^M,\partial_z\rho})$, and $BT = BT_H + BT_V = (T^{0\to1}_{K^M,\zeta} + T^{0\to1}_{K^M,w\partial_z}) + T^{0\to1}_{K^M,\partial_z v}$, which

[10] A bigger $|c_r - \bar{u}(\pm L/2)|$ seems to be hard to achieve, however. In order to generate a dataset such that $|c_i| \ll |c_r|$, the wavenumber should be chosen near the critical $k = \frac{\sqrt{3}}{2}\pi$, which always yields a c_r near $\frac{1}{2}$ and hence two narrowly stripped regimes with negative $\bar{q}(\bar{u} - c_r)$.

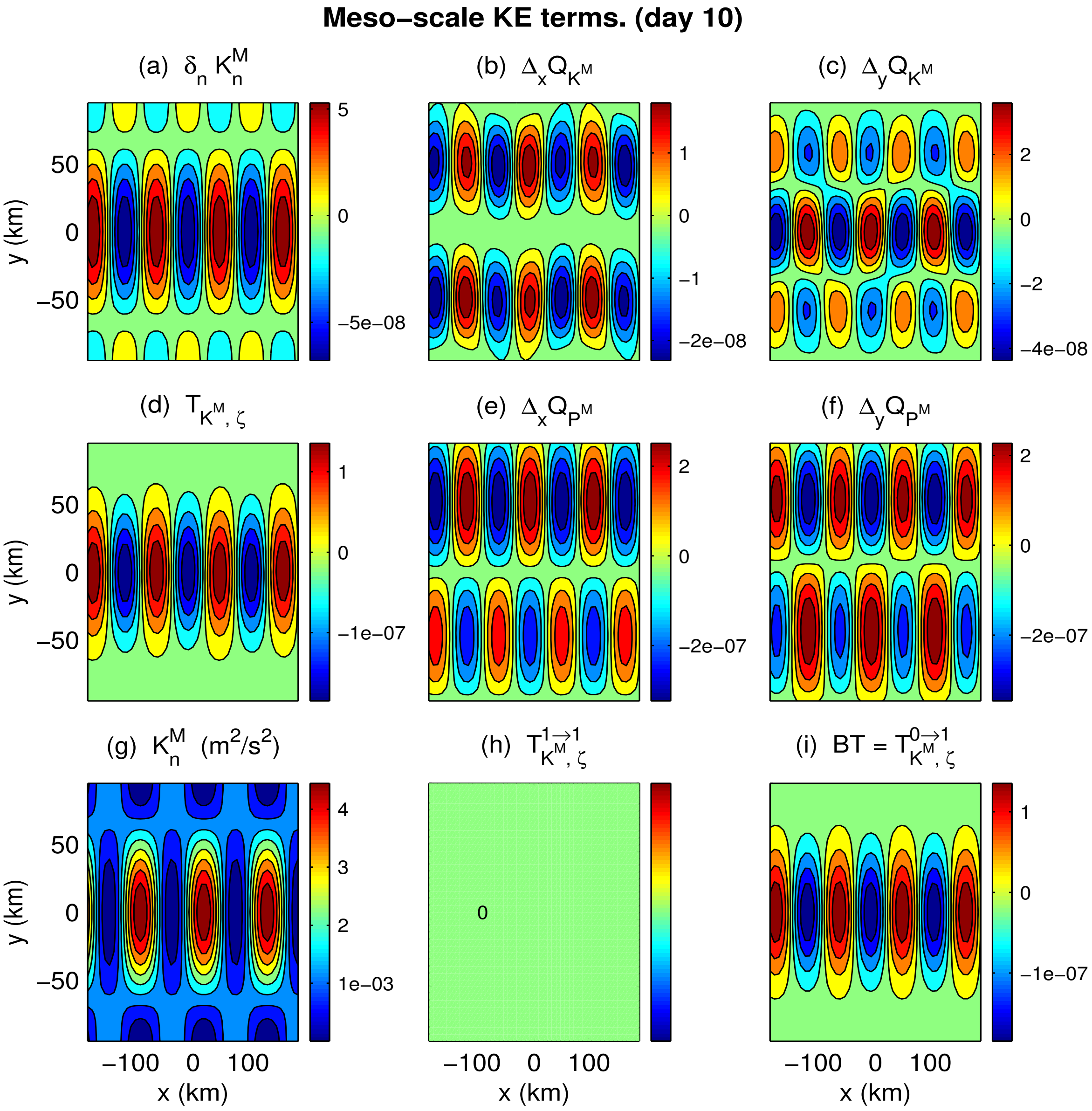

Figure 4.20: The day-10 meso-scale energetics for the Kuo model: the neutrally stable case. (a)-(f) The meso-scale KE terms (in m^2/s^3). (g) The meso-scale KE (in m^2/s^2). (h)-(i) The interaction analysis of $T_{K^M,\zeta}$ (in m^2/s^3). Notice in (h), $T^{1\to 1}_{K^M,\zeta}$ vanishes everywhere in the domain, and in (i), the fact that $BT_V = 0$ has been used.

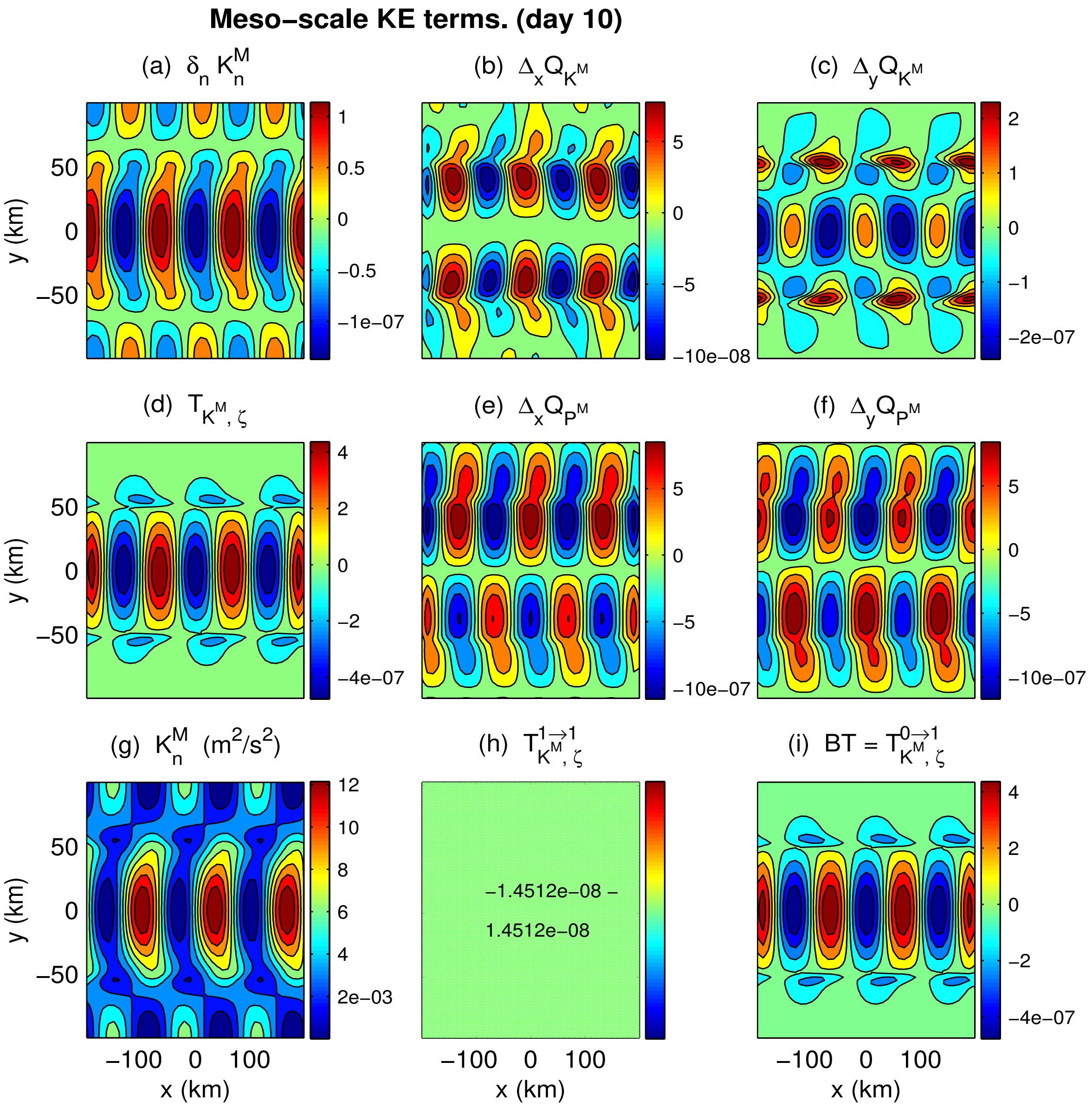

Figure 4.21: The day-10 meso-scale energetics for the Kuo model: the unstable case. (a) - (f) The meso-scale kinetic energetic terms (in $\mathrm{m}^2/\mathrm{s}^3$). (g) The meso-scale kinetic energy ($\mathrm{m}^2/\mathrm{s}^2$). (h) - (i): The interaction analysis of $T_{K^M,\zeta}$ ($\mathrm{m}^2/\mathrm{s}^3$). In (h), $T^{1\to 1}_{K^M,\zeta}$ is not significant compared to $BT_H = BT$ and is hence masked for clarify.

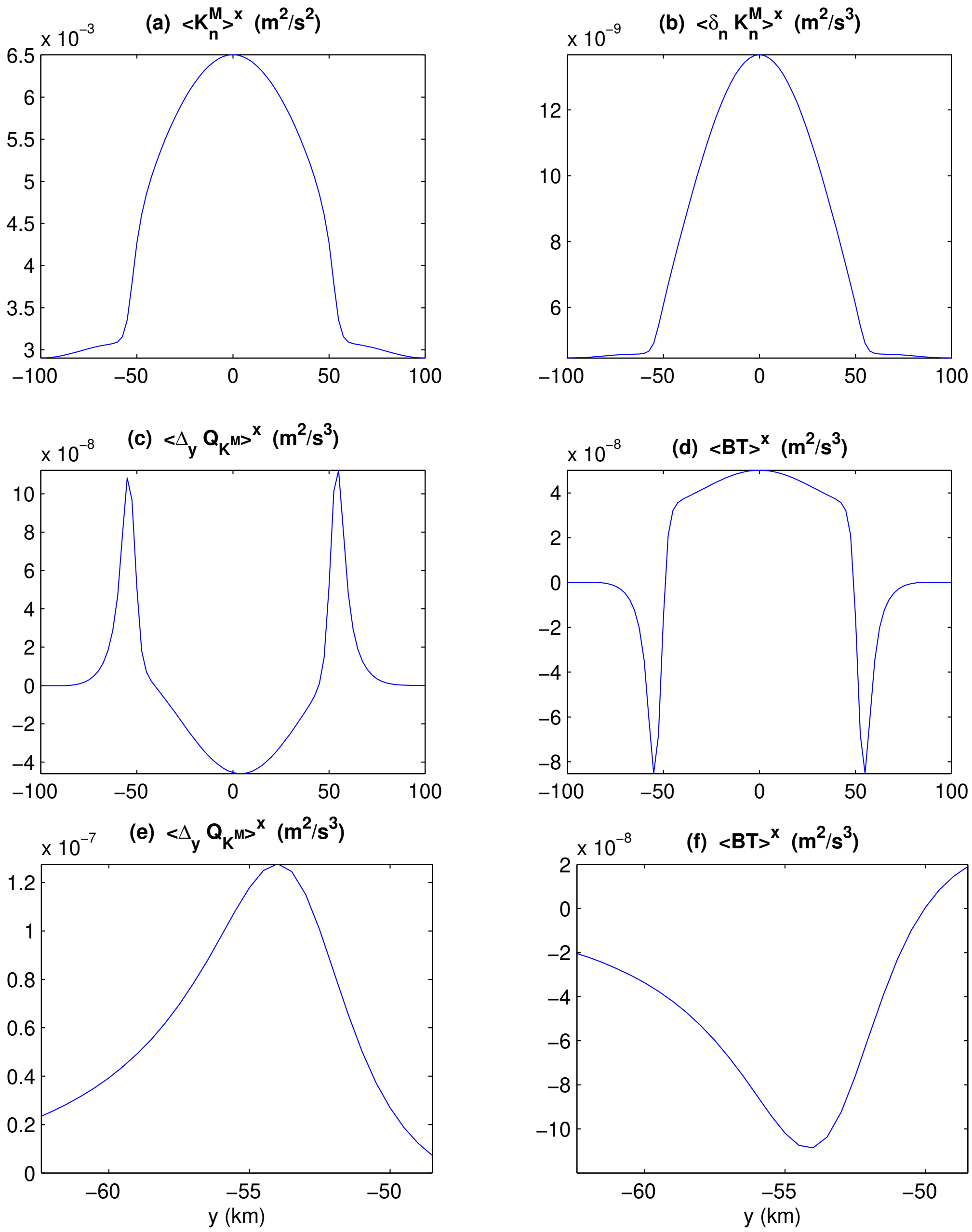

Figure 4.22: The zonal averages (over one wavelength) of a selected group of properties as shown in Fig. 4.21: (a) energy $\langle K_n^M \rangle^x$ (in m^2/s^2), (b)-(d) energetic terms (in m^2/s^3). Subplots (e) and (f) are the close-ups of (c) and (d), respectively.

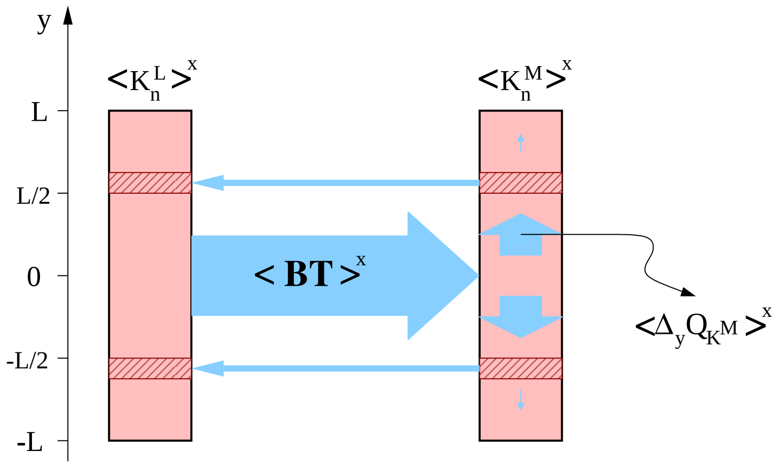

Figure 4.23: A cartoon of the MS-EVA processes that make a barotropic instability. Hatched are the critical regimes where $(\bar{u} - c_r)\bar{q}_y < 0$.

have been shown to be indicative of the two instabilities. With their clear physical interpretation, BC and BT distinguish the MS-EVA from its classical counterpart in that process identification with MS-EVA is made objective. Specifically, when the spatial average of $BC+BT$ is positive, the system is unstable, and for an unstable system, the type of instability is determined by the signs of BC and BT averaged over the domain in study: A positive averaged BC corresponds to a baroclinic instability, while a positive averaged BT implies a barotropic instability. In the cases studied, BT may be simply replaced by BT_H, since BT_V is insignificant in comparison to its peer.

The MS-EVA validation and energetic exploration has been carried out with both the Eady model and the Kuo model. For the former, BT_H averages to zero on all horizontal surfaces (so does BT_V, but this is not shown because of its insignificance). But BC is different. It averages to a positive quantity, just as expected. In fact, for all vertical levels, BC horizontally averages to a positive quantity. The potential energy is continuously transferred from the large-scale reservoir to the meso-scale window, with a (more or less) uniform distribution in the vertical direction. But the meso-scale APE is not the only recipient of the transfer. A significant part of the energy is converted into another form, the meso-scale kinetic energy, upon being transferred. The conversion occurs mainly at the mid-depth, where the converted energy is carried away toward the surface and bottom via pressure work. These are the processes by which an Eady system is energized via a baroclinic instability, a scenario in agreement with that portrayed by Pinardi and Robinson (1986) with their classical QG EVA. The final result of the instability is a dumbbell-shape profile for either the horizontally averaged meso-scale APE or the horizontally averaged meso-scale KE, implying a trapping of disturbances at the two vertical boundaries for both the density and velocity fields, which is indeed the case with the analytical results.

For Kuo's cosine jet model, only barotropic instability exists, so one expects a positive averaged BT (actually BT_H, since $BT_V = 0$ for a 2-D flow) when an unstable mode is concerned. This has been exhibited in the MS-EVA diagnosis. Moreover, if the spatial average is performed in x only, this averaged BT has an interesting meridional profile. This profile is shared simultaneously by the y-transport, $\Delta_y Q_{K^M}$, except for having the opposite sign. The x-averaged BT and $\Delta_y Q_{K^M}$ jointly make the energy circulate for a barotropically unstable mode in a physical-phase space. Specifically, the large-scale KE is first transferred toward the meso-scale window. But this transfer is limited within two critical latitudes, which resemble the two "walls" of an "energetic well" around the jet core. The regions outside the well are not involved. The transferred energy is redistributed by $\Delta_y Q_{K^M}$, which directs it poleward but it also effectively stops near the walls, barely going further. The incessant transfer (provided the instability continues) plus the transport would make the perturbation energy accumulate near the two critical latitudes $y = \pm y_c$, should an inverse transfer not come into play, which brings part of energy back to the basic flow within two narrow bands near those critical latitudes. This inverse transfer is actually implied in Kuo's theorem, and agrees well with what Pedlosky has predicted with analytical arguments about the effect of barotropic instabilities on the basic jet.

The MS-EVA is thereby validated and what it predicts is in agreement with the known analytical results. It must be pointed out, however, the spatial averages we have adopted for the Eady and Kuo models are too special to be of practical use. In real problems, it is usually not feasible to have dynamics in precisely a sinusoidal form, so it is very difficult to cut out a suitable subdomain for averaging. We have to find a general way to make the desired energetic features come out. This is achieved by the multivariate large-scale window synthesis or reconstruction.

The multivariate or multi-dimensional large-scale window synthesis is the multivariate generalization of the large-scale window synthesis introduced in Chapter 2. It may be understood as a low-pass filter (in a multi-dimensional setting) which conserves total energy, or in a rather loose sense, it is just a local average. With such a synthesis, the averaging domain is not an issue any more, as now the average is performed at the location of concern. Clearly, this will work for our purpose so long as the large-scale window (in the spatial direction) for averaging is set larger than the scale of the dynamic features under study. In Chapter 6, a two-dimensional version of this synthesis will be applied to study the Iceland-Faeroe Frontal variability. For an introduction of the two-dimensional multiscale window transform, and an example of the large-scale synthesis, refer to §2.7.

Bibliography

[1] Charney, J. G., 1947: The dynamics of long waves in a baroclinic westerly current. *J. Meteorol.*, 4, 135-163.

[2] Cummins, P. F., and L. A. Mysak, 1988: A quasi-geostrophic circulation model of the northeast Pacific. I. A preliminary numerical experiment. *J. Phys. Oceanogr.* 18, 1261-1280.

[3] Eady, E. T., 1949: Long waves and cyclone waves. *Tellus*, 1: 258-277.

[4] Harrison, D. E., and A. R. Robinson, 1978: Energy analysis of open regions of turbulent flows-mean eddy energetics of a numerical ocean circulation experiment. *Dyn. Atmos. Oceans*, 2: 185-211.

[5] Holton, J R., 1992: *An Introduction to Dynamic Meteorology.* Academic Press. 496pp.

[6] Kuo, H. L., 1949: Dynamic instability of two-dimensional non-divergent flow in a barotropic atmosphere. *J. Meteorl.* 6, 105-122.

[7] Kuo, H. L., 1973: Dynamics of quasigeostrophic flows and instability theory. In: *Advances in Applied Mechanics*, Vol. 13. 247-330 pp. Chia-Shun Yih, ed.. Academic Press.

[8] Pedlosky, J., 1979: *Geophysical Fluid Dynamics.* Springer-Verlag.

[9] Pinardi, N., and A. R. Robinson, 1986: Quasigeostrophic energetics of open ocean regions. *Dyn. Atmos. Oceans* 10, 185-219.

[10] Pinardi, N., and A. R. Robinson, 1987: Dynamics of deep thermocline jets in the POLYMODE region. *J. Phys. Oceanogr.*, 17, 1163-1188.

[11] Press, W. H., S. A. Teukolsky, W. T. Vertterling, and B. P. Flannery, 1992: *Numerical Recipes in C - The Art of Scientific Computing.* 2nd Ed., Cambridge University Press. 965 pp.

[12] Robinson, A. R., 1993: Physical processes, field estimation and interdisciplinary ocean modeling. Reports in Meteorology and Oceanography: Harvard Open Ocean Model Rep. 51. 80 pp., Division of Applied Sciences, Pierce Hall, Harvard University, Cambridge, MA 02138.

[13] Robinson, A. R., 1996: EPS202 notes. Dept. of Earth and Planetary Sciences, Harvard University, Cambridge, MA 02138.

[14] Saltzman, B., 1957: Equations governing the energetics of the larger scales of atmospheric turbulence in the domain of wave number. *J. Meteor.*, 14, 513-523.

[15] Shapiro, J. M., 1994: Adaptative McClellan transformations for quincunx filter banks. *IEEE Trans. on Signal Processing*, vol. 42, No. 3, 642-648.

[16] Spall, M., 1989: Regional primitive equation modeling and analysis of the POLYMODE data set. *Dyn. Atmos. Oceans*, 14, 125-174.

Chapter 5

Real-time forecasting of the Iceland-Faeroe Frontal variability

The Multiscale Energy and Vorticity Analysis (MS-EVA) will now be applied to a real problem: the dynamical study of the Iceland-Faeroe Front (IFF). In this chapter, we make an operational forecast for the IFF variability based on an unprecedented dataset collected on the 1993 R/V *Alliance* cruise (see Robinson et al., 1996; Miller and Cornuelle, 1999). The forecast result will serve as the input for the MS-EVA application in the next chapter.

5.1 Introduction

5.1.1 The Iceland-Faeroe Front and its hydrographic environment

The Iceland-Faeroe Front (IFF) is a narrow band with sharp gradients of temperature and salinity over the Iceland-Faeroe Ridge (IFR), the widest and shallowest gap among the three major channels that connect the North Atlantic and the Arctic. In close vicinity to the IFF are the Greenland Sea, Iceland Basin, and Norwegian Sea (GIN). Fresh and cold Arctic waters and warmer and salty North Atlantic waters from the GIN meet together along the ridge, forming a zone containing several fronts with distinct hydrographic properties on either side. Identifiable on a temperature map like Fig. 5.1 are three such fronts, the IFF being the strongest and most pronounced.

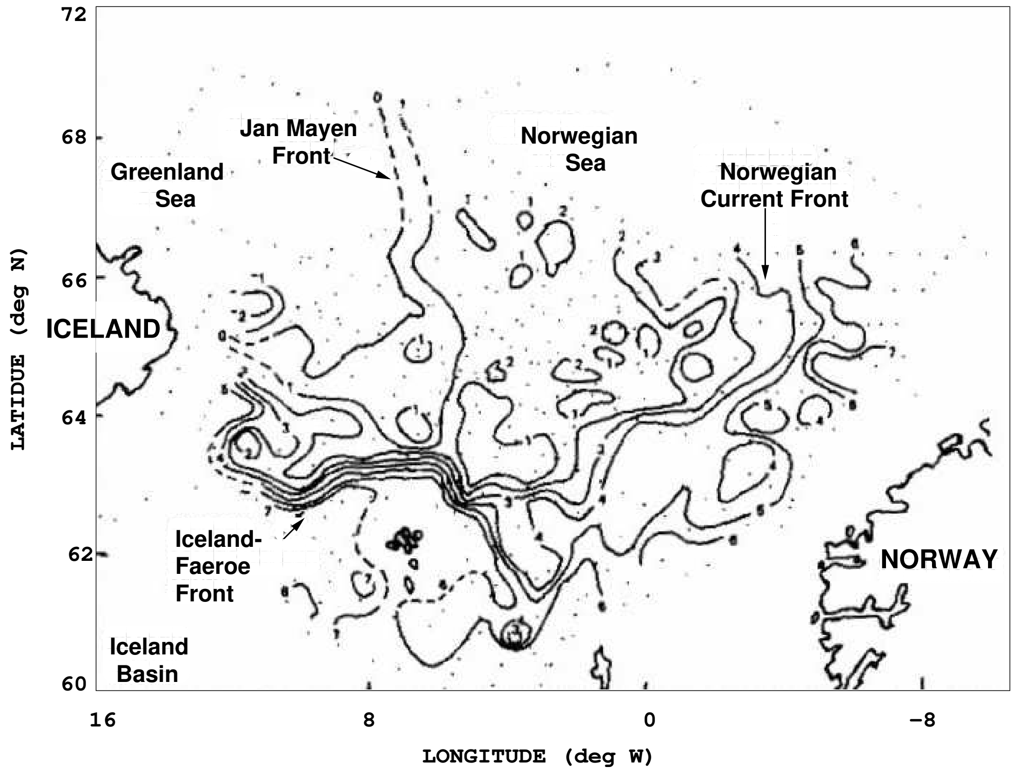

Figure 5.1: Typical temperature distribution (depth 350 m) for the Iceland, Norwegian, and Greenland Seas. Indicated on the map are the three major fronts in this region: the Jan Mayen Front, the Norwegian Current Front, and the Iceland-Faeroe Front. (Adapted from Smart, 1984).

The IFF emerged from among many ocean frontal regions as a focus of attention a long time ago. (For a historical account, see Hopkins, 1988.) Incessant research efforts have been made ever since by the commercial and military, as well as the scientific communities. The intense vertical motion, which is associated with the divergence/convergence along the front, is a driving force for the fishery industry of the surrounding countries; the energetic current system and accordingly the sound speed variation, on the other hand, makes this region an ideal testing ground for the military who rely on sonar to detect sound signals under the sea; in oceanography, this region's richness in relevant scales is a challenge to seek better understanding of the fundamentals of frontal dynamics; and, most importantly, the IFF sets a boundary for the Arctic and the North Atlantic oceans, and hence its variability plays a decisive role in the communication between two of the major oceans of the world. For all these reasons, the IFF has earned its niche in the temple of regional ocean problems. In the following, we give a brief description of the water environment and the frontal variability in this region as known to date.

Around the IFF many water masses with distinct hydrographic properties have been identified, as summarized by Allen et al. (1994), and Read and Pollard (1992). Roughly speaking, waters from the Greenland Sea and waters from the Iceland Basin have T-S values that form the two extremes. The former are generally cold and fresh, while the latter are much warmer and saltier. Those from the Norwegian Sea lie in between. A more careful study reveals that there actually exist seven such water masses. By the naming convention of Allen et al. (1994), they are, respectively, the North Atlantic Water (NAW), the Modified North Atlantic Water (MNAW), the Norwegian North Atlantic Water (NNAW), the Norwegian Sea Deep Water (NSDW), the Arctic Intermediate Water (AIW), the East Icelandic Water (EIW), and the East Icelandic Current Water (EICW). Among these waters, the EIW and EICW are two fresh varieties from the west. Another variety of EIW, the Modified EIW (MEIW), has also been identified (Read and Pollard, 1992) but is not included here. The NNAW and EIW appear in the literature by many different names (Muller et al., 1979; Becker and Hansen, 1988; Blindheim, 1990; Meincke, 1978). We adopt Allen et al.'s convention because it seems that they have given the most relevant and complete account so far.

Closely related to the distribution and formation of the water masses in the region is the frontal variability. As is well known, the IFF varies on a variety of distinct scales. On average, it is believed to lie approximately parallel to the IFR (Smart, 1984). Either the thermally defined gradient (2-7^{o}C) or the 35.00 psu isohaline (which has been used as an alternative IFF definition) has been observed to follow the bathymetry well on a large scale, though the definition by salinity might have less seasonal variation. (See, for example, Hansen and Meincke, 1979.) On a smaller scale, however, the IFF is quite energetic and intermittent. Excursions from its mean state of up to 200 km in several months are not uncommon (Smart, 1984). In comparison to the long-term average, the meso-scale variability, which is characterized by the internal Rossby radius of deformation (Robinson, 1983), is usually of more interest, because of its role in the heat/energy and mass transport between the Arctic and the Atlantic (Allen et al, 1994; Hallock, 1985; Willebrand and Meincke, 1980). With the synoptic or quasi-synoptic surveys conducted since 1960, the IFF variability has gradually come clear. Coherent structures have been identified and temporal and spatial scales estimated. For example, Hansen and Meincke's (1979) dataset revealed a number of eddy-like features and meanders; Meincke and Ross observed a cold anomaly and a warm meander on either side of the front (see Willebrand and Meincke, 1980); Gould et al. (1987) located 7 cold centers in their study; Scott and Lane (1990) found ten cold eddies south of the front; and so forth. These meso-scale events are quite episodic and intermittent. In the frequency spectrum of energy, Willebrand and Meincke (1980) found that there usually exist two peaks, one at scale 2-5 days and another at scale 8-11 days. This was partially verified by Allen et al. (1994), who observed a temporal scale of 2-4 days for the variability. Spatially, Niiler et al. (1992) obtained a scale of 10-17 km (corresponding to an eddy size 30-50 km) from a combination of airborne (AVHRR and AXBT) and in-water observations (CTD) together; similar scale is also seen in Scott and Lane (1990). A slightly larger estimate is made by Allen et al (1996), 15-17 km. Both 10-17 km and 15-17 km are in good agreement to the internal radius of deformation, which is approximately 10 km for this region (Robinson et al., 1996).

The short time and space scales of the frontal variability raise an issue for the IFF study. Because of the stringent synoptic constraint, it is quite challenging to obtain a hydrographic dataset without missing significant energetic events. This challenge had seriously limited the understanding of the dynamic processes for this region, which are highly energetic and intermittent, until the advent of a new approach, real-time operational forecasting.

5.1.2 A brief history of the forecasting for this region

Real-time operational forecasting was initially introduced by Robinson (1992). Based on the Harvard Ocean Prediction System (HOPS), it aims at predicting those fast changing and highly energetic events, either for oceanographic research or for the rapid assessment of regions for naval operations. Since its introduction, operational forecasting has been launched in many regions of the world ocean, and has become an important methodology in oceanographic studies, many of which would be impossible with traditional approaches.

For the IFF region, a quasi-geostrophic (QG) forecast (Miller et al., 1995a) and a primitive equation (PE) forecast (Robinson et al., 1996) had been made before the present study was undertaken. The QG forecast was launched with the data acquired during the hydrographic survey conducted in October 1992 by SACLANTCEN (Supreme Allied Commander, Atlantic

Undersea Research Center) and Harvard University, while the PE forecast was set up during the 1993 R/V *Alliance* cruise, which was also conducted by Harvard in cooperation with SACLANTCEN. The PE forecast was formulated as a real-time shipboard problem, as introduced by Robinson (1992) for naval operations. The unprecedented IFF'93 dataset provides this forecast with a good initialization and an objective validation. The basic features observed during the survey have been captured, particularly in the run with all available observational data assimilated. As explained in the paper of Robinson et al. (1996), the results are satisfactory, either by visual inspection or by quantitative validation (in terms of the anomaly correlation coefficient and the root-mean-square error).

Besides these two previous forecasts, we noticed, while the present study was in progress, that Miller and Cornuelle (1999) had completed another one, also with the 1993 R/V *Alliance* dataset. An inverse scheme was exploited to initialize their model runs, which helps overcome the synopticity problem with the observations. Their result is dynamically consistent and qualitatively and quantitatively successful. But they also found that there is no clear superiority of the inversely initialized scheme over the less sophisticated initialization method used by Robinson et al. (1996), in terms the quantitative measures of forecast skill defined above.

Notice that for the MS-EVA application we need a PE model dataset of the IFF variability which faithfully reproduces the dynamics of the meso-scale window and is robustly tested against parameters. The available PE forecasts, either by Robinson et al. or by Miller and Cornuelle, are both performed using a model with a horizontal grid spacing $\Delta x = \Delta y = 5$ km and only 5 levels in the vertical direction. Considering the 10-km radius of deformation, reproduction of the meso-scale processes might be affected by the coarse grid resolution. Moreover, the HOPS code has been upgraded since Robinson et al.'s study, the current version having many new capabilities, and it is anticipated that a better forecast could result by using the new version, including an improved PE model. In a word, there is still a room for improvement in the 1993 IFF forecasting, and that motivates the study in this chapter.

5.1.3 An outline of this chapter

We therefore need a physically consistent forecast of our own for the MS-EVA application. As no superiority has been found to adopting a sophisticated initialization, we keep using the

Robinson et al. (1996) forecast strategy, and as many of their parameters as possible, to set up the PE model for this region, employing the current version of HOPS with modifications as appropriate. Note that in Robinson et al.'s forecast, the old PE model is used. Adopting their methodology thus allows one to see the HOPS improvements since our forecast will be made comparable to their benchmark study. For the same reason, this chapter is also arranged similarly to their paper, albeit everything is started from scratch. Immediately following this section is a preparation of the original data for the PE model, which is set up in §5.3. The forecast results are described thereafter, followed by a section of sensitivity study which is lacked in their study. After that we give an evaluation of our results with the forecast skill quantitatively defined before. This chapter closes with a summary and a brief discussion of some issues raised during the forecasting.

5.2 Data pre-processing

5.2.1 The dataset

The IFF93 dataset was acquired from the R/V *Alliance* cruise in a rectangular domain centered at 64.25°N, 10.75°W (Fig. 5.2), with a zonal extent of 140 km and a meridional extent of 190 km. The cruise began on August 14, and lasted through August 23. During the 10-day period three surveys were conducted. They are, as named for convenience by Robinson et al. (1996), the *initialization survey* (August 14-16), the *zigzag survey* (August 18-20), and the *validation survey* (August 20-23). The initialization survey is intended for the forecast initialization, while the validation survey is for verification. The zigzag survey, which earns its name from its zigzag cruise track, aims at gleaning more close-up information about the meandering stream to update the model in the vicinity of the frontal axis. The data acquired during the first and third surveys include CTDs, XCTDs, and XBTs, while along the zigzag track, only XBTs are available (see Table 5.1). These hydrocasts were supplemented with the deployment of two sets of surface drifters and current meter moorings. The current data retrieved from these sets have been exploited for absolute velocity inference by Robinson et al. (1996). But in the present study, we will use them only for reference whenever necessary.

The locations of the data points are shown on the sampling maps in Fig. 5.3. Two almost

Table 5.1: Number of the hydrocasts during the IFF93 surveys.

	CTD	XCTD	XBT
Initialization survey	16	19	125
Zigzag Survey	0	0	116
Validation Survey	37	9	125

identical sampling patterns are seen for the initialization and validation surveys, with stations roughly 7 km apart in the north-south direction and 24 km in the west-east direction. Drawn beneath these maps are the corresponding cruise tracks. Initially the ship entered the domain from the eastern boundary. It took about three days to complete the data acquisition for the first leg of the cruise. As we will see later, this data nonsynchronization raises an issue for the nowcast, and something in compensation must be done with the unsampled western region for the initial days. During August 20-23, the ship followed the same cruise track as the initialization survey, in the hope of gathering a dataset comparable with the first one. The zigzag survey was conducted during August 18-20. It does not have a regular track and the northern region has an apparent lack of sampling points. The scarcity of data there proves to be a problem when attempting to extract the domain-scale features for that duration, but for the stream along the frontal axis, the data coverage should be dense enough to reflect the corresponding processes.

5.2.2 Salting and smoothing

From Table 5.1 we see that a large number of hydrocasts are XBTs. No salinity data were collected for the corresponding stations. To make the dataset forecast-ready, these XBT casts have to be *salted*, i.e., to be augmented by consistent salinity information. Salting is justified by the observation that temperature and salinity are more often than not statistically highly correlated. T-S diagrams, for example, often allow one property to be derived from the other. In the present problem, salting is achieved through T-S diagram inference (typical of the IFF region) guided by comparisons with the surrounding CTD/XCTD profiles and the climatological data for the ambient stations. The typical diagram used for the inference is

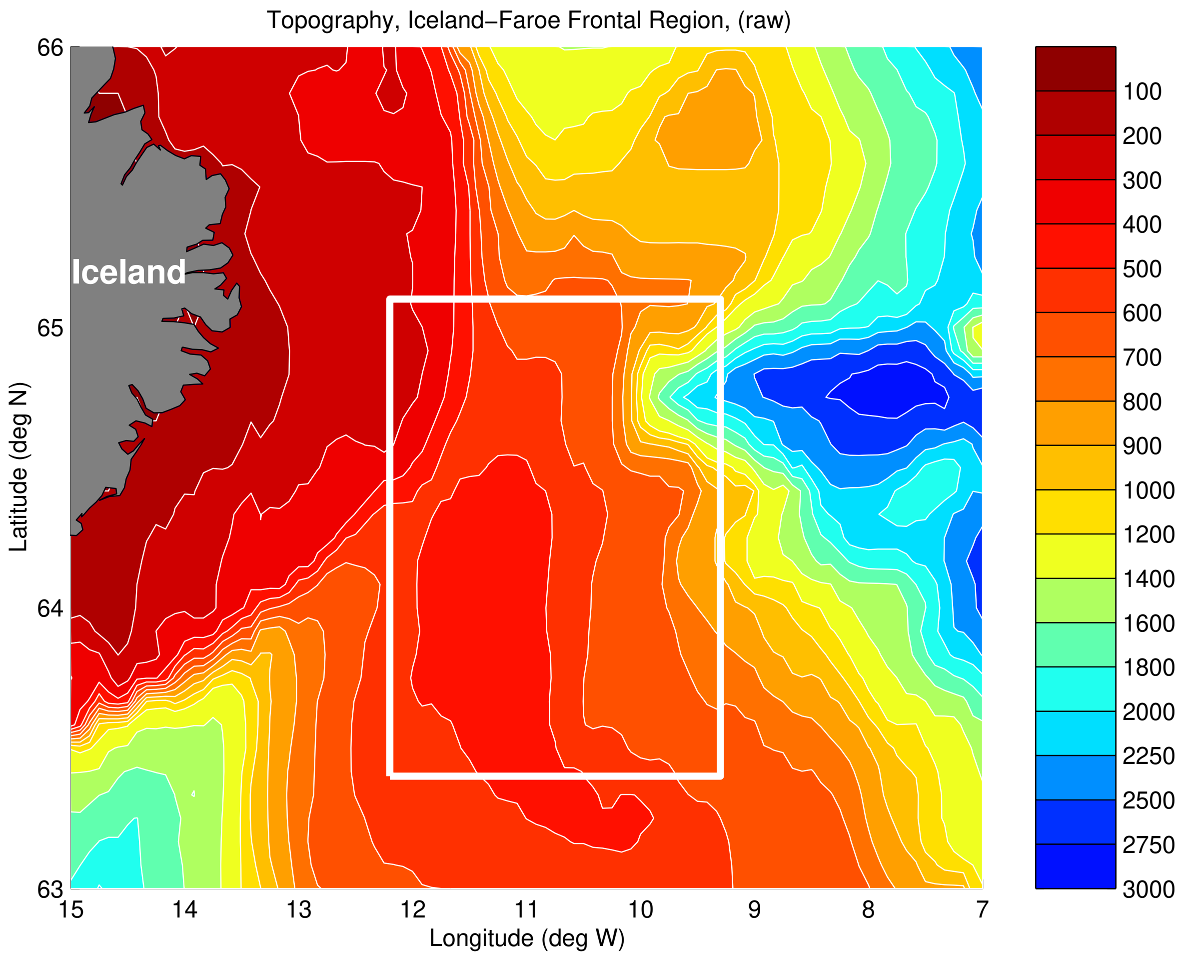

Figure 5.2: The IFF93 experimental domain and its underlying bottom topography (depth in meters).

shown in Fig. 5.4, which is simplified from Read and Pollard (1992). We use the historical result instead of the current dataset for this purpose in order to make the forecast a real operational one. (We don't have the IFF93 data in advance.) As a verification, Fig. 5.5 depicts the same diagram with the IFF93 CTD/XCTD data (from all the three surveys). Clearly, most of the seven water masses are present (except the East Iceland Current Water and the Modified North Atlantic Water), though their varieties may not coincide with Fig. 5.4 exactly. The salting procedure is simply an exercise to derive from Fig. 5.4 the S (salinity) given a T (temperature) as an input. As is seen on the diagram, S is a multi-valued function of T. When salting, we first pick for an XBT datum an S_m on the main branch (the thick curve), then compute for this point and this level the linear interpolation S_i of the available

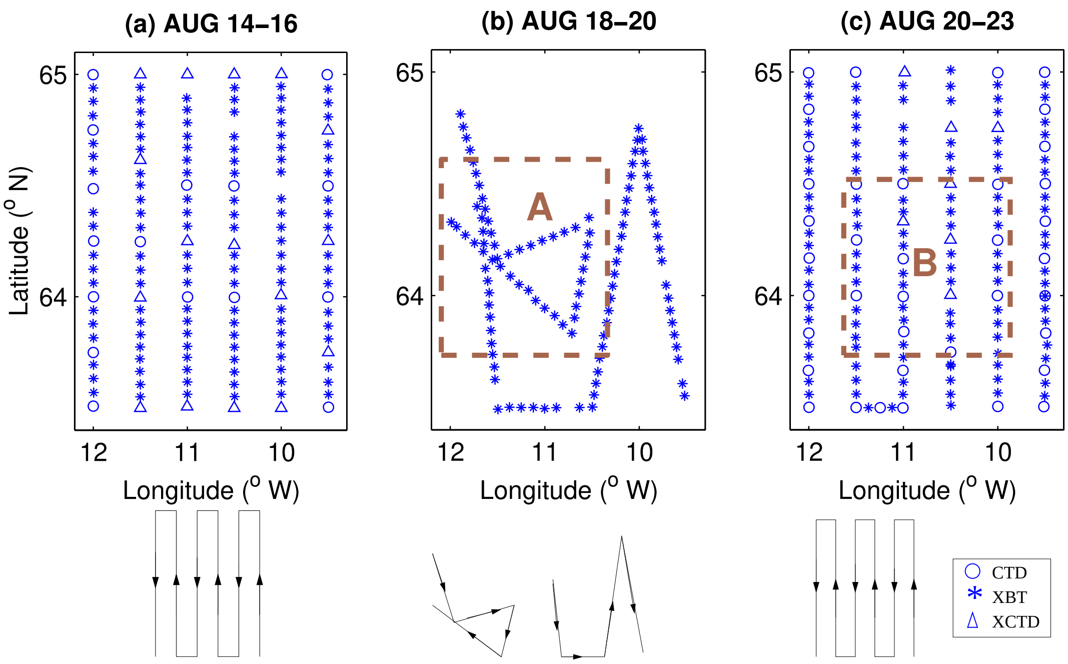

Figure 5.3: Hydrocast locations for (a) the *initialization survey*, (b) the *zigzag survey*, and (c) the *validation survey*. Drawn beneath the figures are the corresponding cruise tracks. The subdomains A and B will be used for validation. (Adapted from Robinson et al., 1996)

salinity values at the same level from the surrounding stations. If S_m differs from S_i by a significant amount (define this amount to be 0.1 psu), try all the branches of Fig. 5.4 until an S is found which is closest to the interpolated value S_i. This S is then taken to be the desired salinity corresponding to the given T.

The salted dataset allows one to build two temperature/salinity profiles, which are typical of the hydrographic properties south and north of the front. These profiles are to be used later in a feature model construction to initialize the forecast. Fig. 5.6 shows such profiles for the western region of the domain which are obtained in a manner similar to that used by Robinson et al. (1996). It is of interest to notice that, while the temperatures are distributed in a similar trend, the salinities have completely different vertical structures. More interesting is that fact that, these temperature/salinity profiles, distinct as they may be, result in two almost identical density structures for the regions north and south of the front (Fig. 5.8). The lack of density gradient across the front implies to some extent the importance of barotropicity, as will be clear in the next chapter.

In distinction to the typical smooth profiles, the observed temperature and salinity sig-

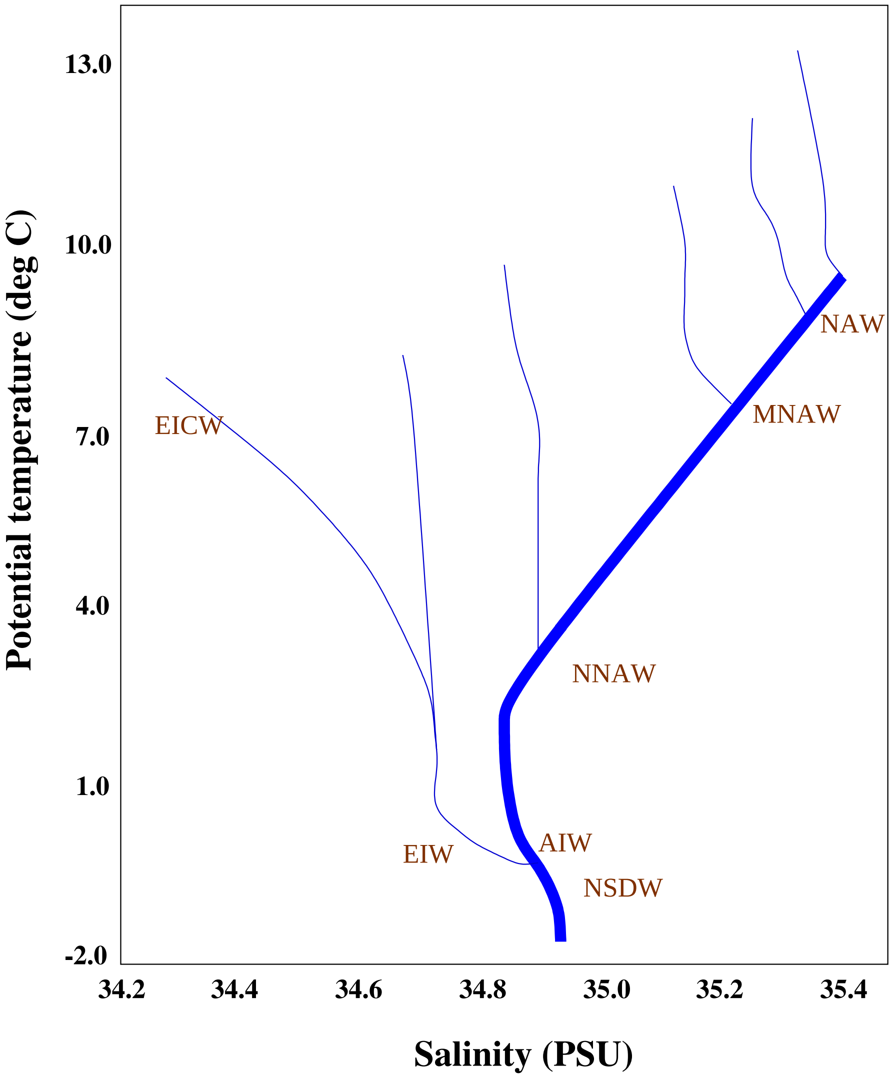

Figure 5.4: A simplified T-S diagram used for the salting of the IFF93 XBT data (adapted from Read and Pollard, 1992). Indicated in the figure are the water masses identified in this region, among which NAW stands for the North Atlantic Water, MNAW for the Modified North Atlantic Water, NNAW for the Norwegian North Atlantic Water, AIW for the Arctic Intermediate Water, NSDW for the Norwegian Sea Deep Water, EICW for the East Iceland Current Water, and EIW for the East Icelandic Water.

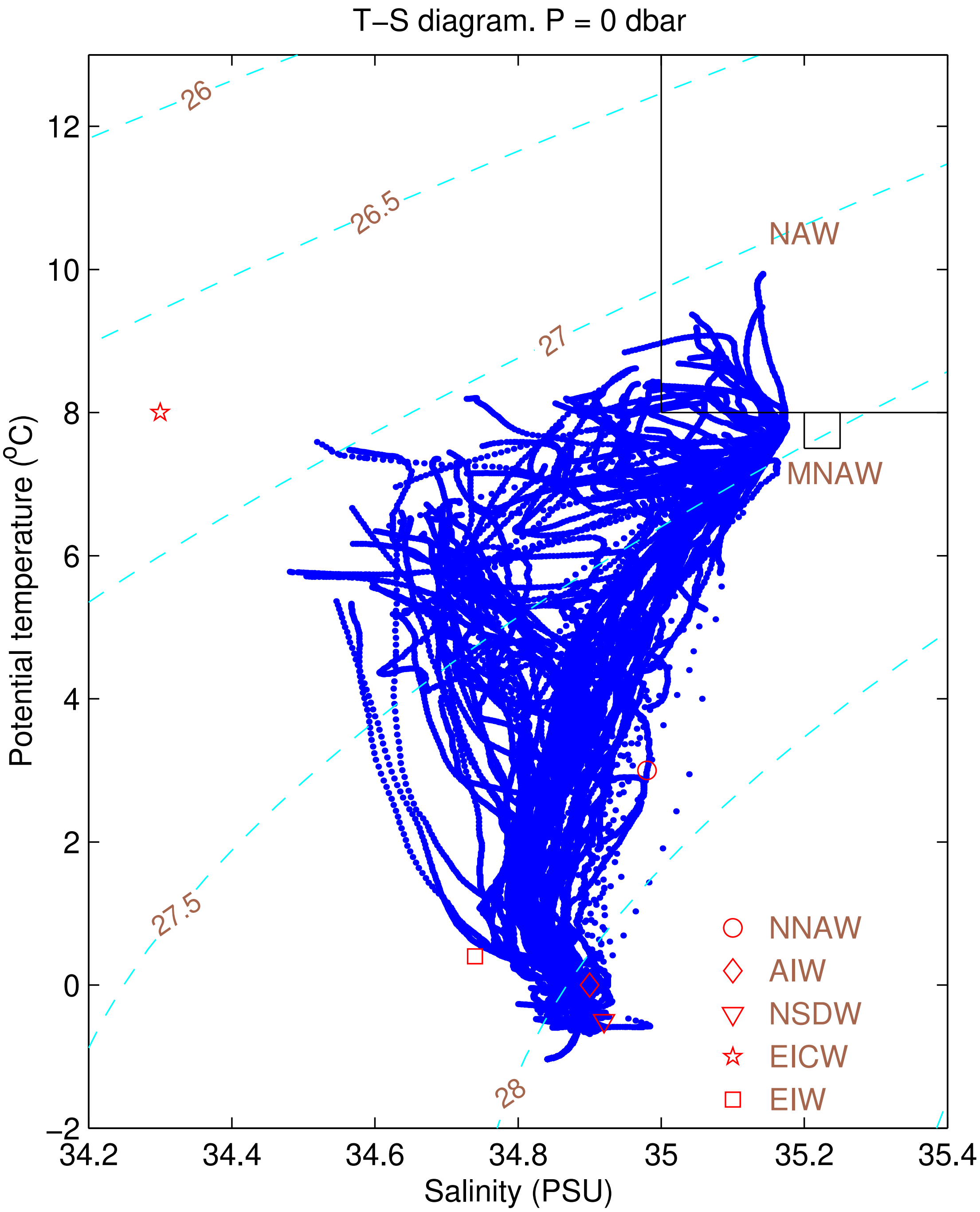

Figure 5.5: T-S diagram based on the CTD and XCTD data acquired during the IFF93 surveys. The water masses are indicated in the boxes. Refer to Fig. 5.4 for an explanation of these shorthands.

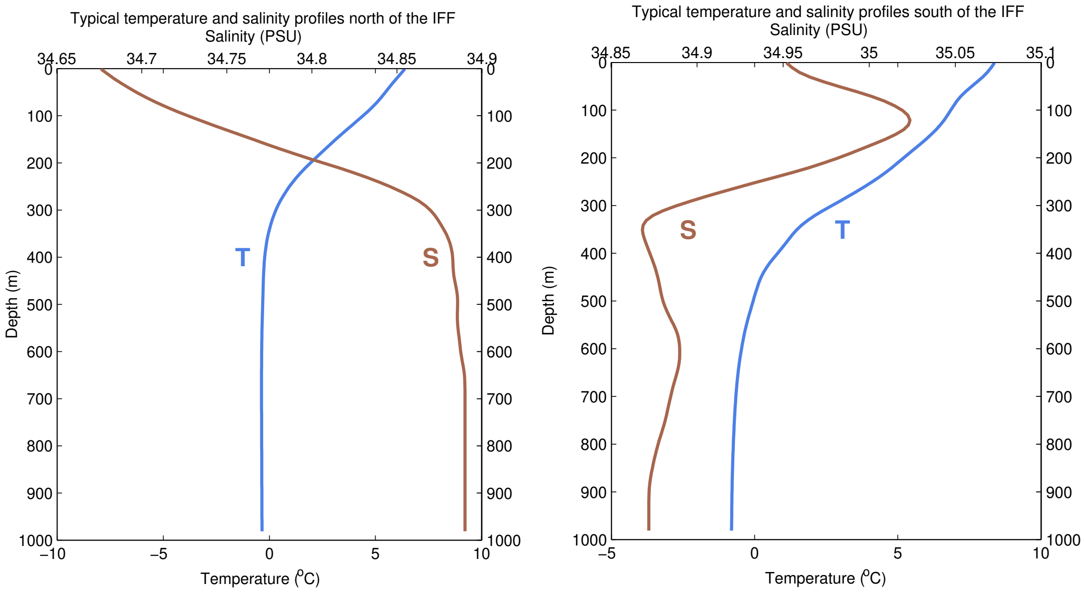

Figure 5.6: Typical hydrographic profiles north (left) and south (right) of the IFF. T indicates temperature whereas S is salinity.

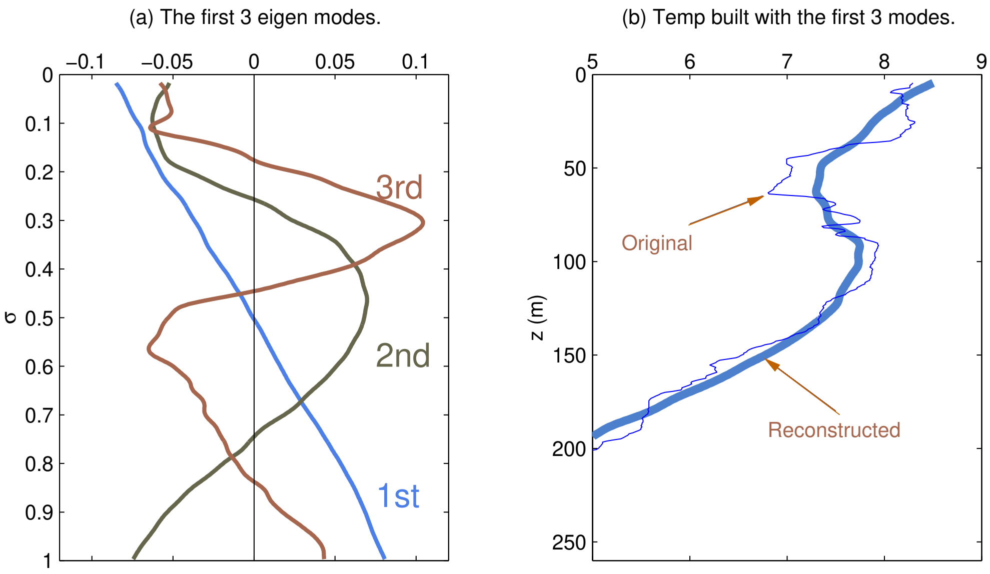

Figure 5.7: (a) The first three EOF functions (the mean removed in advance) versus the scaled vertical coordinate σ, and (b) a typical upper-layer temperature profile (in ^{o}C) and its reconstruction (the thick curve) with these modes. (The mean has been added back to it).

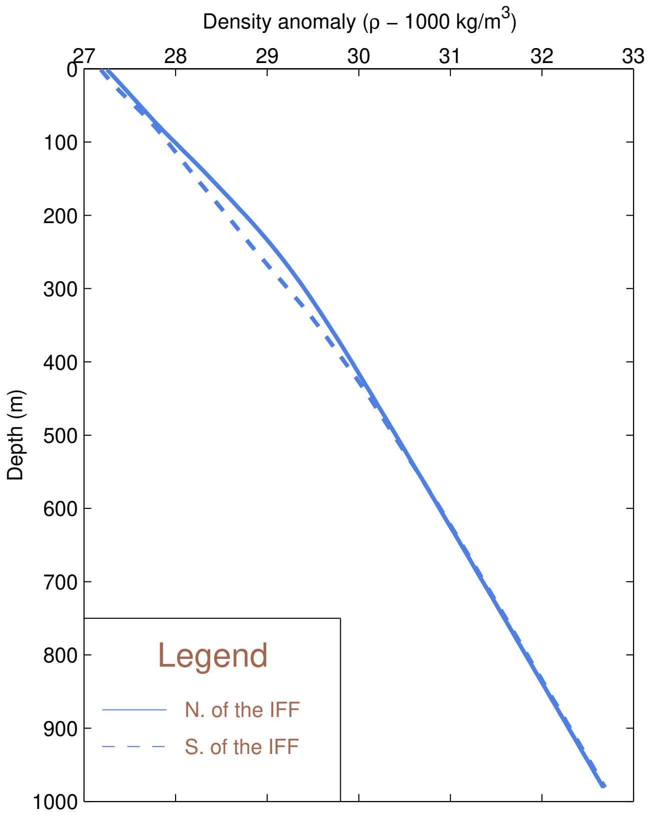

Figure 5.8: Typical density profiles north (solid) and south (dashed) of the IFF.

nals are rich in fine vertical structure which cannot be resolved by the model. These small scale ripples must be eliminated before the data can be adapted into the nowcast. To do this, we fit each signal with the first three empirical orthogonal function (EOF) modes (cf. Appendix A5.1). These EOFs are calculated each time a new set of assimilation fields are requested. (Since the prediction is in a real-time operational mode, we are supposed to use only those hydrographic data available up to the day of forecast.) Different as they may be, it is observed that, for all of the vertically demeaned profiles, the first three modes account for at least 95% of the total variance. Fig. 5.7a is just such an example. It shows these structures derived from the whole initialization data set. Shown together with it is a typical temperature profile and its reconstruction with these modes (Fig. 5.7b).

5.2.3 OA maps

The salted and smoothed hydrographic data are now mapped onto the mesh grid points to prepare for the nowcast and data assimilation. (See §5.3 for a description of the model grid.) The mapping is achieved with objective analysis, or OA for short. OA is a widely used statistical tool based firmly on a few assumptions which have been justified by Gandin (1965) for the atmosphere and oceans (cf. Appendix A5.2). In this application, we use the "Two-stage approach" proposed by Lozano et al. (1996) to corroborate these assumptions. This approach was adopted in Sloan (1996) and further delineated by Lermusiaux (1997). In our problem, it can be re-stated as the following procedures:

(1) Guess a mean of the field in concern. This may be done by fitting the observation with a lower-order polynomial (e.g., a bilinear function $f(x, y) = ax + by + cxy$), or simply by taking the arithmetic mean.

(2) Estimate the covariance function and the scale of correlation. (see Appendix A5.3).

(3) Apply OA to calculate the basin-scale features.

(4) Take the result of (3) as the mean field and repeat (2) and (3).

In Lozano et al. (1996), Steps (1), (2), and (3) are viewed as one stage, and Step (4) as another stage. That is how the "Two-stage approach" earns its name.

It must be pointed out that the so-called "basin scale" is different from that elaborated in the other chapters of this thesis. Here no energy issue is involved, and, as a matter of fact, energy will not even be conserved with the "scale" defined hitherto in this chapter. Since our purpose here is not energy study, but only to map the observational data onto some prepared mesh grid, a practical way like the two-stage OA is just fine.

As shown in Appendix A5.2, the covariance matrix $\underline{\underline{\mathbf{C}}}$ and the error E are two major parameters to be tuned in the objective analysis. $\underline{\underline{\mathbf{C}}}$ reveals the scale of correlation, and E, in analogy to what we know from the multi-resolution analysis, actually determines the detail space— with the exception here that it is not perpendicular to the approximation space. In the current version of HOPS, the covariance is chosen to be[1]

$$Corr(\underline{\mathbf{x}}_i, \underline{\mathbf{x}}_j) = (1 - r_1^2)\exp\left(-\frac{1}{2r_2^2}\right) \tag{5.1}$$

with

$$\begin{aligned}
r_1^2 &= (\underline{\mathbf{x}}_i - \underline{\mathbf{x}}_j)^T \underline{\underline{\mathbf{L}}}_1^{-2} (\underline{\mathbf{x}}_i - \underline{\mathbf{x}}_j) + \left(\frac{\Delta t}{\tau}\right)^2 \\
\underline{\underline{\mathbf{L}}}_1 &= r_0 \underline{\underline{\mathbf{I}}} \\
r_2^2 &= (\underline{\mathbf{x}}_i - \underline{\mathbf{x}}_j)^T \underline{\underline{\mathbf{L}}}_2^{-2} (\underline{\mathbf{x}}_i - \underline{\mathbf{x}}_j) + \left(\frac{\Delta t}{\tau}\right)^2 \\
\underline{\underline{\mathbf{L}}}_2 &= r_d \underline{\underline{\mathbf{I}}}
\end{aligned}$$

where r_0 and r_d are respectively the zero-crossing and the e-folding decay scales, $\underline{\mathbf{x}}_i$ is the model grid point position, and and $\underline{\underline{\mathbf{I}}}$ the identity matrix. This prescription essentially reduces the OA to a problem of how to manipulate the four parameters:

- Zero-crossing r_0,
- Spatial decay scale r_d,
- Temporal decorrelation scale τ,
- Error variance E.

If possible, some relation connecting r_0 and r_d will further reduce the complexity of the problem of parameter manipulation. In practice, a linear relation $r_0 = a \cdot r_d$ with $a \leq 2$ turns out to be a good choice.[2] With these assumptions the parameters are then tuned until an optimal estimation is achieved. An optimal estimation is determined according to the following two criteria:

(1) The correlation coefficient approaches zero at some large distance;

(2) The ratio σ_f^2/σ^2 is consistent with E (ideally they should be the same). Here $\sigma^2 = \overline{(\psi_{OA} - \bar{\psi}_{OA})^2}$ and $\sigma_f^2 = \overline{(\psi - \psi_{OA})^2}$ with the overbar signifying the ensemble mean and ψ_{OA} being the OAed field.

[1] This definition is the same as the one in Sloan (1996), p.41, except that he defines $Corr(\underline{\mathbf{x}}_i, \underline{\mathbf{X}}_j)$ ($\underline{\mathbf{X}}_j$ the data point position) while here we define $Corr(\underline{\mathbf{x}}_i, \underline{\mathbf{x}}_j)$. We invoke only $\underline{\mathbf{x}}_j$ to make the form look simple. But these two are essentially identical.

[2] Carlos Lozano, personal communication.

As there could be scale change in the vertical direction, different r_0 and r_d are generally needed for different levels. But in this study the same parameters are adopted for all fields throughout the water column. The tuning begins with a guess for r_0 made from the correlation analysis (Appendix A5.3). After many experiments (could be tedious and time-consuming), the parameteric combination $r_0 = 60$ km, $r_d = 40$ km, $\tau = 4$ days, and $E = 0.1$ appears to yield a satisfactory estimation.

Some of the OA results are plotted in Fig. 5.9. From left to right are, in their own order, temperature maps for the initialization survey (centered on August 15), zigzag survey (August 19), and validation survey (August 22). From top down, distributions for the 25-m, 80-m, and 125-m levels, and σ-level 7 (cf. Fig. 5.12) are displayed. These maps are generally the same as those of Robinson et al. (1996) except the 80-m-level one, which is not presented in their paper. We keep it here for later use.

We now give a brief summary of the features shown with these maps, though many of them have already been elaborated upon by Robinson et al. (1996). From Fig. 5.9 the three surveys have captured snapshots of how a deep-sock meander was formed in the middle of the domain during the experiment. Initially the fresh cold waters are well separated from the warm salty Atlantic waters. The front is in an intact form, albeit with some wiggles riding on it. Generally, it is strong in the western region, and gets weaker and weaker toward the east. (Similar observations are seen in Smart, 1984.) A cold vortex is found on the northern flank, attached to another warm eddy from further north. This vortex pair is apparent until the 80-meter level. It is not significant in deeper layers, however.

When time reaches August 19, the frontal axis suddenly straightens toward the southeast. Accompanying this event is a strong northwesterly flow, which has also been observed with the drifter trajectories. (Not shown here. See Robinson et al., 1996.) Although in Fig. 5.9, temperature distributions for this period are shown only for those regions with tolerable errors (25% of the maximal variance), this front straightening is easily seen.

Following the re-orientation of the front comes a sudden meandering during August 20-23. By its shape this cold-tongue intrusion has been described as "deep-sock" (Robinson et al., 1996) and "hammer-head" (Miller et al., 1995a; Miller and Cornuelle, 1999). From the satellite infrared image (AVHRR) (Fig. 5.10) of the sea surface temperature (SST), this phenomenon is

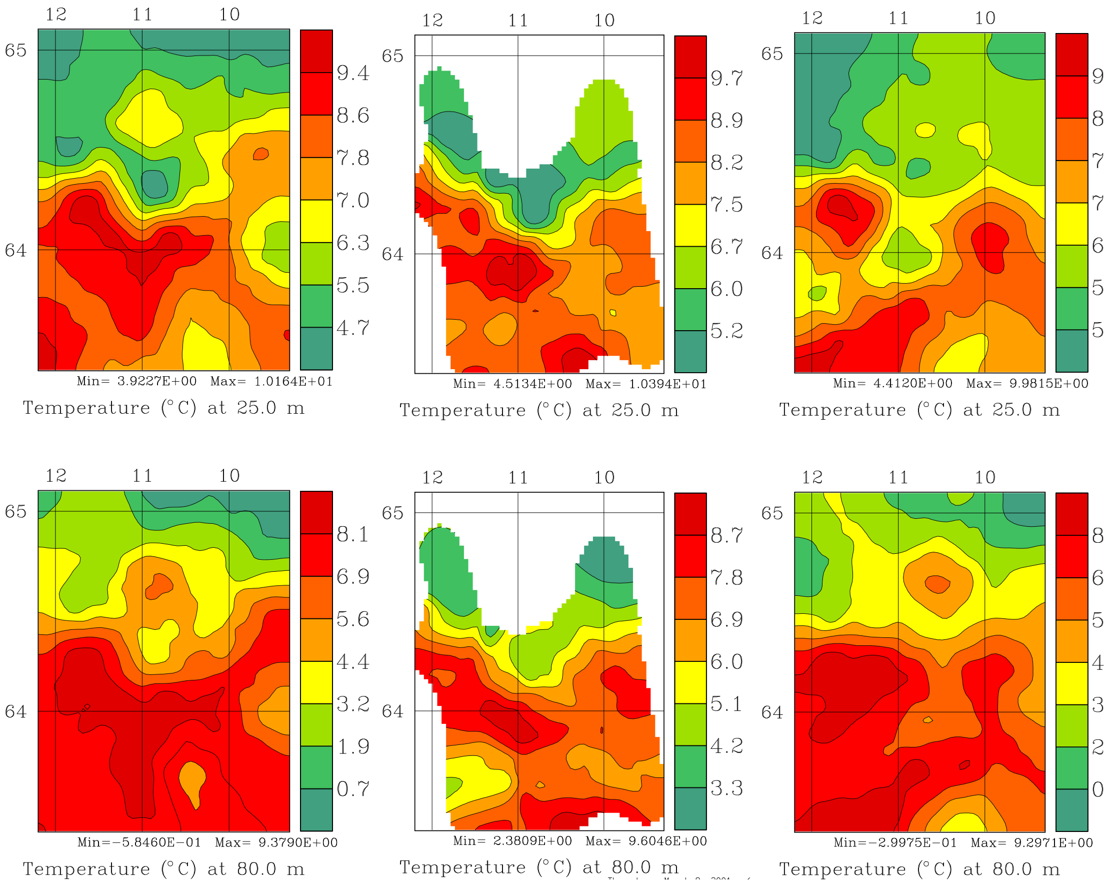

Figure 5.9: OA maps of the observed temperature for August 15 (left panel), August 19 (middle panel), and August 22 (right panel), 1993, which represent the initialization, zigzag, and validation surveys, respectively. The axes are in longitudes (degrees west) and latitudes (degrees north).

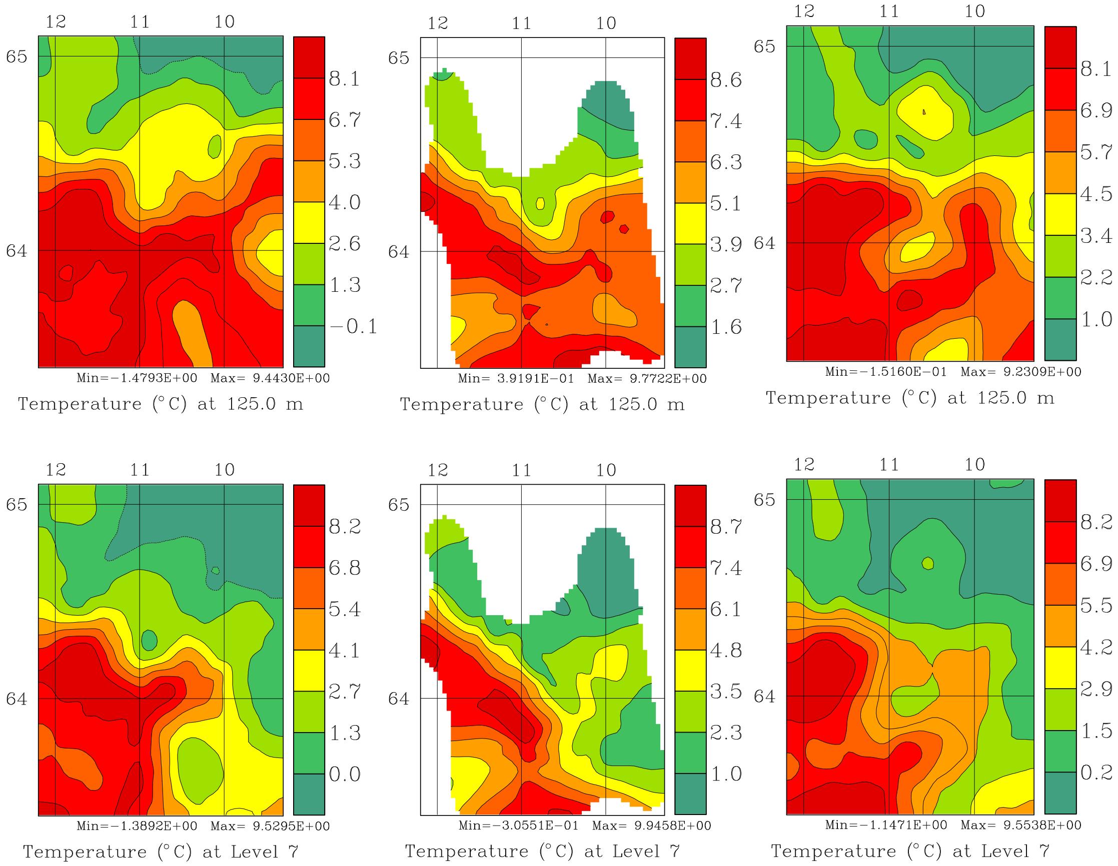

Figure 5.9: (Continued)

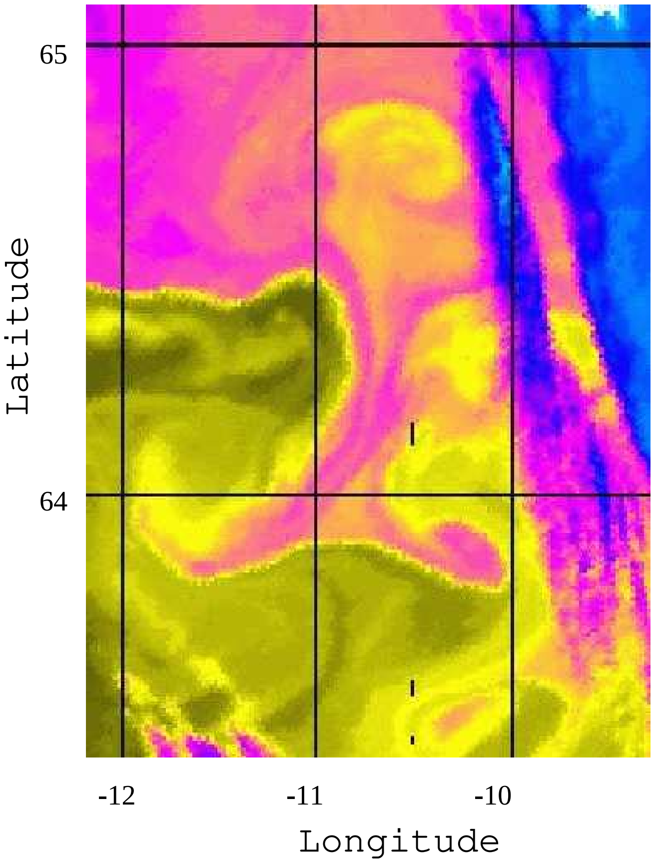

Figure 5.10: A satellite infrared image (AVHRR) for the sea surface temperature of the model domain on August 22, 1993. Temperature is higher in the yellow area than in the pink area. The blue indicates clouds. (Re-drawn from Robinson et al., 1996)

also apparent, with a meander similar to that of Fig. 5.9. In Fig. 5.10, the yellow color indicates warm waters while the pink is cold. The blue area is covered with cloud. The general pattern of the intrusion on this image is actually like a T with upper side down, oriented toward the southeast, with some warm water sheetificated into the north. Superimposed on this pattern we also see features of smaller scales, thanks to the high resolution. These "details" reveal to us an SST map which is rich in self-similarities, implying the possible role of strong nonlinearity in the frontal meandering process (cf. §1.3).

The vertical distributions also show some interesting phenomena. In Fig. 5.11, a meridional section across the center of the deep-sock meander displays how the two waters, one from the north (right) and another from the south (left), interact with each other and eventually lead to an intrusion. The front is initially seen as a interface with a sharp temperature gradient in

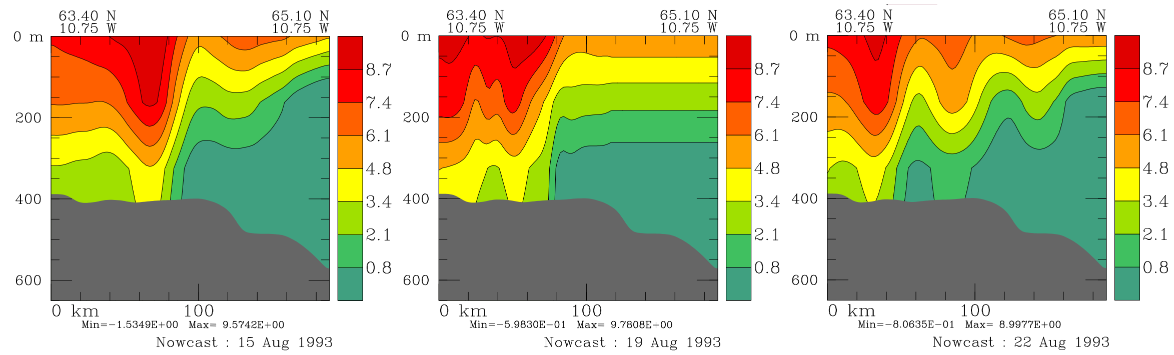

Figure 5.11: Temperature distribution on a meridional section across the deep-sock meander (10.75°W). The unit is °C.

the middle, then on August 19, it appears to be almost perpendicular to the bottom from the north, while in the south, a lot of sub-mesoscale features emerge, reminiscent of the interference with waves reflected from the front. On the same day, the northern part seems to be quiescent. But this might not be true since observational coverage is scarce there. On August 22, the oscillation in the south subsides, but the Arctic waters are advancing southward (the cold intrusion), followed behind by a wake of meso-scale eddy-like fluctuations. Based on what is described above, it is reasonable to postulate that the eddy wake is a result of the north-south negotiation, reflecting the response to the onset of the frontal meandering.

5.3 Model set-up

The IFF93 experiment spans a 140 km $\times$ 190 km rectangular domain which is centered at 64.25°N, 10.75°W (cf. Fig. 5.2). In the forecast model, this rectangle is horizontally discretized into 56 $\times$ 76 mesh grid boxes (57 $\times$ 77 grid points), with an equi-distant spacing $\Delta x = \Delta y = 2.5$ km. In the vertical direction, a hybrid coordinate system is adopted. Above a constant interface depth H_c=150m, the domain is discretized into $k_c = 5$ horizontal levels. Below this depth are 4 terrain-following levels (σ levels). HOPS adopts the Arakawa B-grid (cf., Arakawa and Lamb, 1977), with the tracer and velocity placed on different grid points (refer to Fig. A3.7). In Fig. 5.12, we plot the configuration of the tracer boxes (T-boxes) for two typical vertical sections. The thicknesses of these T-boxes are, from top to bottom, 15 m, 20 m, 25 m, 40 m, 50 m, 78 m, 156 m, 224 m, 312 m. The top 5 horizontal levels are thence

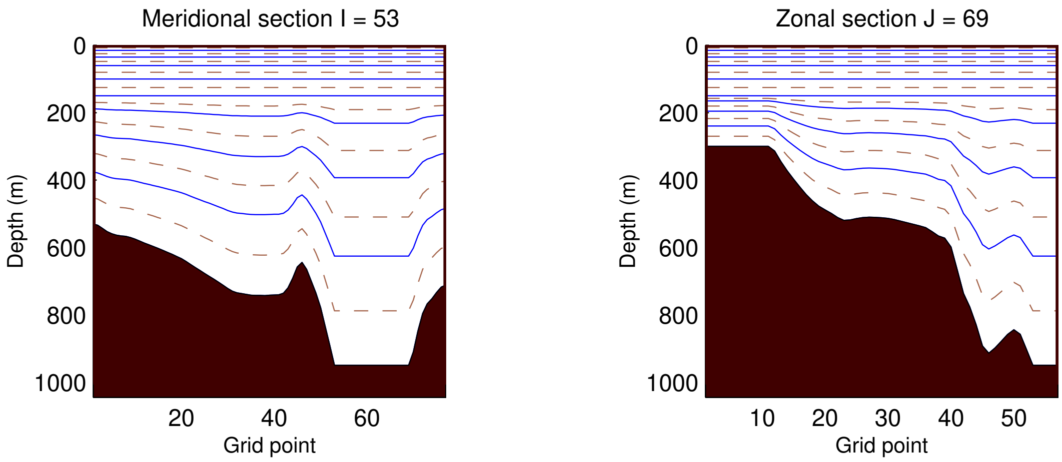

Figure 5.12: σ-levels (dashed lines) for a typical meridional section (left) and a zonal section (right). The solid lines delimit each vertical layer.

at, respectively, 7.5 m, 25 m, 47.5 m, 80 m, and 125 m. Below them are 4 sigma levels.

In order to meet the consistency requirement for the hybrid coordinate system (see Appendix A5.5), the bottom topography is clipped to the range [-950 m, -300 m] (Fig. 5.13a). The hydrostatic consistency factor (HCF) for this conditioned topography and the coordinate system is contoured in Fig. 5.13b. The maximal HCF is $0.548 < 2$, lying within the consistent limits (Lozano, 1995).

The model prognostic fields are now initialized. Since the forecast is an operational one, we pretend to know nothing about the future in setting up the model. Recall that on August 14, 1993 (denoted as day 0 for convenience), the hydrocasts cover only the eastern part of the domain. The western region lacks data, and hence artificial "hydrocasts" are needed to fill the gap. In Robinson et al. (1996), a simple feature model based on the SST image is used to fulfill this goal. We will adopt their approach here. (Other approaches are also available, e.g., Miller and Cornuelle, 1999.) In the feature model, the front sits on an average axis, artificial stations placed on either side in the western region with their corresponding typical T-S profiles obtained before (Fig. 5.6). The station placement is such that the front has a width of roughly 15 km.

The feature model makes it easy for the field initialization. For the temperature and salinity, what is needed is to map the available data, both from the real stations and from the

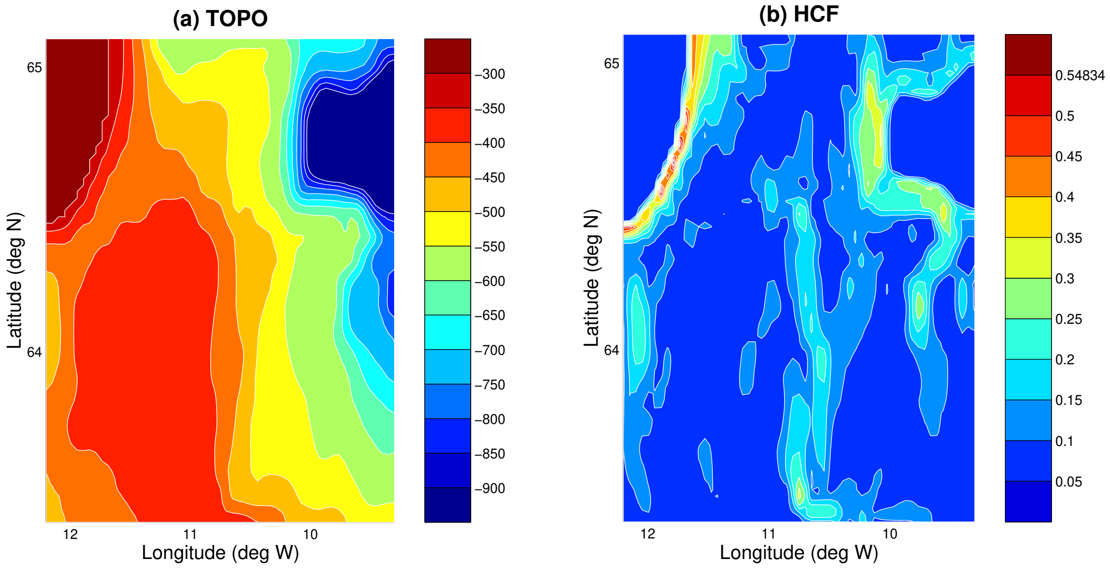

Figure 5.13: (a) Conditioned bottom topography used in the model (min = −950m, max = −300m, CI = 50m). (b): Hydrostatic consistency factor (min = 0, max = 0.548, CI = 0.05).

artificially sampled points, onto the model grid with the objective analysis. The parameters and application procedures are the same as those used in the OA described in the last section. For the velocity field, the baroclinic mode is first calculated, with a reference depth 600 m. The barotropic mode is inferred from the quasigeostropic flow field. (Invert for the barotropic vorticity to obtain the transport. See Lozano, 1995.) In Robinson et al. (1996), the surface drifter information helps determine the external mode. We choose not to do so since it might be inappropriate to have the limited current information extended throughout the whole model domain.

The next step for the model set-up is the specification of boundary conditions. In our problem, there are four horizontal boundaries. We specify implicit Orlanski radiation conditions (Orlanski, 1976; Chapman, 1985; Lozano et al., 1996) on all of them for all the prognostic fields: tracer, velocity, transport streamfunction, as well as vorticity. In the vertical direction, there are also two boundaries where conditions are needed. On the surface, a rigid-lid is imposed (this being the only option for the current version of HOPS), and no external forcing is applied. At the bottom, no momentum, heat or salt flux is allowed through the solid sea floor. These horizontal and vertical conditions, if possible, are updated along with the interior data assimilation, as the model is advanced forward.

In order to stabilize the numerical scheme, spatial and temporal relaxations are applied to each grid point adjacent to the boundaries. That is to say, the boundary conditions are not felt by the interior points immediately. They are relaxed to the interior values within a spatial distance and a designated period. This kind of nudging greatly reduces the possible mismatch in the conditions that are specified and then a longer model integration is ensured. Appendix A5.6 provides some details on this issue. The relaxation parameters used in the model are listed in Table 5.2.

Another issue that the model must resolve is the subgrid process parameterization. In the vertical direction, the effect of these processes are modeled with a Laplacian type diffusion term, while horizontally a Shapiro filter (Shapiro, 1971) is exploited for the parameterization. The equivalence between the Shapiro filtering and Laplacian diffusion can be demonstrated for a single wave with wavenumber κ. In this case, a Shapiro filter with parametric triplet (p, q, r) (refer to Table 5.2 for an explanation) is equivalent to an effective eddy diffusion with diffusivity (Lermusiaux, 1997, p. 136)

$$K = \left[1 - (1 - s^p)^{q/2r}\right] \frac{K_0}{4s}, \tag{5.2}$$

where $K_0 = (\Delta x)^2/\Delta t$, $s = \sin^2(\kappa dx/2)$. In the present problem, the typical spatial scale is about 10-25 km (cf., Hansen and Meincke, 1979; Robinson et al., 1996), the effective diffusivity for the parametric triplet adopted (cf. Tab. 5.2) is estimated to be of order 160-900 m^2/s, which is reasonable for real physics.

To close this section, we give a summary of the adopted model parameters and options, and tabulate them in Table 5.2. The model is now set up and ready for the forecasting.

5.4 The forecasting

5.4.1 General strategy

The strategy of Robinson et al. (1996) is chosen to achieve our forecasting goal. We believe it is the right approach to find a satisfactory forecast, though some modifications may be needed. Generally speaking, it is a kind of field updating which corrects sequentially the model prediction. As the model advances forward, the observations are taken in whenever

Table 5.2: A summary of the PE model parameters.

Parameters	Value
Model configuration and mesh grids	
Grid	$57 \times 77 \times 9$
Mesh	140 km $\times$ 190 km
Time step for	
velocity	180 s
tracer	180 s
transport streamfunction	180 s
Horizontal grid spacing	
Δx	2.5 km
Δy	2.5 km
Vertical grid spacing	
Horizontal level 1	15 m
Horizontal level 2	20 m
Horizontal level 3	25 m
Horizontal level 4	40 m
Horizontal level 5	50 m
Sigma level 6	78 m
Sigma level 7	156 m
Sigma level 8	224 m
Sigma level 9	312 m
Shapiro filters (p, q, r) *(Order, Freq. of appl., No. of time steps between appl.)*	
Velocity	(2, 1, 1)
Tracer	(2, 1, 1)
Vorticity	(2, 2, 1)
Transport streamfunction	No application
Boundary relaxation	
Spatial e-folding distance d_s for	
Velocity	0.7 grid point
Tracer	0.8 grid point
Temporal e-folding distance τ_s for	
Velocity	24000 seconds
Tracer	12000 seconds
Friction parameters	
Vertical eddy viscosity ν	5 cm^2/s
Vertical eddy diffusivity A_v	5 cm^2/s
Bottom drag coefficient	2.5×10^{-3}

available, for both the interior and the boundaries. The whole procedure is therefore totally data-driven. In our problem, recall initially only the eastern region is surveyed. The data-lacking western region is sampled with artificial data from a feature model. As the data coverage builds up, the feature model gradually retreats from its role, until the domain is fully updated. The data for the updating are prepared with objective analysis (OA). When acquired, they are mapped onto the model grid, where corresponding errors are also given. The gridded data are then melded with the forecast using a pointwise interpolation scheme called *optimal interpolation* or simply OI (see Appendix A5.4). OI is actually a linear weighted averaging scheme, with the weights determined from the forecast error field and the OA map error field. This procedure repeats for the first two days (August 15-August 16) until all the initialization data are assimilated. The whole system is then completely initialized and ready for the next-step forecast.

At this stage, the simulation scheme bifurcates on the flow chart of forecasting strategy (cf. 5.14). It goes into two directions, depending on whether the zigzag data are assimilated or not. The branch without further data taken in gives an *F2 forecast*, while the other leads to an *F5 forecast*. The F2 and F5 forecasts were so designated by Robinson et al. (1996) to simplify the otherwise cumbersome terminology. They also labeled the day 14 August 1993 to be forecast day 0, 15 August 1993 to be forecast day 1, and so forth. With the aid of these labels an F2 is simply a forecast based on the observations up to day 2, and an F5 the forecast based on the observations up to day 5 (August 19). Both F2 and F5 are advanced forward at least to day 8 (August 22), the day when a validation is supposed to be conducted.

The general idea of the above methodology is pictorially presented in the flow chart of Fig. 5.14. For clarity, the F2 and F5 schemes are grouped and distinguished by two individual boxes.

5.4.2 F2 forecast

The F2 forecast is a HOPS application based solely upon the data acquired from the initialization survey. As shown in the flow chart of Fig. 5.14, before the model is let go on its own from day 2 (August 16), nowcasts are launched and the model and boundary conditions are sequentially updated. The updating is with a pointwise optimal interpolation, in which the er-

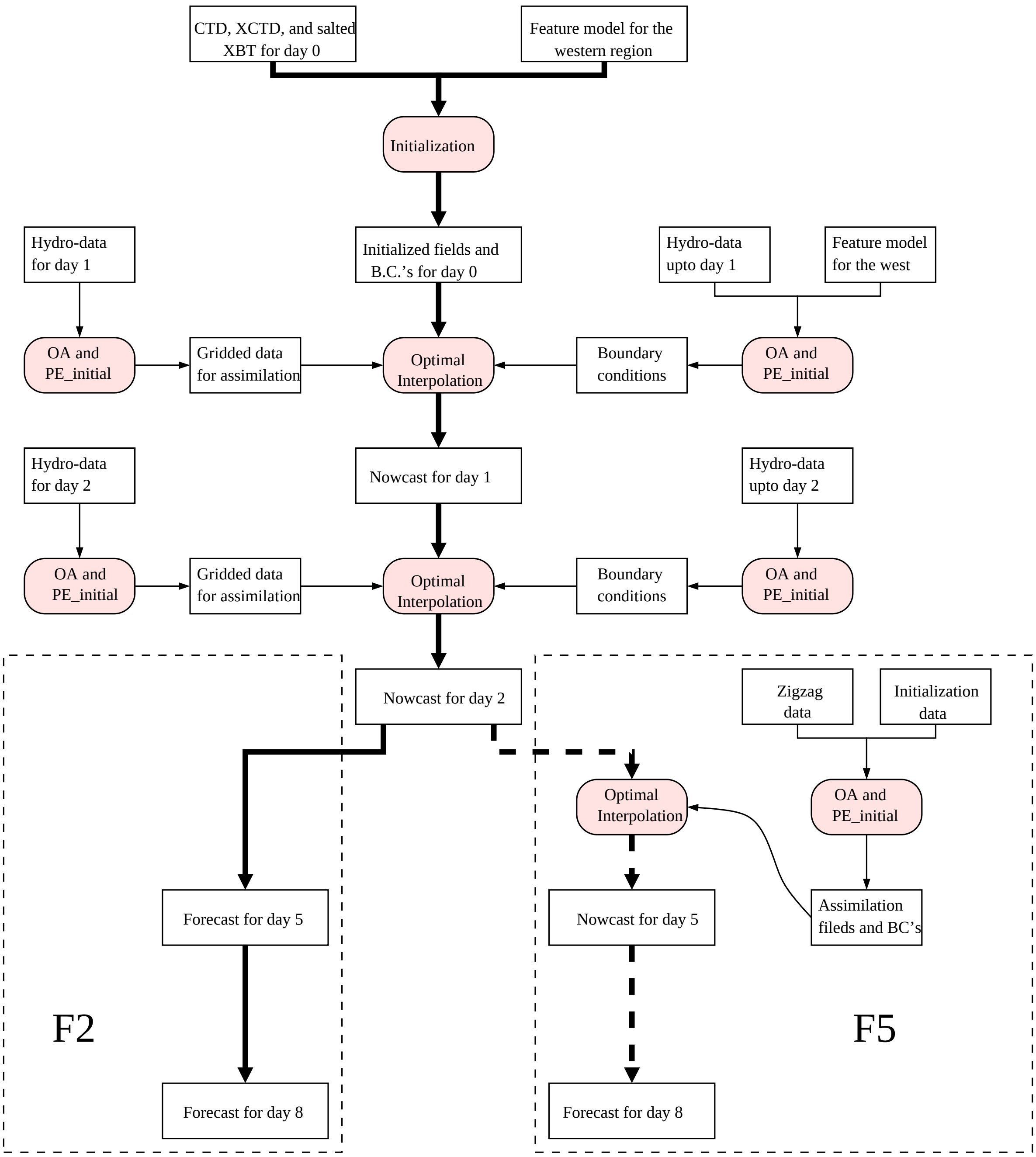

Figure 5.14: Flow chart of the forecasting strategy. The shaded round-corner rectangles represent the HOPS modular applications at that step, while the other boxes show the input/output or forecast status. For an explanation of the terminology, refer to the text of §5.4.1.

ror (variance) field plays a crucial role. The observation error is obtained simultaneously with the OA map (cf. Fig. 5.21), while the forecast error is estimated using some simple model. The exponential decay equation for the prediction certainty in Appendix A5.4 is a choice. But here we use an even simpler one, which appears to be as good as those seemingly more sophisticated schemes, to a large extent. In this model, the forecast error is assumed to be uniform throughout the domain, and is simply taken to be the basin mean of the observational error field. This bold simplification is out of the too much loss of predictability and based on the observation that many other sophisticated but expensive schemes have no convincing superiority by any chance.

Figures 5.15-5.16 show how the forecast initialization is completed with the sequential data updating. These temperature pictures agree well with those of Robinson et al. (1996), and besides, Fig. 5.16 is very similar to the OA map in Fig. 5.9, except that the latter is richer in small-scale features.

Beginning day 2 (August 16), the model is driven entirely by its own physics. The outputs after that form the F2 forecast sequence; shown in Fig. 5.17 is the day-5 temperature in the sequence. As we have described before, a remarkable feature on this day is the re-orientation of the western frontal axis from what it used to be on day 2 toward the southeast. Our forecast has captured this event. It is seen on the maps for all the depths through the water column, and is especially evident on mid-levels such as levels 4 (80 m) and 5 (125 m). A little problem here is that the slope of the front is still too mild and the southeastward current is not strong enough (cross-front temperature gradient not large enough). The maximal speed along the straightened frontal axis is 69 cm/s (figures not shown), about 10 cm/s less than the 79 cm/s inferred from the surface drifters (Miller et al., 1995a,b).

The deep-sock or hammer-head meander of day 8 (August 22) has also been well reproduced. In fact, this is the most successfully simulated phenomenon for the F2 forecast. Comparing the temperature distribution of Fig. 5.9 to that of Fig. 5.18, the forecast meander is correlated well to the observed one, either in its spatial location or its geometric shape. Particularly appealing is the way the meander is oriented. The cold water intrudes westward, in good agreement with both the satellite image and the hydrographic observation. Besides, some indication of scale-similarity, which is lacked in the OA maps (Fig. 5.9) because of the sampling resolution but is present in the AVHRR image (Fig. 5.10), is also significant here. Some other not-that-significant features in the OA maps, such as the cooler tongue intrusion from the northwestern corner, a cold eddy immediately above the southern boundary, etc., are also seen in the forecast result.

If inspected carefully, the time sequence of temperature reveals that a subtle event takes place after the front's southeastward re-orientation. This event, which is characterized by a sudden cresting of the front on the western side of the "valley" (Fig. 5.19), seems to be crucial in bringing about the deep-sock meandering. Most of the failed experiments don't have this phenomenon. In those runs, the front intrudes at the deep valley, making the western flank not able to crest up (northward) back to a state similar to day 2. It is reasonable to conjecture that this cresting may be a pre-condition for the formation of the deep-sock intrusion.

To a satisfactory extent, F2 forecast can be said to be successful. Not as satisfactory, however, is the frontal strength on the western flank of the meander. The contour lines of the highest temperature are not able to crest northeastward as other contour lines do in that region. This is the major forecast-observation discrepancy that affects the forecast skill, with skill scores defined by anomaly correlation (refer to §5.6.2). Many problems could account for this discrepancy, but here the persistence of boundary conditions after day 2 (August 16) might be the main reason. Nonetheless, as displayed in the forecast sequence (plots not shown here), the thermal pattern changes so much from day to day that whether a dataset with a time span as wide as three days can really provide a detailed structure precise enough for the validation is still in question.

Above are the major features evidenced on the horizontal maps. The section temperature distribution also tells something interesting about this forecast. The solitary cold center of 19 August 1993 near the bottom south of the front (Fig. 5.11) shows up on the fifth forecast day (Fig. 5.20), with some smaller features riding on it. On day 8 (August 22), the observed front pushes a little more southward than the forecast one does, but the wake behind (north of IFF) is well reflected on the wavy 1.8^{o}C contour line. In this sense, the F2 forecast results are well correlated to the observations, though the second cold center over the bottom doesn't find its way into Fig. 5.20 (right), and the wake is not strong enough to be quantitatively satisfactory.

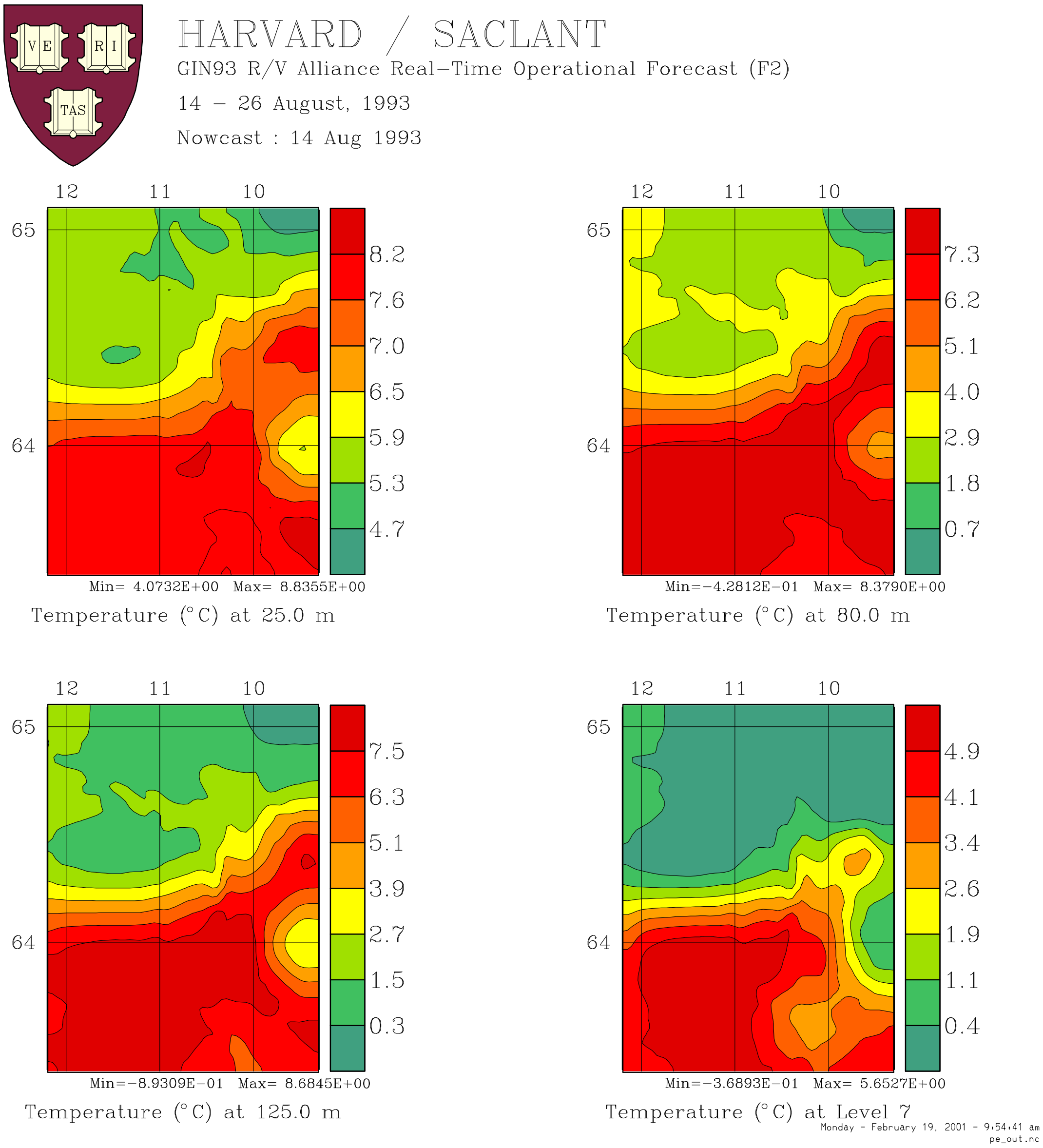

Figure 5.15: F2 nowcast of the day-0 (August 14) temperature.

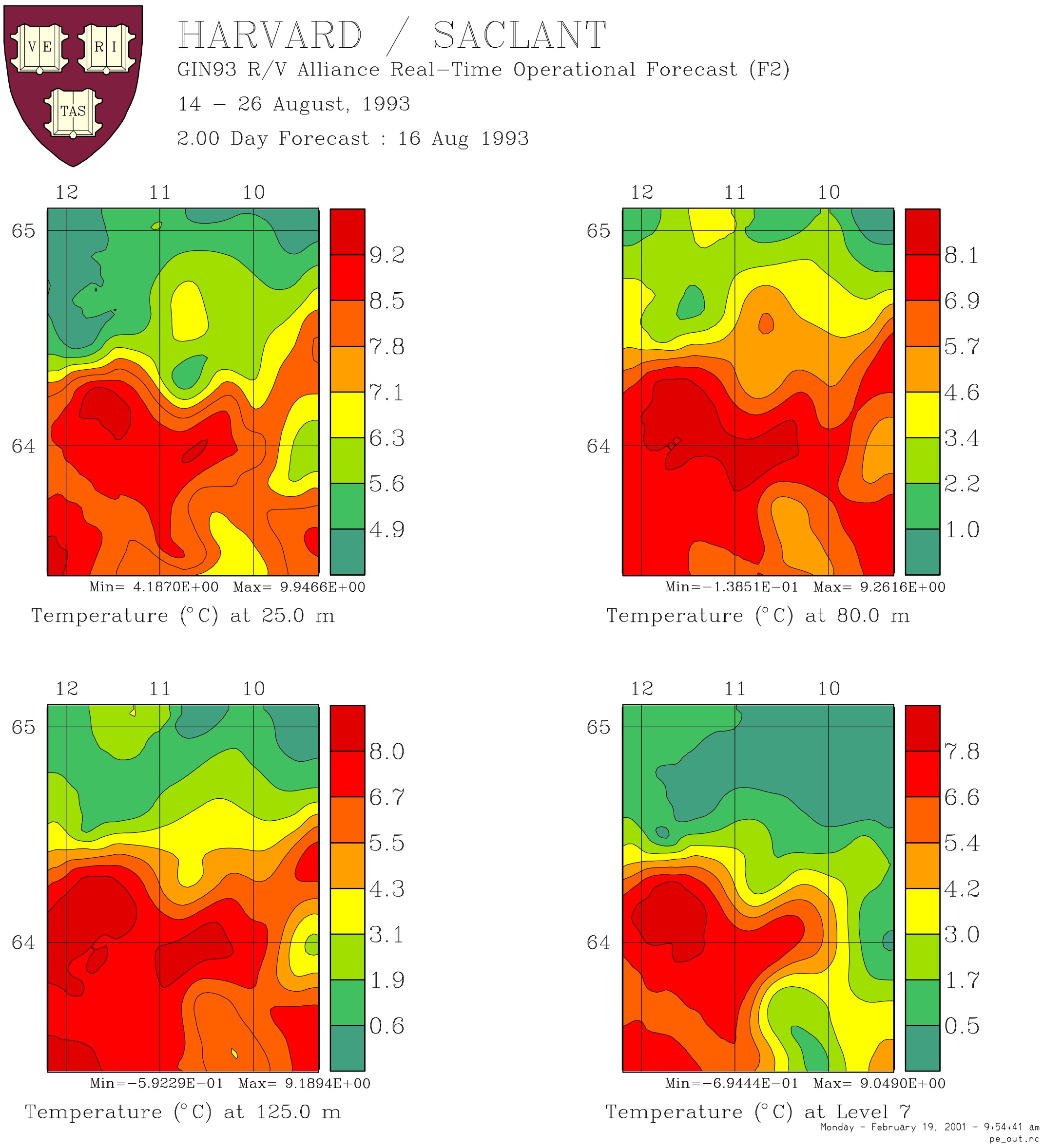

Figure 5.16: F2 nowcast/forecast of the day-2 temperature (August 16). Initialization is completed by this step.

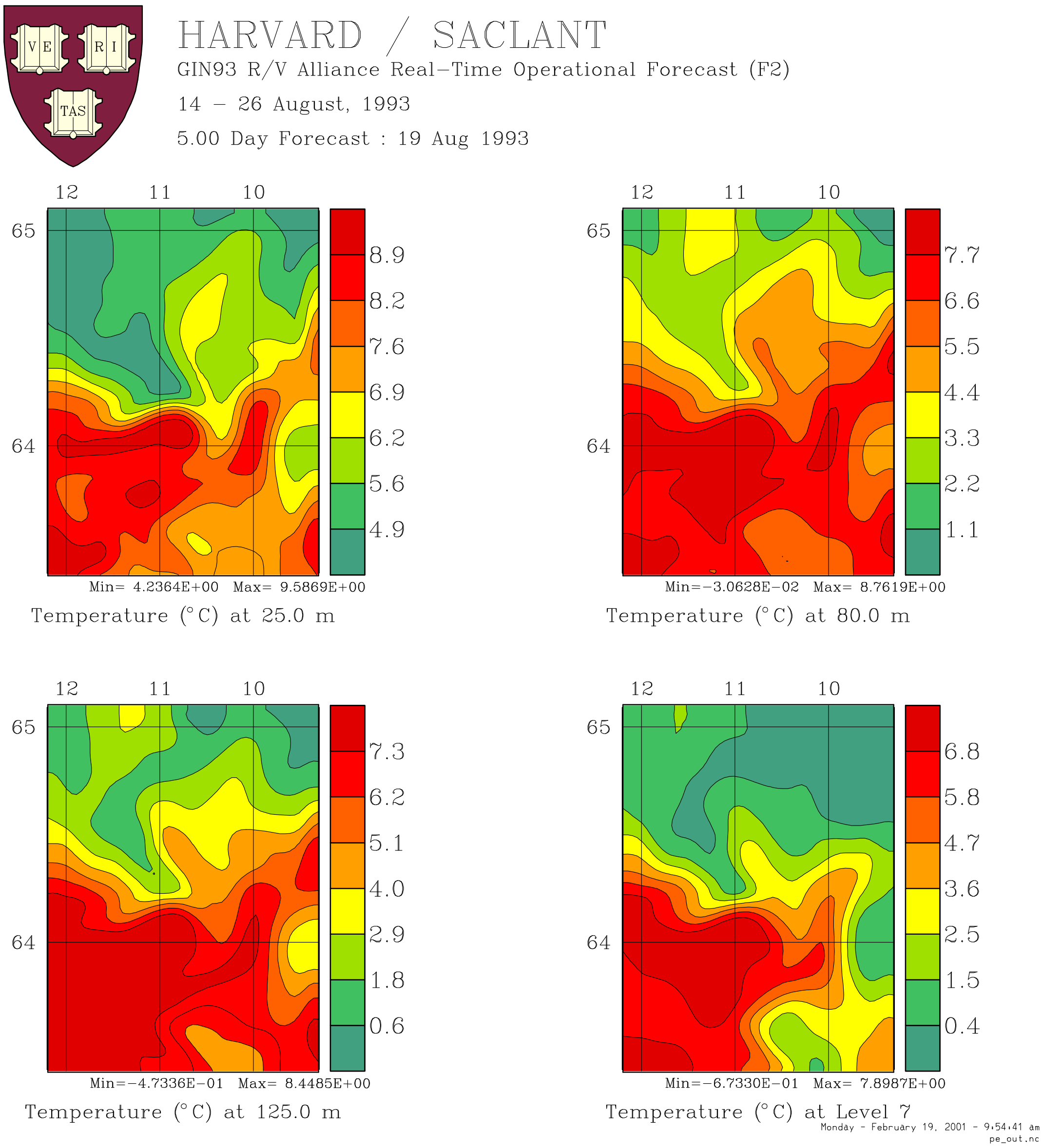

Figure 5.17: F2 forecast of the day-5 (August 19) temperature. The southeastward straightening of the front is clearly seen at all the levels shown here.

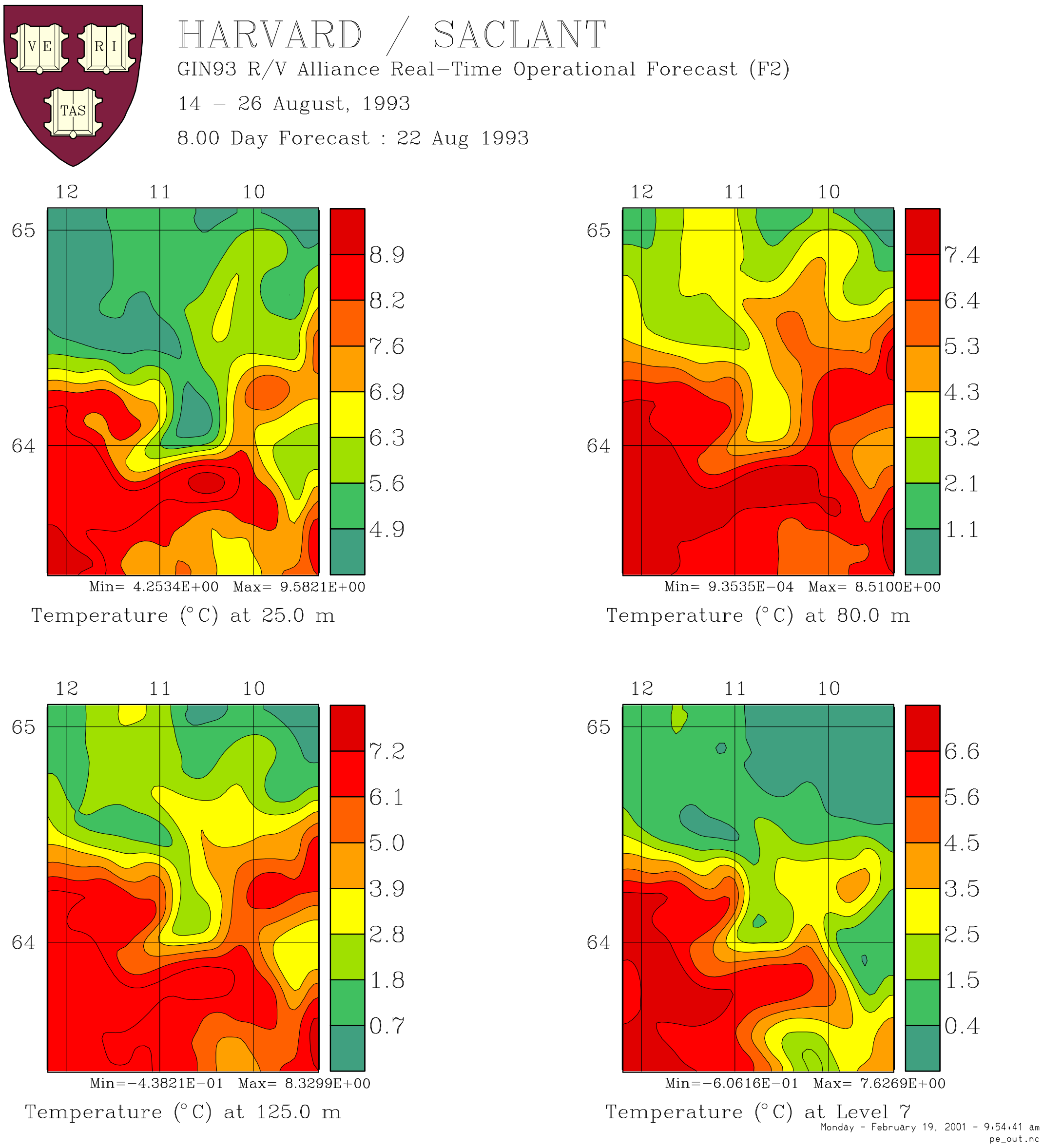

Figure 5.18: F2 forecast of the day-8 (August 22) temperature. Notice the deep-sock meander in the middle of the model domain.

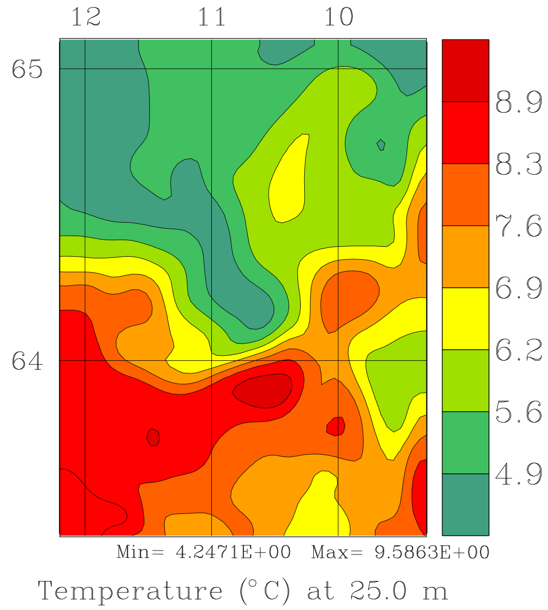

Figure 5.19: F2 forecast of the 25-meter temperature of day 7 (August 21).

5.4.3 F5 forecast

Assimilation of the zigzag data

The F5 forecast starts with the same setting as that of F2. The difference lies in its assimilation of the zigzag data en route to day 8 and beyond. The zigzag dataset, which was acquired on August 18 through August 20, is distinct from the other two sets in that

(1) It does not contain any salinity information,

(2) The hydrocasts cover only the central and southern parts of the model domain (cf. Fig. 5.3), and

(3) A considerable number of probes did not reach the bottom.

The first limitation is removed with the technique of salting described before. (The initialization CTD/XCTD casts are used for comparison.) Robinson et al. (1996) extrapolated the incomplete XBT profiles to the bottom and so the third limitation is not a problem. But here we choose not to apply this extrapolation. (The reason will be seen in the § 5.5.4.) We simply take what they have and prepare the assimilation maps with all the hydrographic data

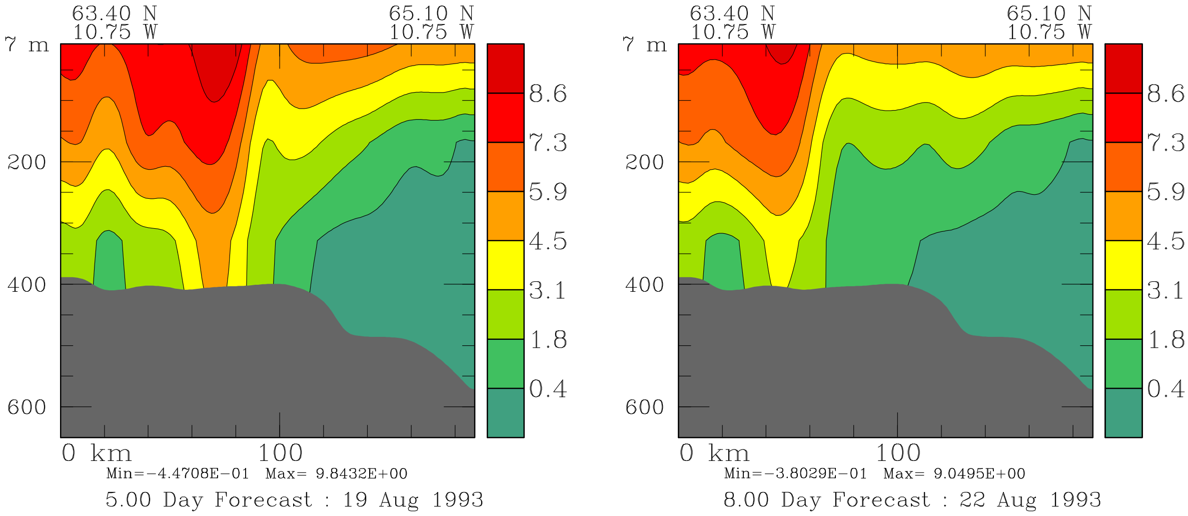

Figure 5.20: F2 forecast temperature on a meridional section (10.75°W) across the deep-sock meander for day 5 (left) and day 8 (right).

available up to forecast day 6 (August 20). In doing so the third limitation is also gone. We are now left with only (2). It is this limitation that makes the zigzag data assimilation an issue for the F5 forecast.

A direct result of the limited XBT coverage is the nonuniformly distributed observation error field. As shown in Fig. 5.21b, the observed temperature error is very large outside the zigzag area defined in Fig. 5.9 (the middle column), compared to either that of the interior or that of the initialization survey (Fig. 5.21a). We will show now, in the following paragraph, that this kind of error nonuniformity may introduce spurious structure to the forecast field if the observation is assimilated with a classical scheme.

Let us illustrate the problem with a simple example. In this example, the physical space is simplified by considering only two locations, location 1 and location 2. Suppose there is an observation T_o and a prediction[3] T_p^- for some field T (not necessarily temperature, but any prognostic variable). They are evaluated at locations 1 and 2, and denoted as $T_o(1)$, $T_o(2)$, $T_p(1)$, $T_p(2)$, respectively. Suppose further that

$$T_o(1) \quad = \quad T_p(1) + C$$

[3]Here the superscript "-" is used to signify a field before any observation is assimilated. Likewise, the superscript "+" will be used for a field after assimilation.

$$T_o(2) \;=\; T_p(2) + C$$

where C is a constant. Then the observed and the predicted fields share an identical spatial *structure*, though the absolute values are not equal. For many fields, one anticipates a correct structure much more than the absolute value. That is to say, the offset may not be that important. Examples of these fields in geophysical fluid dynamics include temperature, salinity, density, to name but a few. In fact, the forecast skill in the next section will be defined in just this spirit (cf., the anomaly correlation coefficient). As the observed and predicted fields have the same structure, ideally we expect the field after assimilation, denoted as $T^+(1)$ and $T^+(2)$, also preserves the structure. However, with a classical assimilation scheme, this is not possible in general. We will soon see why.

By optimal interpolation (Appendix A5.4), when no correlation between the observation and model prediction is considered, the assimilated field is

$$\begin{aligned}
T_p^+(1) &= \left[E_o^{-1}(1) + E_p^{-1}(1)\right]^{-1}\left[E_o^{-1}(1)T_o(1) + E_p^{-1}(1)T_p^-(1)\right], \\
T_p^+(2) &= \left[E_o^{-1}(2) + E_p^{-1}(2)\right]^{-1}\left[E_o^{-1}(2)T_o(2) + E_p^{-1}(2)T_p^-(2)\right].
\end{aligned}$$

where E_o and E_p are the variance for T_o and T_p, respectively. The new structure, which is characterized by the difference of $T_p^+(2)$ and $T_p^+(1)$, is hence

$$\begin{aligned}
T_p^+(2) - T_p^+(1) &= \left[T_p^-(2) - T_p^-(1)\right] + \\
&\quad \left[E_o^{-1}(2) + E_p^{-1}(2)\right]^{-1}\left[E_o^{-1}(2)T_o(2) - E_p^{-1}(2)T_p^-(2)\right] - \\
&\quad \left[E_o^{-1}(1) + E_p^{-1}(1)\right]^{-1}\left[E_o^{-1}(1)T_o(1) - E_p^{-1}(1)T_p^-(1)\right] \\
&= \left[T_p^-(2) - T_p^-(1)\right] + C\left[\left(1 + \frac{E_o(2)}{E_p(2)}\right)^{-1} - \left(1 + \frac{E_o(1)}{E_p(1)}\right)^{-1}\right].
\end{aligned}$$

This difference will not be equal to the previous $[T_p^-(2) - T_p^-(1)]$, unless $C = 0$, or $E_o(2)/E_p(2) = E_o(1)/E_p(1)$. Unfortunately, neither of these conditions are generally satisfied in real problems. That is to say, with the classical OI, assimilation of an observation with the same structure but not identical in values could result in something with a completely different pattern just because of the presence of error nonuniformity!

This problem actually originates from the classical optimal interpolation methodology. The scheme is to minimize a performance or cost functional such that the observed field and the model prediction as a whole are combined into a new field. This does nothing to discriminate features of one scale against others. This proves to be problematic if the error field is not spatially uniform, as exemplified above. Since error inhomogeneity is unavoidable, one has to keep an eye on the physics instead of just doing the mathematics of optimization. In the present study, this problem could be severe, considering that the zigzag error is in a pattern like Fig. 5.21b. The multiscale issue, therefore, must be addressed in devising an appropriate assimilation scheme for the F5 forecast.

An in-depth study of multiscale assimilation is beyond the scope of this thesis. To cope with the problem we just adopt a simple scheme, a scheme aiming at reducing or removing the C in the above example (as the error field is something not alterable). This is achieved by removing the large-scale part of both the model result and the observation. The two fields that are to be melded are therefore the *de-trended* fields. After the assimilation, the trend (of the model prediction) is added back to get the updated field. In doing so, the potentially spurious structure due to the discrepancy of the predicted and observed large-scale features, if any, is then effectively removed.

The large-scale features, by which we mean here the basin-scale trend, is obtained through a reconstruction using the 2-D scaling basis built before (see Fig. 2.11 on p. 90) with a spatial scale level index $j_0 = 1$. Fig. 5.22 are the basin-scale distributions of both the predicted and observed temperatures, $\bar{T}_o$ and $\bar{T}_p$, for the forecast day 5 (August 19). Notice the difference in value (about 0.5°C) albeit the pattern as a whole is similar. This offset would surely lead to spurious patterns (and hence spurious flows by the thermal wind relation) with the zigzag observation errors, should these two basin-scale features be involved in the optimal interpolation. We therefore subtract these features from the originals and then apply the pointwise OI on the de-trended structures. Suppose the so-assimilated field be $T_p'^{+}$. The forecast temperature is then updated with $T_p'^{+} + \bar{T}_p$.

In order not to shock the system, the assimilation is ramped up from day 4 to day 5, instead of getting to the day-5 melding in one step. The assimilation weight ω (See Appendix A5.4) is increased from 0.50 for day 4, to 0.99 for day 5 ($\omega = 1$ means 100% of the observation is assimilated). The forecast fields are then updated and the updated temperature is like Fig. 5.23. After all the zigzag data are taken in, the model is let go on its own.

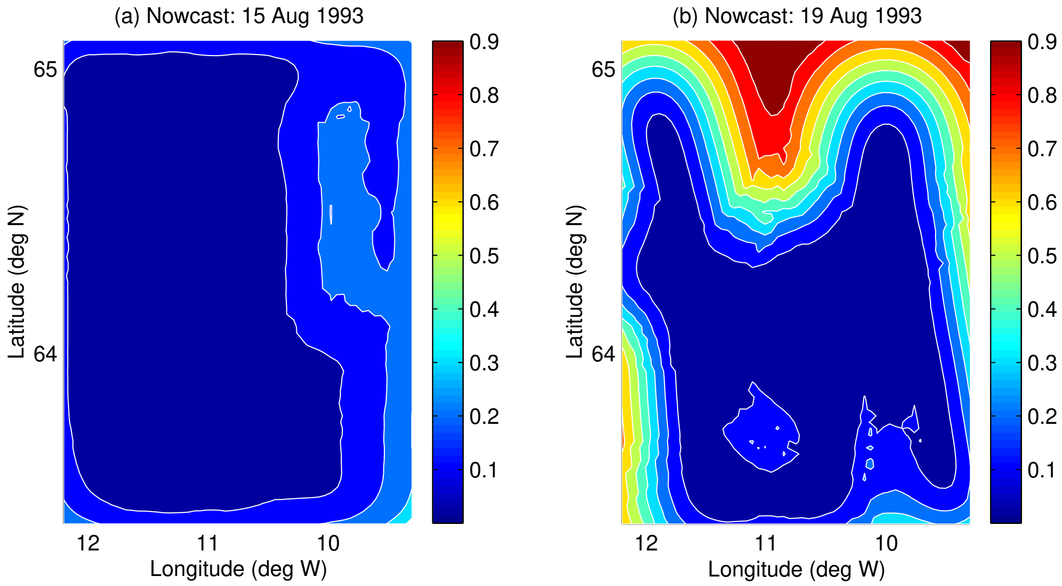

Figure 5.21: The nondimensionalized observational error of the temperature (scaled by its maximal value) for day 1 (a) and day 5 (b). The contour interval is 0.1.

Forecast results

The sequence of the forecast temperature is shown in Fig. 5.23 through Fig. 5.25. Compared to Fig. 5.17, the assimilation of zigzag data introduces many small-scale features, scattered on all but particularly the upper levels. These grid-size events are soon dissipated. They are not significant enough to influence the large scale evolution. On day 8, or August 22, the deep-sock meander is clearly seen in the middle of the domain. The intrusion "depth" (southward extent) is more realistically reproduced than it is in the previous F2 experiment (Fig. 5.18), though the geometric size is somewhat a little smaller. One of the significant improvements here in comparison to the F2 forecast is the reproduction of the frontal strength on the western flank of the meander, and accompanying this is a high temperature pool centered near (11.5°W, 64.2°N) in the surface layers.

As observed (Fig. 5.9), the deep-sock meander is actually not centered at exactly the same position throughout the water column. There is a phase difference between the surface layers and the mid/deep layers. This is particularly clear if one compares the observed 25-meter and the 80-meter temperature distributions. Considering that the flow is generally from the west toward the east, the surface layers lag the deep layers by a significant amount of phase. We

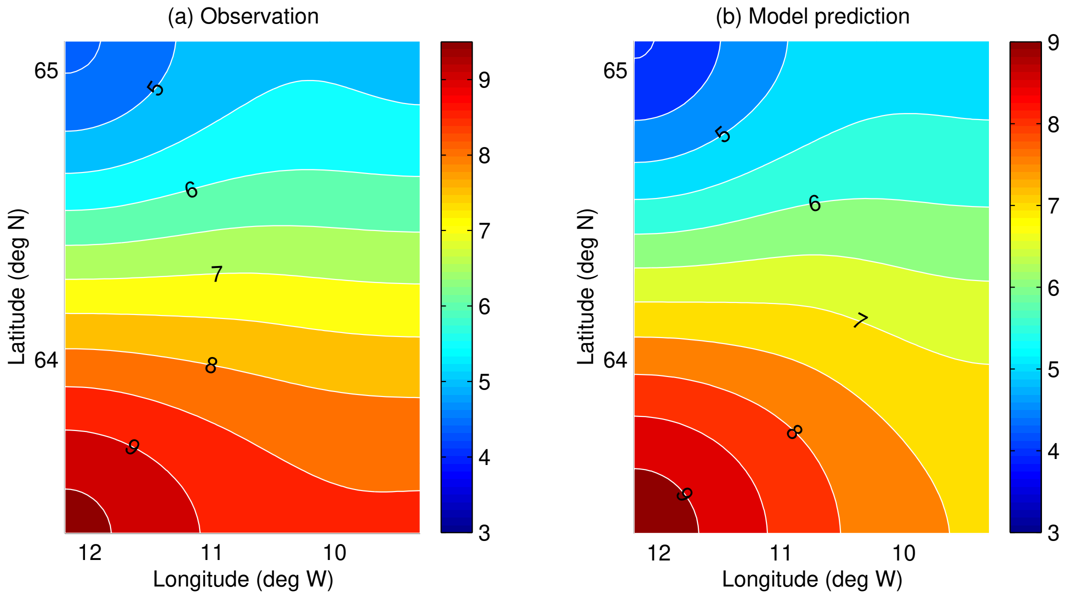

Figure 5.22: The large-scale temperature distribution reconstructed with (a) observation, and (b) model prediction for the fifth forecast day. A window index $j_0 = 1$ is used for the 2-D multiscale window analysis.

know from the linear stability theory that the lag in phase of the upper layers usually implies some mechanism related to baroclinic instability (e.g., Holton, 1992). This phenomenon is thus not trivial, and for this reason, it is interesting to see on Fig. 5.24 there is indeed some indication of such a lag. Upon a close scrutiny, the 80-meter meander is centered at somewhere near 11.5$^{\rm o}$W while the 25-meter meander is a little bit to the west. Our model has at least qualitatively captured the phase lag phenomenon.

The deep-sock meandering intrusion is therefore successfully reproduced. We integrate further to see what happens. After day 8, the model is advanced for another 3 days to day 11 (August 25). The forecast temperature is displayed in Fig. 5.25. From the picture, the meander formed on August 22 now seems to reach the end of its life cycle. An eddy is being snapped off the main stream. The detachment may not be as evident as those observed in the Gulf Stream system (e.g., Robinson et al., 1988; Sloan, 1996), but it is still noticeable in the upper layers (25 m and 80 m).

The eddy detachment is also seen in some section distributions. Particularly in bottom layers, it seems to occur even earlier. This is indicated on the section distributions of the day-8 temperature (Fig. 5.26), where a sequence of the forecast temperature on a north-south section (10.75°W) is plotted. The section runs across the center of the observed deep-sock meander. Observe the small-scale features in the middle layers on day 5. They agree well with those of Fig. 5.11 for the same day. After one day's integration, these features are damped out, but with the southward pushing of the front, some new small features appear again just one day later (August 21). Accompanying these events is the appearance of a solitary benthic cold center, a herald of a series of events on the next day (August 22) which eventually lead to a splitting of the cold center into two just above the bottom, and a "wake" with an amplitude comparable to that of Fig. 5.11. The well-reproduced two cold centers and the following large-amplitude "wake" indicate, at least in the deep-sock meander region, a high correlation between the forecast and the observation, and indeed, this forecast skill gets much improved from the F2 forecast, as will be quantitatively evaluated later.

5.5 Sensitivity study

In order to know whether the forecast is sensitive to the choice of certain parameters, four testing experiments have been conducted. They are, (1) the vertical mixing test (to test the eddy viscosity and diffusivity), (2) the filtering and boundary relaxation effect test, (3) the mesh grid resolution effect test, and for the F5 forecast only, (4) the zigzag data assimilation test. For (1), another factor that affects the vertical mixing is the mixed layer depth prescription. But in this study external forcing is totally neglected, so it is excluded from our consideration.

5.5.1 Vertical eddy viscosity and diffusivity

This study is to test the effect of the two vertical mixing parameters, ν (vertical eddy viscosity for the momentum equations) and A_v (vertical eddy diffusivity for the tracer equations) on the simulation. First ν and A_v are chosen respectively to be 0.5 cm^2/s and 0.1 cm^2/s, reduced by more than an order in comparison to the 5 cm^2/s and 5 cm^2/s for the standard run. The outcome (figures not shown here), however, doesn't show a visible difference from those shown in the above section. Likewise, incrementing ν and A_v by 15 cm^2/s does not significantly

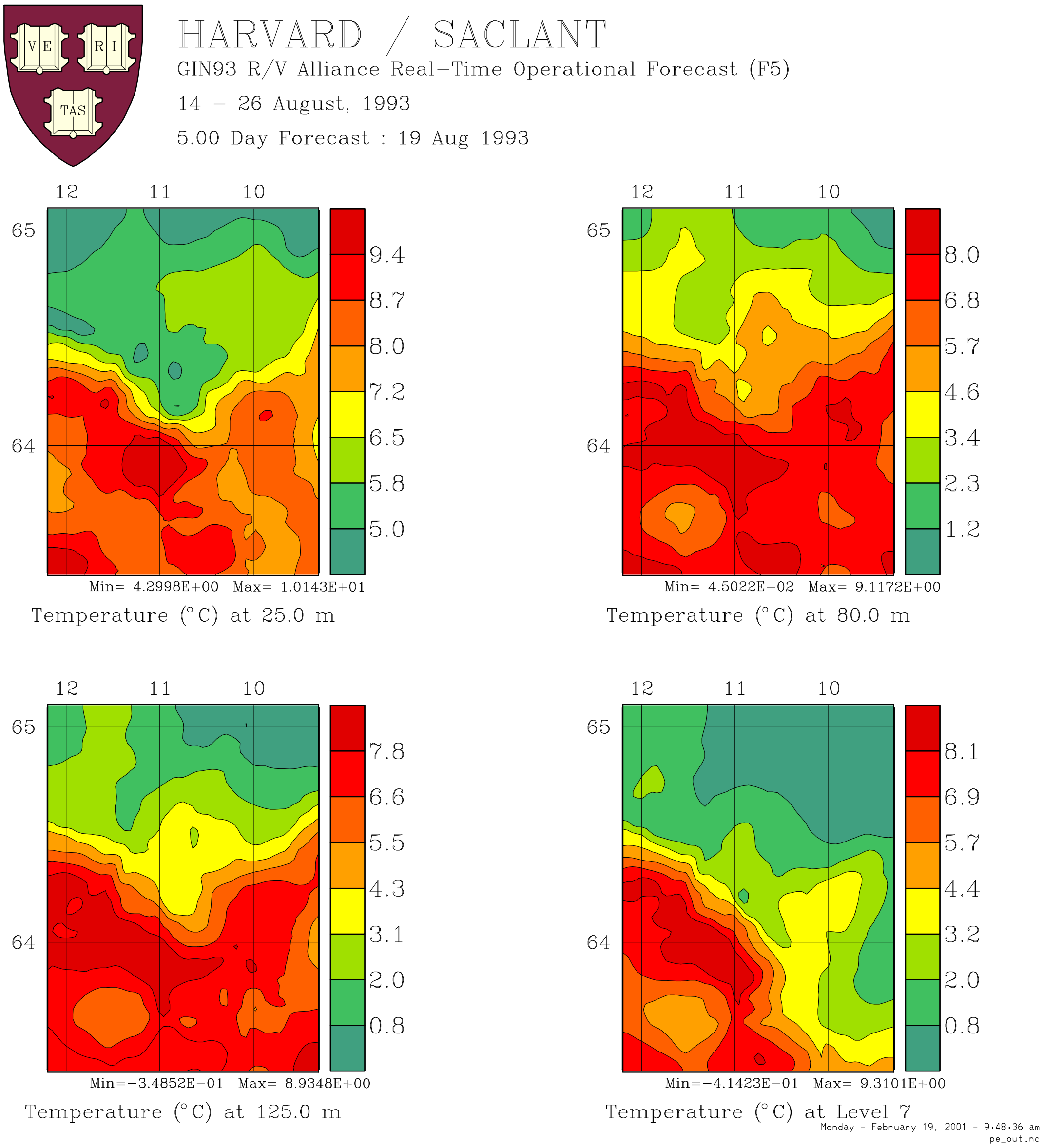

Figure 5.23: F5 nowcast/forecast of the day-5 (August 19) temperature. The small-scale features come in with the *zigzag* data assimilation, which is a melding of the F2 forecast and the zigzag observation by optimal interpolation.

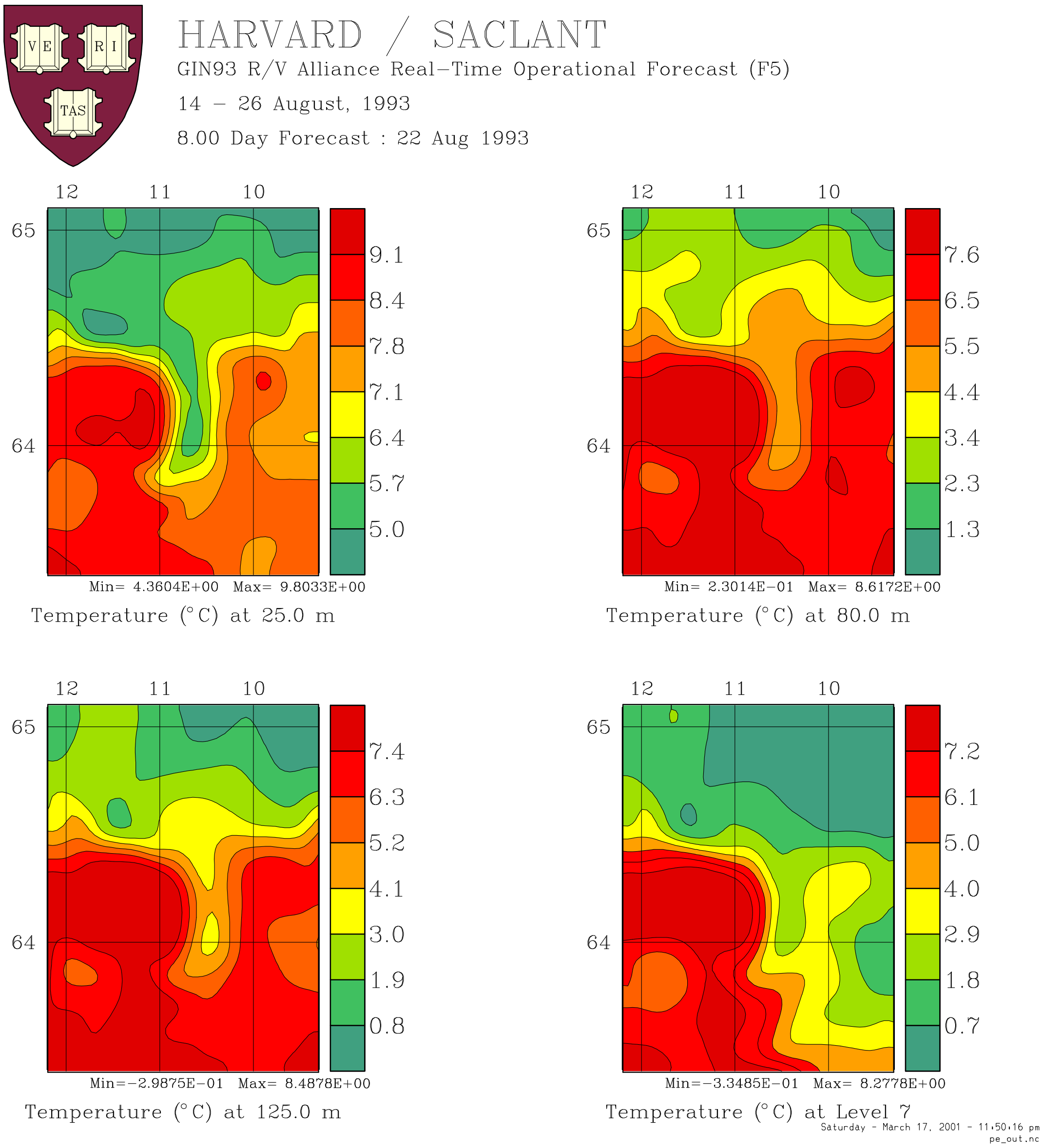

Figure 5.24: F5 forecast of the day-8 (August 22) temperature. The deep-sock meander is in the middle of the domain.

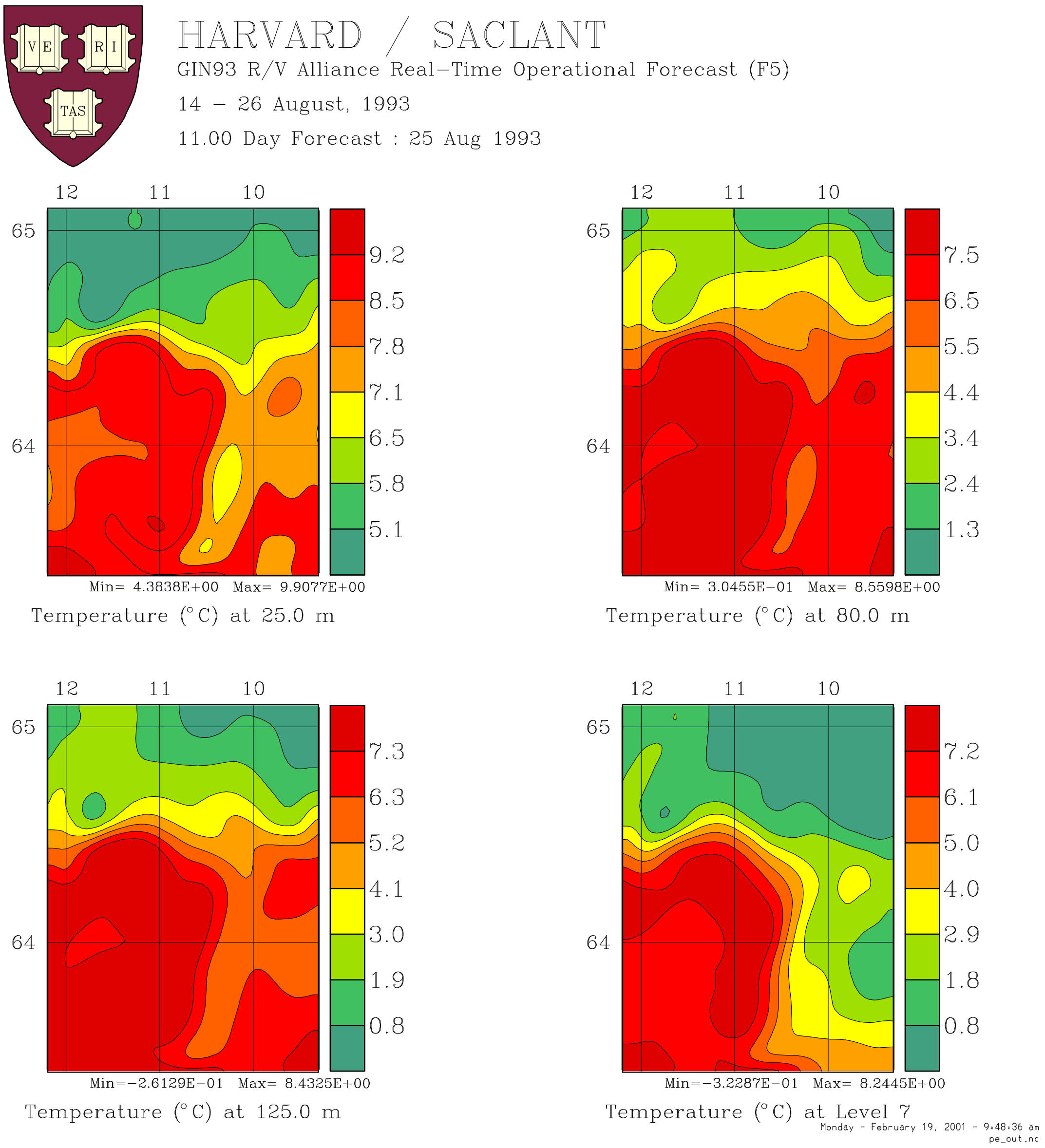

Figure 5.25: F5 forecast of the day-11 (August 25) temperature. This extended forecast reveals an eddy pinch-off in the upper layers.

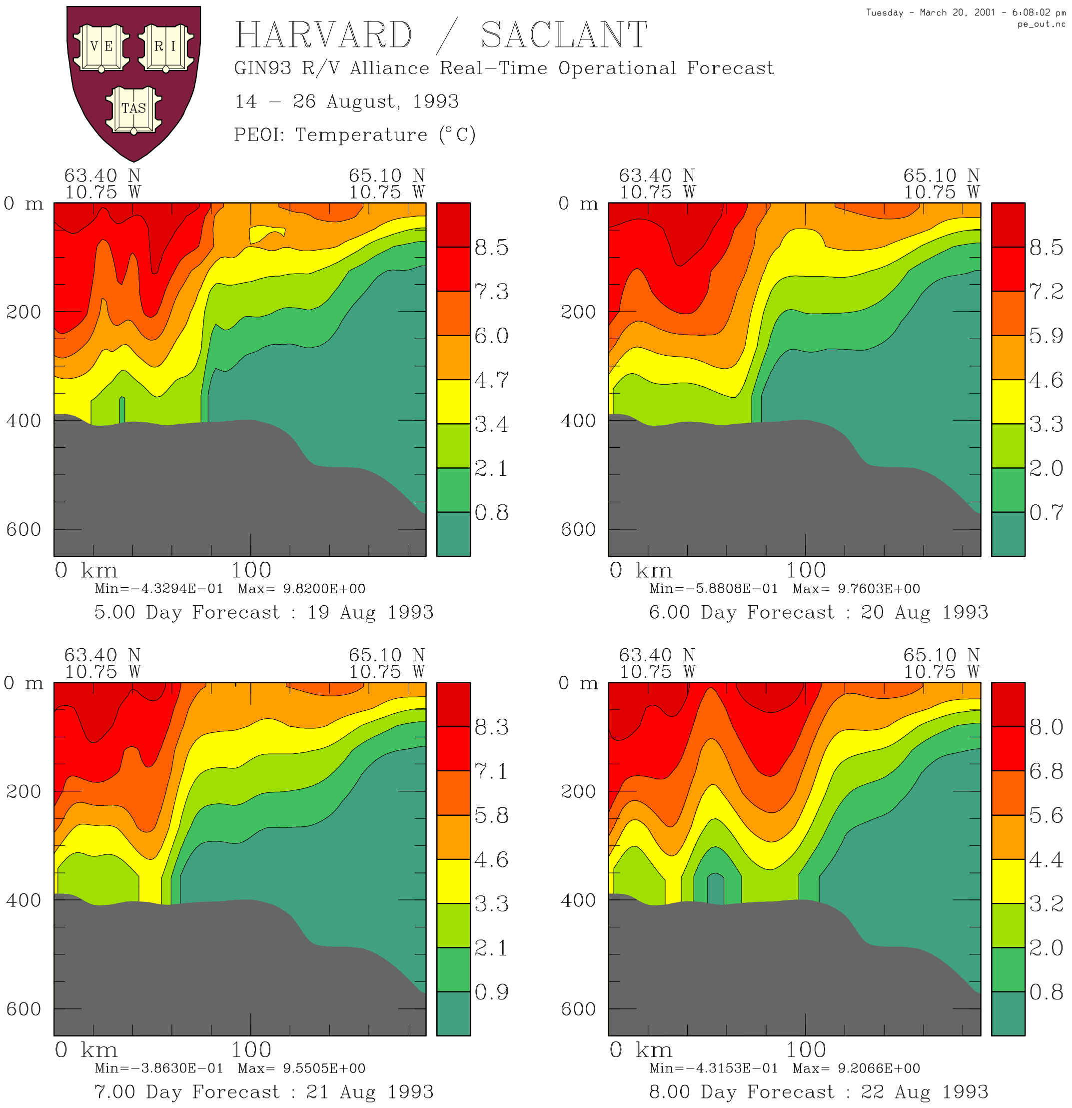

Figure 5.26: The F5 temperature forecast sequence for a meridional section (10.75°W) which runs across the deep-sock meander.

affect the results, either.

5.5.2 Filtering and boundary relaxation

Filtering and boundary relaxation jointly exert effects on the field evolution. Shapiro filters are used to parameterize the subgrid processes. Their application is felt by all points in the domain throughout the integration. The effects of the boundary relaxation, on the contrary, are restricted only to a small region (of scale d_s) in close vicinity to the open boundaries within a limited time (with scale τ_s). Since the model domain of this forecast is small, the large-scale process is on a basin scale, which is driven by the boundary conditions. For this reason, the boundary relaxation should be as small as possible so that these conditions can be felt in time by the interior points. Considering this, it may then be required that the filtering be stronger than usual in order to attain a stable solution. In the standard numerical run, the relaxation is made such that their effects are limited within 2-3 grids immediately close to the boundaries, and filter triplets (2,1,1), (2,1,1), (2,2,1) (see Table 5.2 for interpretation) are chosen for momentum, tracer, and vorticity, respectively. A little stronger than usual as they might be, the filtering effects are within reasonable bounds. By Eq. 5.2, for a typical length scale 25 km, the effective viscosity is estimated to be 160 m^2/s when $\Delta t = 450$ s is used, or 400 m^2/s when $\Delta t = 180$ s, in good agreement with the 300 m^2/s estimated by Willebrand and Meincke (1980) from their observation.

Unlike the vertical eddy viscosity and diffusivity, filtering and boundary relaxation do affect the simulation. Fig. 5.27 is the surface temperature forecast with a spatial decay scale $d_s =$ 3 points and a temporal decay scale $\tau_s = 1800$ s, and a filter triplet $(p, q, r) = (2, 2, 1)$ for both the momentum and tracer equations. It appears on these maps that the day-5 frontal straightening does not occur in the western region, and hence no deep-sock meander shows on day 8 (August 22). Clearly, the meso-scale events have been severely inhibited by the large relaxation and the heavy filtering.

Although changes of boundary relaxation and filter parameters may result in significantly different consequences, our forecast is not that sensitive if these parameters vary within some limits. By experiment, when a second order Shapiro filter is employed, the boundary relaxation scale d_s has to be greater than 0.5 grid point for either momentum or tracer evolution equation.

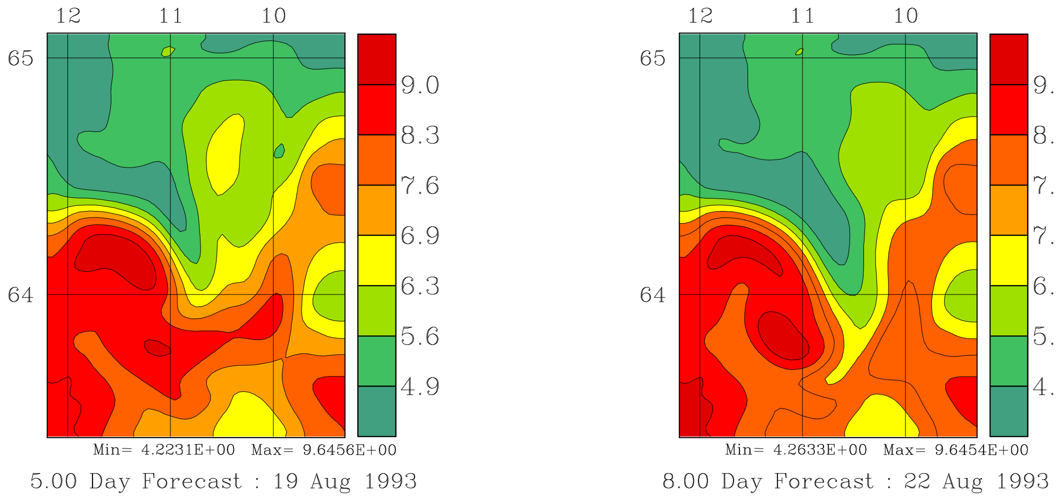

Figure 5.27: Surface temperature predicted with a boundary relaxation d_s=3 points, $\tau_s = 1800s$, and a filter $(p, q, r) = (2, 2, 1)$ for both momentum and tracer equations. The left is the temperature at 25 m for the fifth forecast day (August 19) and the right for the eighth day 8 (August 22).

Otherwise the numerical integration will soon blow up. When $d_s > 1$ point, the resulting day-8 meander pattern appears not that good as the one in Fig. 5.18; When $d_s > 2$ points, the deep-sock meander no longer emerges on that day. If d_s lies on $[0.5, 2]$, change of τ_s does not affect the emergence of the meander, though the shape variation could be significant even by eye inspection. We may then safely conclude that the model is qualitatively insensitive in some parametric neighborhood of the boundary relaxation and filtering.

5.5.3 Mesh grid resolution

Originally, this study began with a model with a coarse resolution ($\Delta x = \Delta y = 5$ km, and 5 levels in the vertical direction), and a satisfactory forecast was obtained. The parameters used in that model are basically the same as those used in the present one, except for a little modification in the choice of (within the "neighborhood" mentioned above) the filter and boundary relaxation in order to ensure stability (filter triplets (2,5,1), (2,1,1), (2,1,1) for momentum, tracer, and vorticity, respectively, and boundary relaxation (d_s, τ_s) = (1, 5400) for momentum and (0.8, 7200) for tracers). Shown in Fig. 5.28 are the surface temperature (25 m) maps for day 5 (August 19) and day 8 (August 22). The two events, the southeastward frontal-axis straightening on August 19, and the deep-sock meandering on August 22, are both

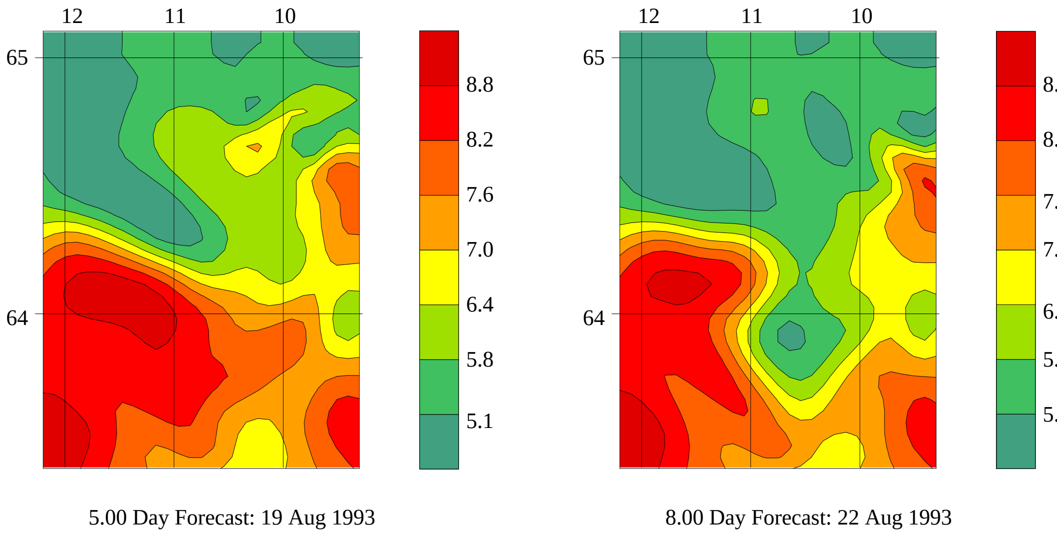

Figure 5.28: The F2 forecast of the 25-m temperature using a model with a grid spacing $\Delta x = \Delta y =$ 5 km in the horizontal and 5 levels in the vertical.

reproduced satisfactorily, though quantitatively it may not be quite as successful as those of Fig. 5.17 and Fig. 5.18. A very impressive observation from Fig. 5.28 is that, in this forecast, the major events are greatly enhanced due to the small-scale-process-removing effect with a coarse resolution. This makes it even easier to identify the events that are essential to the observed phenomena. An examination of the evolution sequence (not shown here) reveals that, from day 2 (Aug. 16) to day 5 (Aug. 19), the front is undergoing an evident re-orientation process until it lies straight toward the southeast. The current attains a speed as strong as 80 cm/s, in good agreement with the observation. On day 6 (Aug. 20), the straightened part of the front suddenly bends from the middle, eventually leading to the formation of a deep-sock meander on day 8 (Aug. 22).

The coarsely resolved model is more sensitive to the choice of boundary relaxation. This is not surprising, since with the same relaxation, a coarsely resolved model has a larger influenced region compared to a finely resolved one. Experiments indicate that a change of d_s by 0.2 point will give rise to a drastic variation in the final meandering pattern.

5.5.4 Zigzag data assimilation

An appropriate zigzag data assimilation scheme is the key to the success of the F5 forecast. Table 5.3 lists nine selected runs for this test. The first is the standard run which has been described in the last section. We know that the zigzag dataset comprises of only XBT measurements, (many casts did not reach the bottom), and some areas are not covered by data. Run 7 and Run 8 are designed to see what if assimilated with only those available data. Unfortunately both of them blow up. In all the remaining runs except Run 2, the assimilation is ramped up from that of day 4 to day 5 (August 19). In Run 2, the assimilation is first ramped up, with the weight ω going up from 0.4 on day 4 to 0.8 on day 5, and then it is ramped down, toward 0.4 again on day 6. This ramp-up and ramp-down strategy was originally used in Robinson et al. (1996), but here it turns out not successful. The meander thus formed is not in the "deep-sock" form.

Table 5.3: Run table for the zigzag data assimilation test.

Run	Assimilation Weights	Fields assimilated	Datasets used	Vert. ext.	Special treatments and comments
1	.50 .99 .00	T S V_i Ψ	Z. + I.	No	Standard run
2	.40 .80 .40	T S V_i Ψ	Z. + I.	No	Intrusion not in a form of deep-sock meander.
3	.50 .99 .00	T S V_i Ψ	Z. only	Yes	Meander shape and orientation not good.
4	.50 .99 .00	T S V_i Ψ	Z. + I.	Yes	Front too sloped west of the meander and meander shape not correct.
5	.50 .99 .00	T only	Z. + I.	No	No deep-sock meander.
6	.50 .99 .00	T and S	Z. + I.	No	Similar to Run 6.
7	.50 .99 .00	T S V_i Ψ	Z. + I.	No	Assimilate in Zigzag domain only. Integration blows up.
8	.50 .99 .00	T S V_i Ψ	Z. + I.	No	Assimilate up to available depth only. Integration diverges.
9	.50 .99 .00	T S V_i Ψ	Z. + I.	No	Classical OI applied. Meander shape and orientation not good.

A note on the symbols or terms used in the table:

T S V_i Ψ: Temperature, salinity, internal velocity, and transport streamfunction.

Assimilation weights: Assimilation weights for days 4, 5, and 6

Vert. ext.: Vertical extension of XBT data downward to bottom

Z. I.: Datasets for zigzag survey and initialization survey, respectively.

As of August 19 (day 5), two datasets, the zigzag dataset and the initialization dataset, have been acquired. Should we use all the data available thus-far or use only the zigzag data for the assimilation? Run 3 is intended to answer this question. Since the XBT dataset is not complete, and a partial assimilation will cause the integration to diverge (Run 8), vertical extension has to be made for all the observation data points. The result of this extension is that, as one might expect, neither the forecast meander shape nor its orientation is good. Even with the initialization data taken in, the extension still yields a forecast result which is not satisfactory (Run 4).

Besides the dataset exploitation, vertical extension, and assimilation weight tests, other testing runs include Runs 5 and 6, which have only part of the fields updated, and Run 9, which does not consider the potential effect of the multiscaling on the assimilation in the presence of error field inhomogeneity. The results of these runs strengthen, from many perspectives, our previous arguments (§ 5.4.3) on how to choose the assimilation scheme and parameters for a successful F5 forecast.

5.6 Quantitative evaluation of the forecast skill

5.6.1 Forecast skill score

As described in the last section, the IFF forecast has qualitatively reproduced with success the major events observed during the zigzag survey and the validation survey. Especially successful is the deep-sock meandering intrusion. By visual inspection, not only its location and its size, but also its geometric shape and orientation have been predicted. This is the first time such a satisfactory operational forecast has ever been made since the acquisition of the IFF '93 dataset. It testifies from this one aspect to the power of the new version of HOPS.

Besides qualitative evaluations, the unprecedented IFF '93 dataset also allows, to some extent, a quantitative validation of the forecast. Following Miller et al. (1995), the forecast skill is measured using the anomaly correlation coefficient (ACC) and the root-mean-square error (RMSE):

$$\mathrm{ACC} = \frac{\overline{T_p' T_o'}}{\left(\overline{T_p'^2}\ \overline{T_o'^2}\right)^{1/2}}, \tag{5.3}$$

and

$$\mathrm{RMSE} = \overline{(T_p - T_o)^2}^{1/2}, \tag{5.4}$$

where the overline denotes an averaging over some designated space domain, and T is the temperature with subscripts p and o representing prediction and observation, respectively. Here by observation we mean the data OA mapped in the mesh grid. In the expression of ACC, T' denotes the mean-removed temperature fluctuation. By these definitions, a high skill should have a high ACC and a low RMSE. We may therefore use the ACC and RMSE of forecast versus the ACC and RMSE of persistence to score the forecast skill, as defined by Miller et al. (1995) and used by Robinson et al. (1996).

5.6.2 Validation of the F2 forecast

Tables 5.4 and 5.5 give the calculated forecast skills against persistence for four different domains: the *Full domain, Subdomain A, Subdomain B*, and the *Zigzag area*. By "full domain" here we actually mean the entire model domain with 3 grids close to the four open boundaries excluded. These excluded points belonging to the relaxation-affected regions, where the forecast fields cannot be compared with real observations. Subdomains A and B are marked in Fig. 5.2. They are designed for the validation of the two major events, the frontal reorientation, and the deep-sock meandering. The zigzag area is the one shown in Fig. 5.9. This is the region where a comparison with the zigzag data is possible. In Table 5.4, no calculation is available below sigma level 8, since the XBT probes have a vertical extent of only 400 m or so.

Generally speaking, Tables 5.4 and 5.5 indicate a successful forecast for day 8 (August 22), especially in the deep ocean. For day 5 (August 19), the ACC is increased versus the persistence for most of the vertical levels, though the RMSE does not see a decrease as a whole. This is actually as expected, for the southeastward current is too weak to make the "valley" deep (south) enough. The southernmost tip of the front lies at too high a latitude. Contrast to this is the day-8 forecast. In Subdomain B, where the deep-sock meander is identified, virtually at all levels the ACC gets increased and the RMSE decreased except at levels 3 and 4 (the 47.5-m and 80-m levels). Even at these two levels, they are much more improved in

Table 5.4: F2 forecast skill vs. persistence (assimilating 16 August /predicting 19 August).

	Subdomain A		Zigzag area		Full domain	
Temp at	ACC change	RMSE change	ACC change	RMSE change	ACC change	RMSE change
7.5 m	+0.017	+10.8%	+0.052	-11.9%	+0.034	+10.7%
25 m	+0.007	+2.4%	+0.041	+5.5%	+0.036	+3.3%
47.5 m	+0.028	-14.6%	+0.022	-0.6%	+0.035	-3.4%
80 m	-0.011	+10.1%	-0.008	+5.3%	-0.013	+3.1%
125 m	-0.005	+2.5%	+0.008	+1.5%	-0.006	-0.4%
σ_6~166	+0.006	-2.2%	+0.020	-3.1%	+0.007	-4.3%
σ_7~215	+0.007	-2.8%	-0.005	-0.6%	-0.006	-1.9%
σ_8~300	N/A	N/A	N/A	N/A	N/A	N/A
σ_9~405	N/A	N/A	N/A	N/A	N/A	N/A

Table 5.5: F2 forecast skill vs. persistence (assimilating 16 August /predicting 22 August).

	Zigzag area		Full domain		Subdomain B	
Temp at	ACC change	RMSE change	ACC change	RMSE change	ACC change	RMSE change
7.5m	+0.029	-0.64%	0	+1.3%	+0.206	-14.0%
25m	+0.015	+8.9%	-0.012	+11.2%	+0.171	-2.5%
47.5m	-0.033	+14.3%	-0.024	+16.3%	-0.036	+21.7%
80m	-0.052	+20.7%	-0.042	+24.7%	-0.062	+32.1%
125m	-0.001	+5.7%	-0.002	+4.0%	+0.033	-1.5%
σ_6~166	+0.063	-17.5%	+0.038	-17.1%	+0.108	-31.6%
σ_7~215	+0.083	-23.7%	+0.057	-23.9%	+0.014	-47.8%
σ_8~300	+0.062	-13.1%	+0.049	-14.6%	+0.012	-45.3%
σ_9~405	+0.019	-2.5%	+0.019	-4.0%	0	-15.8%

comparison to the forecast skill of August 19, if one examines carefully the time sequences of the skills of forecast and persistence from August 16 to August 22 (Figs. 5.29-5.30). In fact, a process with a period of 5-6 days is seen from these sequences. The recurrence of some processes on August 22 makes the persistence highly correlated to the observation. But even with this, the forecast skill still exceeds the persistence skill at most of the levels.

5.6.3 Validation of the F5 forecast

The ACC and RMSE for the F5 forecast are tabulated in Tab. 5.7 for the three validation domains. In this case, only day 8 is used for the verification. Compared to Tab. 5.5, this time it is on the two top levels that the ACC and RMSE are not as good as desired. All the other levels see an increase in ACC and a decrease in RMSE. On levels 3, 4, 5, and 9, the correlation increase and error decrease are particularly significant. The time sequences of ACC and RMSE for some typical levels over Domain B are plotted (thin solid lines) in Figs. 5.31 and 5.32. Level 4 (80 m) is perfect in terms of the forecast skill as defined. The ACC keeps going up until toward the eighth day, while the RMSE continues to go down until reaching its lowest point on August 22. Level 5 (125 m) is also good in this regard. The other two levels, level 2 (25 m) and sigma level 7 are examples with not-so-good quantitative skills. We will show, however, even for these two levels their skill scores will be significantly increased after some phase corrections.

The quantitative evaluation with phase correction is also from Robinson et al. (1996). Here what we need to do is to have the fields lag in the x direction by some distance (moved westward). The lag differs level by level. For all those well forecasted, no lag is necessary; For others such like levels 1 and 2, a little more lag is needed. A detailed report of the necessary lags is provided in Tab. 5.6, with grid interval being the unit.

The spatially lagged F5 skill scores of forecast versus persistence for all levels are calculated in Tab. 5.8. Now on each level, either ACC increase or RMSE decrease is improved, by a significant amount. The time sequences, plotted as the thick solid lines in Figs. 5.31 and 5.32, also indicate a greatly improved forecast skill. Except for the 25-m level, where ACC is a little bit bigger on August 21 than on August 22, all other sequences reveal a maximal ACC on August 22, the very day when forecast is validated. Even for the 25-m level, the 21 August

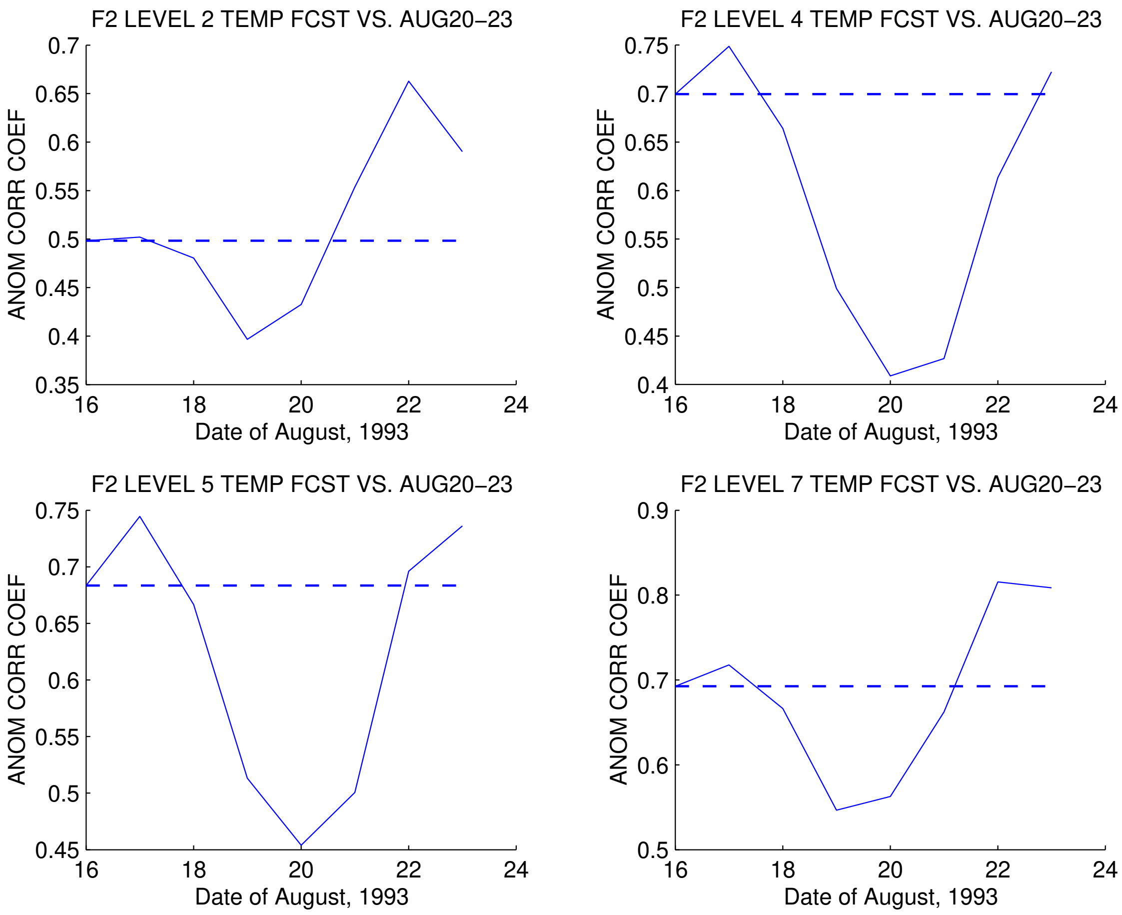

Figure 5.29: Anomaly correlation coefficient for the F2 forecast (solid line) versus the persistence (dashed line).

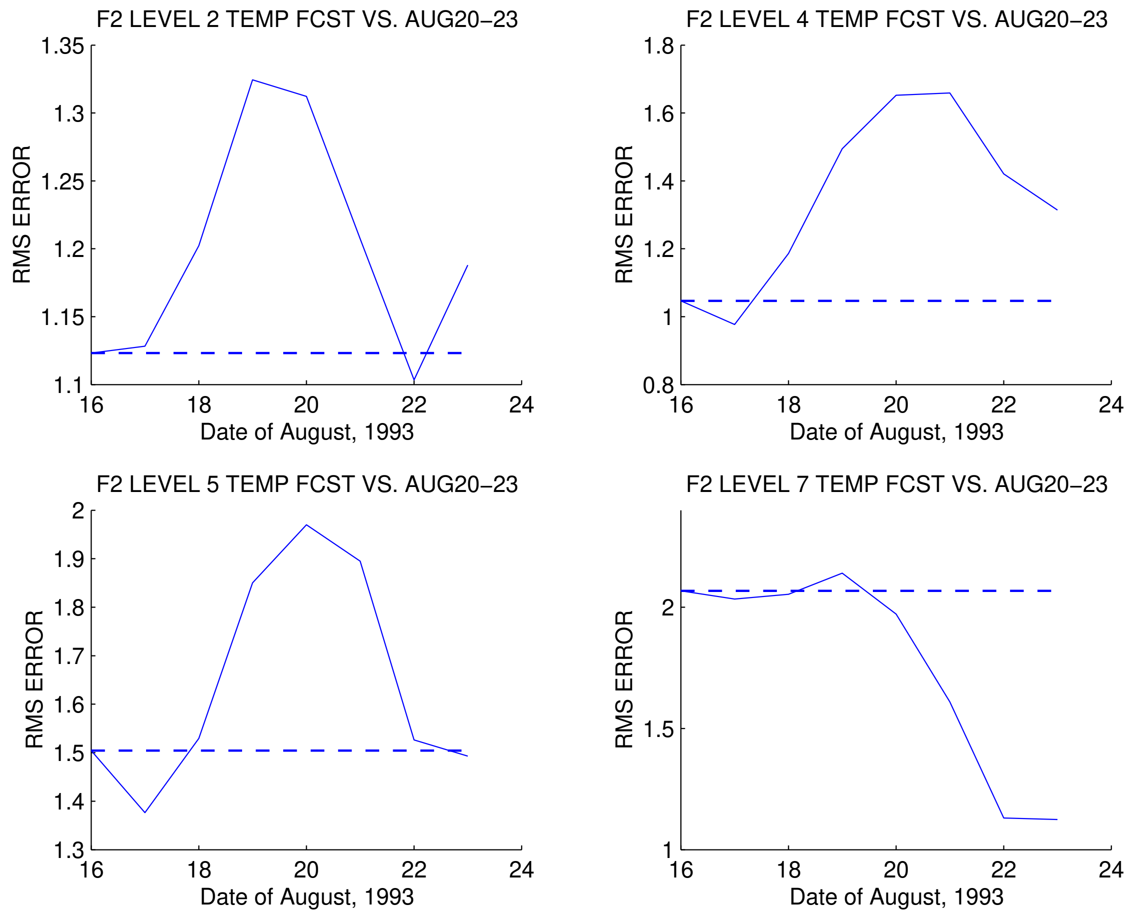

Figure 5.30: Root-mean-square error for the F2 forecast (solid line) versus the persistence (dashed line).

Table 5.6: Spatial lag in x (positive if moved westward). $\Delta x = 2.5$ km.

Level	Spatial lag	Level	Spatial lag
Level-1	8 points	σ-6	3 points
Level-2	6 points	σ-7	6 points
Level-3	0 points	σ-8	6 points
Level-4	0 points	σ-9	2 points
Level-5	2 points		

ACC exceeds the 22 August ACC only by just a small amount, and anyhow the latter is well above the persistence. By all accounts, the forecast skill scores are high for all levels. In other words, the results are highly relevant and correlated to the observations. Our forecast is thus satisfactorily validated.

5.7 Summary

During 14 August-23 August, 1993, an unprecedented dataset with three (the initialization, zigzag, and validation) surveys was acquired on board R/V *Alliance* from the Iceland-Faeroe Frontal (IFF) region. This chapter presents a real-time forecast with this dataset for the IFF variability using the Harvard Ocean Prediction System (HOPS). The model is set up on a hybrid coordinate system. It is totally data-driven, sequentially updated until it is fully initialized. The forecast is launched with two schemes: the F2 scheme which is based solely upon the initialization data, and the F5 scheme which updates the field with the zigzag data. The zigzag updating requires some multiscale treatment because of the apparent inhomogeneity in the error field.

The events observed include a re-orientation of the curved frontal axis on August 16 to a straightened one on August 19, and the formation of a deep-sock or hammer-head cold water intrusion on August 22. These events are observed from the surface to deep levels, with an apparent indication of barotropicity. On an N-S vertical section across the intrusion center, the front is observed pushing southward, and in the process two cold pools are generated just above the bottom, leaving behind a wake-like fluctuating isothermal structure. These events have been reproduced satisfactorily, in either the F2 forecast or the F5 forecast, if

Table 5.7: F5 forecast skill vs. persistence (Assimilating 19 August /Predicting 22 August).

	Zigzag area		Full domain		Subdomain B	
Temp at	ACC change	RMSE change	ACC change	RMSE change	ACC change	RMSE change
7.5m	-0.103	+3.9%	-0.073	+5.6%	-0.032	+4.9%
25m	-0.036	+0.5%	-0.028	+3.8%	-0.018	+0.48%
47.5m	+0.100	-19.5%	+0.057	-16.3%	+0.173	-31.7%
80m	+0.088	-15.1%	+0.038	-14.4%	+0.189	-32.4%
125m	+0.059	-12.3%	+0.027	-10.6%	+0.136	-21.5%
σ_6~166	+0.044	-10.3%	+0.019	-7.5%	+0.076	-8.1%
σ_7~215	+0.026	-5.9%	+0.012	-2.8%	+0.049	+1.6%
σ_8~300	+0.024	-6.5%	+0.015	-4.2%	+0.086	-2.0%
σ_9~405	+0.086	-29.6%	+0.078	-29.3%	+0.186	-37.0%

Table 5.8: F5 spatially lagged forecast skill vs. persistence (Assimilating 19 August /Predicting 22 August).

		Zigzag area		Full domain		Subdomain B	
Temp	moved w. by	ACC change	RMSE change	ACC change	RMSE change	ACC change	RMSE change
7.5m	8 pts	+0.040	-16.5%	+0.015	-13.5%	+0.076	-29.2%
25m	6 pts	+0.060	-18.8%	+0.033	-15.0%	+0.093	-31.1%
47.5m	0 pts	+0.100	-18.8%	+0.057	-16.3%	+0.173	-31.7%
80m	0 pts	+0.088	-15.1%	+0.038	-14.4%	+0.189	-32.4%
125m	2 pts	+0.071	-17.1%	+0.033	-14.9%	+0.150	-25.6%
σ_6	3 pts	+0.070	-21.8%	+0.032	-17.3%	+0.124	-23.4%
σ_7	6 pts	+0.074	-28.9%	+0.042	-23.1%	+0.119	-36.2%
σ_8	6 pts	+0.091	-30.0%	+0.066	-26.8%	+0.173	-50.8%
σ_9	2 pts	+0.126	-40.3%	+0.106	-40.1%	+0.228	-52.8%

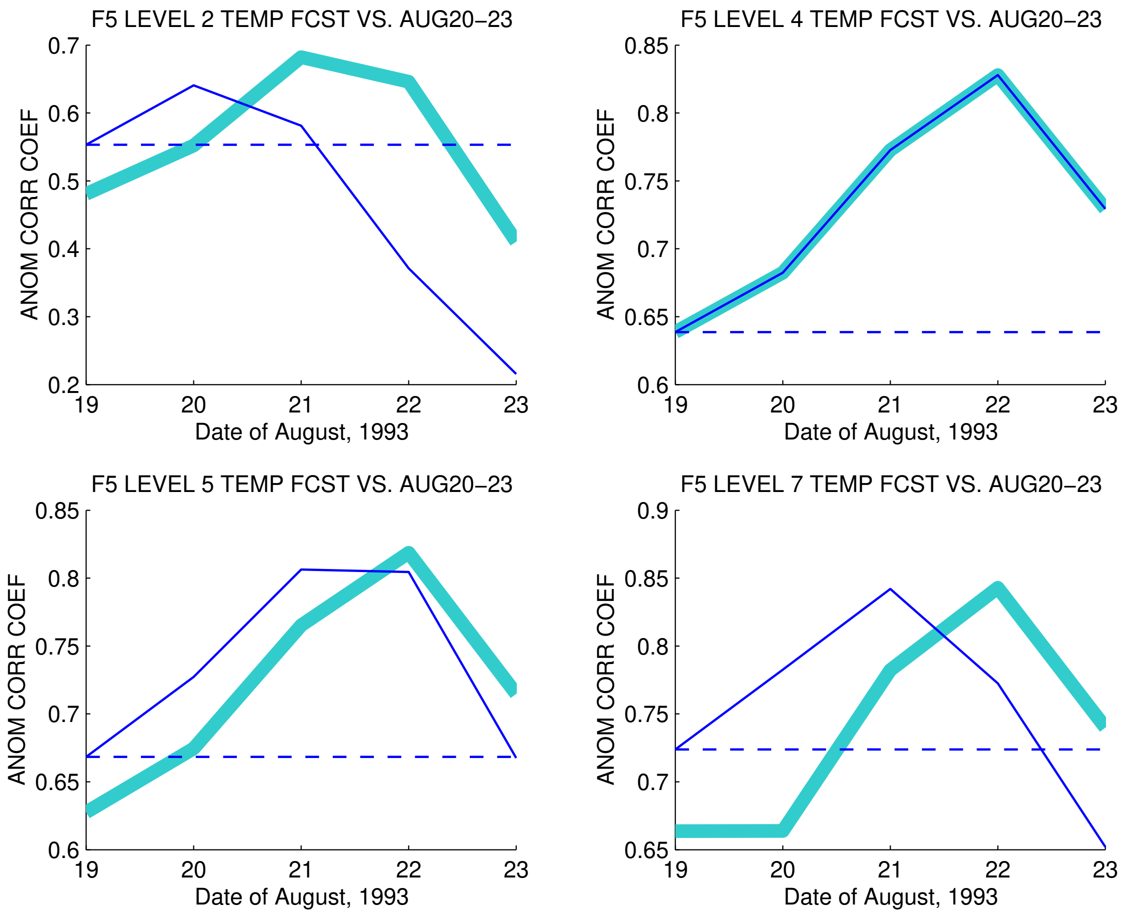

Figure 5.31: Anomaly correlation coefficient for the F5 forecast (solid lines) versus persistence (dashed line). The thick solid line indicates the spatially lagged skill score.

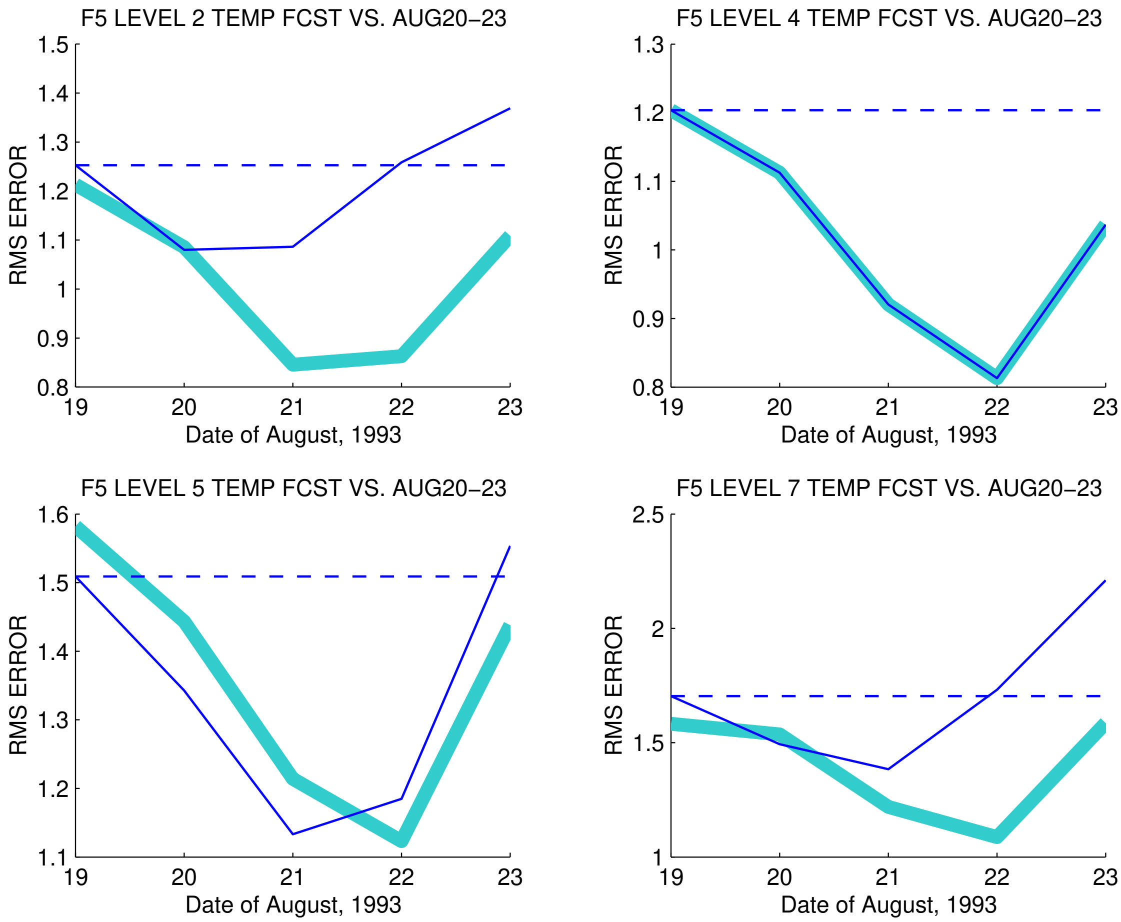

Figure 5.32: Root-mean-square error for the F5 forecast (solid lines) versus persistence (dashed line). The thick solid line indicates the spatially lagged skill score.

inspected qualitatively. Quantitatively, the F5-produced events are also highly correlated to the observation. The forecast skill score is high, as evaluated against persistence, with the anomaly correlation coefficient and the root-mean-square error.

The success of the forecasting is not contingent. Parametric study indicates that the model is not sensitive to the choice of vertical eddy diffusivity and viscosity. The grid spacing is not an issue, either. With the resolution halved, the major events are still there, agreeing qualitatively with the observation. The boundary relaxation jointly with the Shapiro filtering could be a little problematic, but there is no drastic change in the major features produced if these engineering parameters are allowed to vary within reasonable bounds. By all accounts, we safely conclude that the PE model does reflect correctly the fundamental dynamics that lies behind the observed phenomena.

Something that merits more attention is the multiscale data assimilation. We have shown, with a simple example, that in the presence of spatial error inhomogeneity large-scale prediction-observation discrepancy may introduce spurious structures into the fields after assimilation. This problem arises from a lack of multiscale consideration in the classical optimal interpolation scheme. Given that the multiscaling is a reflection of the underlying dynamics, the classical OI may have to be replaced by a new scheme which addresses the multiscale issue as well as the variance minimization. This chapter gives only facets of this problem in carrying forth the F5 forecast. Much ground work is needed before a sophisticated scheme can be made for more generic applications.

Appendix to Chapter 5

A5.1 Principal component analysis

The principal component analysis (PCA), also known as the empirical orthogonal function (EOF) analysis, has been widely used in the science of atmospheres and oceans (see von Storch and Frankignoul, 1998; Preisendorfer, 1988, and references therein). In an elegantly simple way, it transforms a series which might otherwise have a large number of elements into a small set of uncorrelated entities. These entities are referred to as *principal components* (PCs). (An exact definition will be given soon.) With its correlated structures grouped together in these PCs, an efficient representation for a field variable becomes possible.

The PCA can be introduced in a variety of frameworks. In the calculus of variations, it is stated as a problem of first maximizing a performance functional that measures the variability the first PC accounts for, then maximizing that functional for the succeeding PC, and so forth. Suppose we have a field variable $\Psi(t, \underline{\mathbf{x}})$, which is defined on a spatial domain $\mathcal{D}$ and time interval $\mathcal{T}$. The physical space (described by $\underline{\mathbf{x}}$) could be one-dimensional, two-dimensional or three-dimensional. Let the sequence

$$\{\Psi_j(\underline{\mathbf{x}})\}_{j\in J} = \{\Psi(t_j, \underline{\mathbf{x}}), j \in J\}, \qquad J = \{0, 1, 2, ..., n\},$$

($t_j \in \mathcal{T}$, $\forall j \in J$) be a realization of $\Psi(t, \underline{\mathbf{x}})$. Further suppose that,[4]

$$\sum_{j=0}^{n} \Psi_j(\underline{\mathbf{x}}) = 0.$$

Consider the space spanned by the realization, $H = \text{span}\ \{\Psi_j(\underline{\mathbf{x}})\}$. Apparently $H \subset L_2(\mathcal{D})$, and $\dim\ H := m + 1 \leq n + 1$ since the spanning set might not be linearly independent.

[4]This is not essential. An ensemble mean can always be removed before the analysis is applied.

The finite dimensionality of H implies H is a closed subspace of $L_2(\mathcal{D})$ (cf. Chapter 2). Our objective is to find for H an orthonormal basis, $B = \{e_i(\underline{\mathbf{x}}), i = 0, 1, ..., m\}$.

There could exist many such B's in H. We choose a $\{e_i(\underline{\mathbf{x}})\}$ which maximizes the energy (variance) first of the zeroth mode, then of the first, then of the second, and so forth. Elements of such a basis are called *empirical orthogonal functions* (EOFs), and corresponding to them are the EOF modes. Any function in H can be represented with these modes. Particularly,

$$\Psi_j(\underline{\mathbf{x}}) = \sum_{i=0}^{m} \alpha_i^j e_i(\underline{\mathbf{x}}) \tag{A5.5}$$

where

$$\alpha_i^j = \langle \Psi_j(\underline{\mathbf{x}}), e_i(\underline{\mathbf{x}}) \rangle \,. \tag{A5.6}$$

The energy of the i^{th} mode averaged over $\mathcal{T}$ is

$$E_i = \frac{1}{n+1} \sum_{j=0}^{n} \left(\alpha_i^j\right)^2 . \tag{A5.7}$$

It is easy to prove (Aubry et al, 1988), by using the calculus of variations, that the maximization of E_0, then E_1, and so on, is is equivalent to solving a Fredholm integral equation eigenvalue problem (cf. Kreyszig, 1989)

$$\int_{\mathcal{D}} R(\underline{\mathbf{x}}, \underline{\mathbf{x}}')\ e(\underline{\mathbf{x}}')\ d\underline{\mathbf{x}}' = \lambda e(\underline{\mathbf{x}}), \tag{A5.8}$$

with the kernel

$$R(\underline{\mathbf{x}}, \underline{\mathbf{x}}') = \frac{1}{n+1} \sum_{j=0}^{n} \Psi_j(\underline{\mathbf{x}}) \Psi_j(\underline{\mathbf{x}}') \tag{A5.9}$$

being the element of a (symmetric) autocorrelation matrix. The eigenvalues λ and eigenfunctions $e(\underline{\mathbf{x}})$ can be found easily with the aid of computers. The obtained eigenfunctions $e_i(\underline{\mathbf{x}})$ form a total and orthogonal set, spanning the whole space H (Preisendorfer, 1988). If $\{e_i\}$ is normalized, the eigenvalues λ_i are just the energy E_i attached to the corresponding $e(\underline{\mathbf{x}})$. They are called the *principal components* (PCs) for $\{\Psi_j(\underline{\mathbf{x}})\}$. Since the PCs are maximized toward the low orders, functions in H are expected to be effectively represented with just a few EOF modes.

A5.2 Objective analysis

The objective analysis, or OA for short, is a powerful tool in meteorology and physical oceanography which has been particularly useful for data assimilation and field estimation. Its derivation can be found easily in literature (e.g., Gandin, 1965; Bretherton et al., 1976). In this appendix, we provide only a brief introduction. Note the notation here may conflict with that elsewhere. The reader is supposed to take this section as a self-contained one.

OA is a direct corollary of the Gauss-Markov theorem in statistics. For a scalar variable $\psi = \psi(\underline{\mathbf{x}})$, $\underline{\mathbf{x}} = (x, y) \in \mathcal{D}$, with a finite number of measurements, Ψ_r, at N locations $\underline{\mathbf{x}}_r$, $r = 1, 2, ..., N$, the problem is stated as follows:

Find the best linear estimator (least square) of $\psi(\underline{\mathbf{x}})$, given that

(i) ψ is a random variable, and its statistical properties are homogeneous and isotropic on $\mathcal{D}$.

(ii) $\psi(\underline{\mathbf{x}}) \sim N(\underline{\mathbf{0}}, \underline{\underline{\mathbf{C}}})$, *i.e.*, $\psi(\underline{\mathbf{x}})$ obeys a multivariate Gaussian distribution.

(iii) The errors $\epsilon_r = \Psi_r - \psi_r$, $r = 1, 2, ..., N$, are uncorrelated with each other and with the field ψ. They are also normally distributed, $\underline{\epsilon} \sim N(\underline{\mathbf{0}}, \underline{\underline{\mathbf{E}}})$, and $\underline{\underline{\mathbf{E}}} = E\underline{\underline{\mathbf{I}}}$, where $E > 0$ is a given constant.

Assumption (i) was justified by Gandin (1965). It allows an explanation of ψ as a realization of a statistical ensemble, and hence allows a replacement of the ensemble mean by the mean over domain $\mathcal{D}$. In (ii), the zero-mean assumption is actually not essential. It can be always relaxed by a translation. In order not to compound the notational problem, here it is set to be zero. Usually the entries of the covariance matrix $\underline{\underline{\mathbf{C}}}$ is assumed to depend only on the distance between the two points considered, which correponds to the stationary assumption in time series analysis. The third assumption states that ϵ is independent of ψ, since for normal distribution, uncorrelation implies independence (Gut, 1998).

Let an overbar denote the mean over $\mathcal{D}$, and let subscripts signify where a variable is evaluated. The problem becomes, in a more formal language, to find a $\hat{\psi}_x$ to minimize

$$\overline{\left(\psi_x - \hat{\psi}_x\right)^2}$$

subject to

$$\overline{\psi_x} = 0 \tag{A5.10}$$

$$\overline{\psi_x \psi_{x+\xi}} = \delta_{x,x+\xi} \tag{A5.11}$$

$$\overline{\epsilon_r \psi_s} = 0, \qquad \epsilon_r = \Psi_r - \psi_r \tag{A5.12}$$

$$\overline{\epsilon_r \epsilon_s} = E\delta_{rs} \qquad \text{for } r, s = 1, ..., N \tag{A5.13}$$

By the Gauss-Markov theorem (see Bretherton et al, 1976), this results in a linear estimator $\hat{\psi}$ of ψ given the observation Ψ:

$$\hat{\psi}_x = \sum_{r=1}^{N} C_{xr} \sum_{s=1}^{N} A_{rs}^{-1} \Psi_s. \tag{A5.14}$$

In a matrix form, this is

$$\underline{\hat{\psi}} = \underline{\underline{\mathbf{C}}}\, \underline{\underline{\mathbf{A}}}^{-1} \underline{\mathbf{\Psi}} \tag{A5.15}$$

where

$$A_{rs} = \overline{\Psi_r \Psi_s} = \sigma_{rs} + E\delta_{rs} \tag{A5.16}$$

$$C_{xr} = \overline{\psi_x \Psi_r} = \sigma_{xr} \tag{A5.17}$$

with the variance of the error

$$\overline{(\psi_x - \hat{\psi}_x)^2} = C_{xx} - \sum_{r,s=1}^{N} C_{xr} C_{xs} A_{rs}^{-1}, \tag{A5.18}$$

which may be intuitively interpreted as that the reduction of uncertainty is due to the addition of information through observation.

Eqs. (A5.14) and (A5.18) are the two equations for the objective analysis. A detailed derivation of them can be found in Bretherton et al. (1976). In general the so-obtained estimator does not satisfy the orthogonality condition: The difference between the observation and the estimator is correlated to the estimator itself. This is quite different from the linear regression, which always yields estimators independent of the measurement-estimator residual for normally distributed variables (Gut, 1998).

An issue that must be resolved before any application is made is how the field can be translated so that the homogeneity and isotropicity assumption is met. Apparently the field itself, like temperature or salinity, is in general not homogeneous or isotropic at all. A good choice is the "eddy field" (Gandin, 1965), the residual with the large-scale features removed. For this reason, the OA application in real ocean problems is usually implemented with a *two-stage approach*, which was first proposed by Lozano et al. (1996) and now has been adopted in the Harvard Ocean Prediction System. For the details of this approach, refer to the original paper of Lozano et al.. A brief description of the general procedure is given in §5.2.3.

A5.3 Structure and correlation analysis

In Chapter 5, we need to perform structure and correlation analysis to estimate for the OA the scaling parameters. Specifically, we need to find two functions, the structure function and the autocovariance function (or simply correlation function).

Suppose we have a realization, $T(\underline{\mathbf{x}})$, of some state variable, with $\underline{\mathbf{x}}$ being the position vector. Here T is taken as a random variable. Following Gandin (1965), the structure function B and correlation function C are defined (without much rigor),

$$B(\underline{\mathbf{x}}_1, \underline{\mathbf{x}}_2) = E\left[T(\underline{\mathbf{x}}_1) - T(\underline{\mathbf{x}}_2)\right]^2, \tag{A5.19}$$

$$C(\underline{\mathbf{x}}_1, \underline{\mathbf{x}}_2) = E\left[T(\underline{\mathbf{x}}_1) \cdot T(\underline{\mathbf{x}}_2)\right]. \tag{A5.20}$$

Estimation of B and C encounters difficulty when statistical properties of T vary in space. In that case, the ensemble for a location is usually too small to make the estimation statistically significant. A usual way to circumvent this problem is to introduce an assumption of homogeneity and isotropicity. A homogeneous and isotropic T has statistical properties translationally and rotationally invariant. The resulting $B(\underline{\mathbf{x}}_1, \underline{\mathbf{x}}_2)$ and $C(\underline{\mathbf{x}}_1, \underline{\mathbf{x}}_2)$ thus depend upon only the distance between the two locations, $r = |\underline{\mathbf{x}}_1 - \underline{\mathbf{x}}_2|$, unrelated to the locations $\underline{\mathbf{x}}_1$ and $\underline{\mathbf{x}}_2$ themselves.

In analogy to the stationary assumption in time series analysis, homogeneity and isotropicity allow a great reduction of the complexity of the estimation problem. Unfortunately, they usually do not hold directly for T. As elaborated by Gandin (1965), they make some sense only for the "eddy" part. That is to say, if $T(\underline{\mathbf{x}})$ is decomposed as

$$T(\underline{\mathbf{x}}) = E(T(\underline{\mathbf{x}})) + T'(\underline{\mathbf{x}}),$$

observational facts indicate that the above assumption applies to some extent to $T'(\underline{\mathbf{x}})$ rather than $T(\underline{\mathbf{x}})$ itself. For this reason, re-define the statistical structure as

$$b(r) = E\left[T'(\underline{\mathbf{x}}_1) - T'(\underline{\mathbf{x}}_2)\right]^2, \quad \text{(A5.21)}$$

$$c(r) = E\left[T'(\underline{\mathbf{x}}_1) \cdot T'(\underline{\mathbf{x}}_2)\right], \quad \text{(A5.22)}$$

with $r = |\underline{\mathbf{x}}_1 - \underline{\mathbf{x}}_2|$. In many situations, we may prefer the normalized version

$$\beta(r) = \frac{b(r)}{b(\infty)}, \quad \text{(A5.23)}$$

$$\gamma(r) = \frac{c(r)}{c(0)}. \quad \text{(A5.24)}$$

Here $\gamma(r)$ is the correlation coefficient function. It is actually all we have to estimate since $\beta(r)$ can be derived from it once it is obtained:

$$\begin{aligned} b(r) &= E\left[T'(\underline{\mathbf{x}}_1)\right]^2 - 2E\left[T'(\underline{\mathbf{x}}_1) \cdot T'(\underline{\mathbf{x}}_2)\right] + E\left[T'(\underline{\mathbf{x}}_2)\right]^2 \\ &= 2\ c(0)\ \left[1 - \gamma(r)\right], \end{aligned} \quad \text{(A5.25)}$$

and

$$b(\infty) = 2\ c(0), \quad \text{(A5.26)}$$

$$\beta(r) = 1 - \gamma(r). \quad \text{(A5.27)}$$

The estimation of $\gamma(r)$ begins with transforming the multi-dimensional data $T(\underline{\mathbf{x}})$, which might be irregularly distributed, into a one-dimensional series. Since the γ to be estimated is dependent on the distance r only, we may take it as the only coordinate and carefully discretize it into many distance gradations called "bins". For each bin there must be some data falling in it, and an average of them is taken as the transformed value at the location r corresponding to that bin. The next step is simply to estimate the γ for a equi-distance 1-D series. For convenience, the distance dependences are now replaced by integer indices.

To make an estimator ideal, it should be consistent, unbiased, and positive definite. Using a caret for an estimation, and an overbar for a mean, the estimator provided by Fuller (1983),

$$\hat{c}(h) = \frac{1}{N} \sum_{i=1}^{N-h} \left[T(i) - \bar{T}(i)\right] \cdot \left[T(i+h)\bar{T}(i+h)\right], \quad \text{(A5.28)}$$

$$\hat{\gamma}(h) = \frac{\hat{c}(h)}{\hat{c}(0)}, \quad \text{(A5.29)}$$

$$\hat{\gamma}^*(h) = \hat{\gamma}(h) + \frac{N-h}{(N-1)^2}\left(1 - [\hat{\gamma}(h)]^2\right), \quad \text{(A5.30)}$$

(where N is the size of the sample) proves to be positively definite and only slightly biased for a relatively large sample volume. The bias can be measured approximately

$$E\left[\hat{\gamma}^*(h)\right] = O\left(\frac{1}{N^2}\right) \tag{A5.31}$$

$$Cov\left[\hat{\gamma}^*(h), \hat{\gamma}^*(q)\right] \approx \begin{cases} \frac{N-h}{N^2} + O\left(\frac{1}{N^2}\right) & h = q > 0 \\ O\left(\frac{1}{N^2}\right) & \text{else} \end{cases} \tag{A5.32}$$

and $\hat{\gamma}^*(h)$ is approximately normally distributed. Thus 95% of the estimated correlation coefficient would fall between

$$\pm 1.96 \frac{(N-h)^{1/2}}{N}.$$

With $\gamma(r)$ estimated, $b(r)$ and $c(r)$ are easily obtained from the relations (A5.23), (A5.24), and (A5.25) developed before.

A5.4 Pointwise optimal interpolation

In this study, observational data are assimilated into the model using a pointwise scheme called *optimal interpolation* (OI) (see Lozano, 1996; Lermusiaux, 1997). It melds observations and forecasts with a linear weighted average. The theoretical foundation of this melding is the Bayes rule in statistics.

Suppose, for this moment, we have a random variable ψ, and an observation of ψ, Ψ,

$$\psi + \epsilon = \Psi,$$

with ϵ the mismatch or error. The problem of OI reads:

> *If the statistical properties of ϵ are known, find the expected ψ given the observation Ψ, i.e., find the expectation of $\psi|\Psi$.*

Let $\rho(\psi)$ and $\rho(\epsilon)$ be, respectively, the probability density functions (pdf) of ψ and ϵ. According to the Bayes rule, we have

$$\rho(\psi|\Psi) = \frac{\rho(\Psi|\psi) \times \rho(\psi)}{\int_\Omega \rho(\Psi|\psi) \times \rho(\psi)\ d\psi},$$

where the integeration is taken over Ω, the whole probability space of ψ. In the above equation, $\rho(\Psi|\psi) = \rho_\epsilon(\Psi - \psi)$, and $\rho(\psi)$ is given. The conditional density of ψ and hence the expected value of ψ after the observed information comes in can be easily calculated. Suppose further that ψ and Ψ are normally distributed (the distribution most commonly used in estimation theory), then the ψ given the observation Ψ must be also a normal or a Gaussian, with the expected value and variance determined by the variance of ψ, E_ψ, and the variance of Ψ, E_Ψ. And when ψ and Ψ are not correlated (correlation and dependence are equivalent for normal distribution), the expectation of $\psi|\Psi$, $\bar{\psi}|\Psi$ (an overbar will be used hereafter in this appendix to symbolize the mean), is just equal to a weighted mean of ψ and Ψ, with weights inverse proportional to the corresponding variances (cf., Gut, 1998, Chapter V). More precisely, if

$$\begin{bmatrix} \psi \\ \Psi \end{bmatrix} \sim N\left(\begin{bmatrix} \bar{\psi} \\ \bar{\Psi} \end{bmatrix}, \begin{bmatrix} E_\psi & 0 \\ 0 & E_\Psi \end{bmatrix}\right),$$

then

$$\psi|\Psi \sim N\left(\bar{\psi}|\Psi, E_{\psi|\Psi}\right),$$

with a mean

$$\bar{\psi}|\Psi = E_{\psi|\Psi}\left(E_\psi^{-1}\bar{\psi} + E_\Psi^{-1}\bar{\Psi}\right), \tag{A5.33}$$

and a variance

$$E_{\psi|\Psi} = \left(E_\psi^{-1} + E_\Psi^{-1}\right)^{-1}. \tag{A5.34}$$

Note $E_\psi|\Psi$ is smaller than either E_ψ or E_Ψ. The addition of information (observations) thus reduces the uncertainty by a significant percentage. If there is a correlation between ψ and Ψ, the correlation coefficient γ[5] will be involved in the above equations (cf., Gut, 1998, p.132; Brockett, 1996):

$$\bar{\psi}|\Psi = E_{\psi|\Psi}\left[E_\psi^{-1}\bar{\psi} + E_\Psi^{-1}\bar{\Psi} - 2\gamma\sqrt{E_\psi^{-1}E_\Psi^{-1}}(\psi + \Psi)\right], \tag{A5.35}$$

and

$$E_{\psi|\Psi} = \left(E_\psi^{-1} + E_\Psi^{-1} - 2\gamma\sqrt{E_\psi^{-1}E_\Psi^{-1}}\right)^{-1}. \tag{A5.36}$$

[5]This γ is different from that of Appendix A5.3 in that here only a single point is involved.

The above result can be easily generalized to problems with arbitrarily many statistical variables, with a simply replacement of ψ and Ψ with two vectors, and the E's with the corresponding matrices of variance $\underline{\underline{\mathbf{E}}}$. Since in the current study we care about only pointwise interpolation, a vectorized version is not necessary.

For our problem, there is a model predicted quantity T_p^-, plus an observation T_o. We are asked to find the best estimator T_p^+. Substituting T_p^- for ψ, T_o for Ψ in Eqs. (A5.35) and (A5.36), we obtain, after some algebraic manipulation, the best estimator

$$T_p^+ = \alpha T_o + (1-\alpha) T_p^-, \tag{A5.37}$$

where

$$\alpha = \frac{E_{p-} - \gamma\sqrt{E_o E_{p-}}}{E_o + E_{p-} - 2\gamma\sqrt{E_o E_{p-}}}, \tag{A5.38}$$

and the variance with T_p^+

$$E_{p+} = \frac{E_o E_{p-}(1-\gamma^2)}{E_o + E_{p-} - 2\gamma\sqrt{E_o E_{p-}}}. \tag{A5.39}$$

This is the linear blending scheme for the OI used in HOPS (Lermusiaux, 1997; Sloan, 1996), which is also seen in Dombrowsky and de Mey (1992). The correlation γ is a user supplied parameter. It has been proposed[6] that we may choose

$$\gamma = \frac{\sum T_o' T_p'}{\sqrt{\sum T_o'^2}\sqrt{\sum T_p'^2}}$$

where the sum is over the whole model domain. However, Dombrowsky and de Mey (1992) argue that there is no advantage for a delibrately chosen γ over the simplest case $\gamma = 0$, and as a matter of fact, $\gamma = 0$ could even be the best choice. For this reason, we adopt a zero correlation between T_p^- and T_o. The OI scheme now reduces to those of Eqs. (A5.33) and (A5.34).

In the HOPS, the variance of T_o, E_o, is obtained through the objective analysis OA (cf. Appendix A5.2). The other variance, E_p, has to be provided by the model. Since solution of an evolution of the uncertainty is expensive, a simple empirical model is used instead. The forecast variance (error) is increased like

$$E_p^-(t) = E_p^+(t-\tau) + 2\left(1 - E^{-\tau^2/\tau_p^2}\right).$$

[6] Carlos Lozano, personal communication

In the equation τ is the time since the last assimilation. The e-folding decay scale τ_p measures the time for a loss of predictability.

For real problems, a *ramped OI* is a common practice in data assimilation to avoid shocking the already ajusted system. In a ramped OI, the weight α in the linear interpolation is pre-multiplied by a factor, ω, $0 \leq \omega \leq 1$, such that no assimilation is performed when $\omega = 0$, and all observation is assimilated when $\omega = 1$ (ω hence reflects the percentage of assimilation). With this Eq. A5.37 now becomes

$$T_p^+ = \omega\alpha T_o + (1 - \omega\alpha) T_p^- . \tag{A5.40}$$

It should be pointed out that, in order for the OI module to be reusable, the ramping in HOPS is actually not carried out directly with Eq. A5.40. The procedure is fulfilled through inverting Eq. A5.38 for an effective observation error $\hat{E}_o$ with the new weight $\omega\alpha$, followed by an application of the original interpolation scheme, Eq. A5.37.

A5.5 Hydrostatic consistency factor

The hydrostatic consistency factor (HCF) is a quantity that measures the consistency for a hybrid or double sigma coordinate system when handling the computation of pressure gradient (see Lozano, 1995; Haney, 1991). It is defined as (Haley, 1995):

$$\mathrm{HCF} = \max\left(\frac{\sigma}{\Delta\sigma}\frac{\Delta x}{|H(x,y) - H_c|}\left|\frac{\partial h}{\partial x}\right|,\quad \frac{\sigma}{\Delta\sigma}\frac{\Delta y}{|H(x,y) - H_c|}\left|\frac{\partial h}{\partial y}\right|\right) \tag{A5.41}$$

for a hybrid system ($0 \leq \sigma \leq 1$), and

$$\mathrm{HCF} = \max\left(\left|\frac{(2-\sigma)f_x + (\sigma - 1)H_x}{H(x,y) - f(x,y)}\right|\frac{\Delta x}{\Delta\sigma},\quad \left|\frac{(2-\sigma)f_y + (\sigma - 1)H_y}{H(x,y) - f(x,y)}\right|\frac{\Delta y}{\Delta\sigma}\right) \tag{A5.42}$$

for a double sigma system ($1 \leq \sigma \leq 2$). In these equations, $H < 0$ is the topography, $H_c = \text{const}$ (hybrid system) or $f = f(x,y)$ (double sigma system) the interface depth, σ the scaled vertical coordinate, and Δx, Δy, and $\Delta\sigma$ are the horizontal and vertical grid spacings, respectively. A hydrostatically consistent coordinate system requires that $\mathrm{HCF} \leq 2$ (Lozano, 1995).

A5.6 Boundary relaxation

In HOPS, boundary relaxation is a nudging approach to stablize the numerical scheme. The following note is from some written material by Haley,[7] and private communications from him.

For a prognostic variable u (velocity or tracer), the boundary relaxation adds sponge layers to the near-boundary points. It appears on the right-hand side of the prognostic equation as a Newtonian damping term

$$-\frac{1}{\tau_s}e^{-\left(\frac{d}{d_s}\right)^2}(u-u_0),$$

where d is the distance to the nearest boundary, τ_s and d_s the temporal and spatial decay scales, respectively. This forcing alone gives an exponential decay of u to u_0

$$u = u_0 + Ce^{-\alpha t},$$

where

$$\alpha = \frac{1}{\tau_s}e^{-\left(\frac{d}{d_s}\right)^2}$$

and C a constant.

In the HOPS PE model, the relaxation is turned on only when both τ_s and d_s are positively evaluated (off if set zero). Besides, τ_s must be set above some critical value, τ_c, in order for the integration to be stable. For a simple system with both nudging and rotation,

$$\begin{aligned} \frac{du}{dt} - fv + \frac{1}{\tau_s}u &= 0 \\ \frac{dv}{dt} + fu + \frac{1}{\tau_s}v &= 0 \end{aligned}$$

this τ_c can be found analytically with von Neumann analysis, and it proves to be a function of differencing scheme, implicit Coriolis weighting factor,[8] a_{cor}, Coriolis parameter, f, and time step, Δt. Tab. A4.9 tabulates four expressions for τ_c's with different parameters. For the current PE model set-up, $a_{cor} = 0.5$ (semi-implicit treatment of the Coriolis term), $\Delta t = 180s$, $f \approx 1.318 \times 10^{-4}$, and Leap-Frog differencing scheme is used. The critial value τ_c is thus calculated to be 180 s. Our choice of τ_s, either for momentum or for tracers, greatly exceeds this value (Refer to Tab. 5.2).

[7] It is maintained together with the HOPS documentation by Patrick Haley, Jr., and available if requested.

[8] If $a_{cor} = 0$, the Coriolis term is treated explicitly; If $a_{cor} = 1$, a fully implicit scheme is used. The treatment lies in between when $0 < a_{cor} < 1$.

Table A4.9: Critical decay times τ_c for the boundary relaxation.

	Leap-Frog	Euler
$a_{cor} > 0$	$\frac{2\Delta t}{1+\sqrt{1+(2a_{cor}-1)(2\Delta t f)^2}}$	$\frac{\Delta t}{1+\sqrt{1+(2a_{cor}-1)(\Delta t f)^2}}$
$a_{cor} = 0$	$\frac{\Delta t}{1-\Delta t f}$	$\frac{\Delta t}{1+\sqrt{1-(\Delta t f)^2}}$

A5.7 The pressure computation

It is required that the pressure be given prior to any MS-EVA application. The current version of HOPS, however, does not have this field computed during the integration (cf., Lozano et al., 1994). We need to find a way to evaluate it based on other state variables.

From the hydrostatic assumption (3.6),

$$\frac{\partial P}{\partial z} = -\rho g,$$

the pressure

$$P = P(0) - g \int_0^z \rho \, dz \tag{A5.43}$$

can be written as a sum of a *barotropic* part $P_{BT} = P(0)$ and a *baroclinic* part $P_{BC} = -g \int_0^z \rho \, dz$. P_{BC} is easy to compute as the density anomaly ρ is already known (ρ is understood to have been removed by a constant vertical profile $\bar{\rho}(z)$. See §3.1). The problem here is how to find P_{BT}, the barotropic pressure.

Observe that P_{BT} doesn't have z-dependence. We then only need to consider the vertically integrated equations. Look at the continuity equation (3.5) first. It becomes, if integrated from the bottom $z = -H$ to the (rigid) surface $z = 0$,

$$\frac{\partial}{\partial x} \int_{-H}^0 u \, dz + \frac{\partial}{\partial y} \int_{-H}^0 v \, dz = 0. \tag{A5.44}$$

Eq. (A5.44) allows us to define a transport streamfunction Ψ, which is such that,

$$\int_{-H}^0 \underline{\mathbf{v}} \, dz = \underline{\mathbf{k}} \wedge \nabla \Psi. \tag{A5.45}$$

In the HOPS output, Ψ is an evaluated field.

The momentum equation is treated similarly. Rewrite (3.4) as

$$\frac{\partial \underline{\mathbf{v}}}{\partial t} + \nabla \cdot (\underline{\mathbf{v}\mathbf{v}}) + \frac{\partial w\underline{\mathbf{v}}}{\partial z} + f\underline{\mathbf{k}} \wedge \underline{\mathbf{v}} = -\frac{1}{\rho_0}\nabla P + \frac{\partial}{\partial z}\left(\gamma \frac{\partial \underline{\mathbf{v}}}{\partial z}\right) + \underline{\mathbf{F}}_h, \tag{A5.46}$$

where γ is the vertical kinematic viscosity coefficient, and the dot product of $\nabla \cdot \underline{\mathbf{v}\mathbf{v}}$ is understood to apply only on the the first vector of the dyad $\underline{\mathbf{v}\mathbf{v}}$. Integrating both sides with respect to z from $-H$ to 0, and substituting $\underline{\mathbf{k}} \wedge \nabla\Psi$ for $\int_{-H}^{0} \underline{\mathbf{v}}\, dz$ according to (A5.45), we get

$$\begin{aligned} \underline{\mathbf{k}} \wedge \nabla \frac{\partial \Psi}{\partial t} + \int_{-H}^{0} \nabla \cdot (\underline{\mathbf{v}\mathbf{v}})\, dz + (w\underline{\mathbf{v}})|_{-H}^{0} - f\nabla\Psi = \\ -\frac{H}{\rho_0}\nabla P_{BT} - \frac{1}{\rho_0}\int_{-H}^{0} \nabla P_{BC}\, dz \\ +\frac{\underline{\tau}_s - \underline{\tau}_b}{\rho_0} + \int_{-H}^{0} \underline{\mathbf{F}}_h\, dz. \end{aligned} \tag{A5.47}$$

($\underline{\tau}_s$ and $\underline{\tau}_b$ are stresses at the surface and the bottom, respectively) In order to avoid the cumbersome computation of the horizontal gradient in a σ-coordinate grid, use the Leibniz rule, which here implies

$$\int_{-H}^{0} \nabla\phi\, dz = \nabla \int_{-H}^{0} \phi\, dz - \phi(-H)\nabla H, \qquad \text{for all } \phi \text{ with dependency of z,}$$

to integrate it into

$$\begin{aligned} \underline{\mathbf{k}} \wedge \nabla \frac{\partial \Psi}{\partial t} + \nabla \cdot \int_{-H}^{0} \underline{\mathbf{v}\mathbf{v}}\, dz - \nabla H \cdot (\underline{\mathbf{v}\mathbf{v}})|_{z=-H} + (w\underline{\mathbf{v}})|_{-H}^{0} - f\nabla\Psi \\ = -\frac{H}{\rho_0}\nabla P_{BT} - \frac{1}{\rho_0}\left[\nabla \int_{-H}^{0} P_{BC}\, dz - P_{BC}|_{z=-H}\ \nabla H\right] \\ +\frac{\underline{\tau}_s - \underline{\tau}_b}{\rho_0} + \int_{-H}^{0} \underline{\mathbf{F}}_h\, dz. \end{aligned} \tag{A5.48}$$

In a HOPS output, Ψ, $\underline{\mathbf{v}}$, and w are known, P_{BC} can be computed from (A5.43), and the bottom friction stress is parameterized as

$$\underline{\tau}_b = \rho_0 C_B |\underline{\mathbf{v}}_b| \underline{\mathbf{v}}_b \tag{A5.49}$$

($\underline{\mathbf{v}}_b = \underline{\mathbf{v}}|_{z=-H}$, and C_B some prescribed constant) while $\underline{\tau}_s$ is pre-specified by the user. So all the terms in (A5.48) other than $-\frac{H}{\rho_0}\nabla P_{BT}$ are easy to calculate except $\int_{-H}^{0} \underline{\mathbf{F}}_h\, dz$, the horizontal dissipation. Since we use filtering instead of an explicit formula for its parameterization, this term is hard to evaluate directly. Nonetheless, it is generally believed that the horizontal dissipation is small in comparison to the vertical friction, while here the latter, by our result, is at least two orders smaller than the Coriolis term and the baroclinic pressure gradient. Notice that we are not repeating the HOPS scheme to make the computation (actually there is no pressure computation in the HOPS algorithm), and the difference in scheme will generally give rise to discrepancy in results. This discrepancy, though not significant if controlled, may be larger in magnitude than that of $\int_{-H}^{0} \underline{\mathbf{F}}_h dz$. It is therefore meaningless to fuss with an insignificant term like the horizontal dissipation, and we simply drop it out from the momentum equation. The resulting (A5.48) now involves only one unknown. Put the term with it on one side, and all else on the other. The equation becomes

$$\begin{aligned} \nabla P_{BT} &= -\frac{\rho_0}{H}\left[\underline{\mathbf{k}} \wedge \nabla \frac{\partial \Psi}{\partial t} + \nabla \cdot \int_{-H}^{0} \underline{\mathbf{v}}\underline{\mathbf{v}} \, dz \right. \\ &\quad \left. -\nabla H \cdot (\underline{\mathbf{v}}\underline{\mathbf{v}})|_{z=-H} + (w\underline{\mathbf{v}})|_{-H}^{0} \right] + \frac{f\rho_0}{H}\nabla\Psi \\ &\quad -\frac{1}{H}\left[\nabla \int_{-H}^{0} P_{BC} \, dz - P_{BC}|_{z=-H} \, \nabla H\right] \\ &\quad +\frac{\underline{\tau}_s - \underline{\tau}_b}{H} \\ &\equiv \underline{\mathbf{J}}, \end{aligned} \tag{A5.50}$$

where $\underline{\mathbf{J}} = (J_x, J_y)$ is totally determined by the HOPS output.

Taking divergence on both sides of Eq. (A5.50), we obtain a Poisson equation

$$\nabla^2 P_{BT} = \nabla \cdot \underline{\mathbf{J}} \equiv \mathcal{D}, \tag{A5.51}$$

which, together with conditions

$$\frac{\partial P_{BT}}{\partial x} = J_x, \tag{A5.52}$$

$$\frac{\partial P_{BT}}{\partial y} = J_y, \tag{A5.53}$$

specified respectively at the meridional and zonal boundaries, forms a Neumann problem for P_{BT}. It has a solution if and only if the flux of $\underline{\mathbf{J}}$ across the boundaries vanishes, i.e., $\oint_S \underline{\mathbf{J}} \cdot \underline{\mathbf{n}} dS = 0$. By our computation, this condition is satisfied to a tolerance of error. So this Neumann problem is well defined. But the solution is underdetermined. Another condition, which corresponds to the arbitrary additive constant for the barotropic pressure field, must be supplemented. Here we simply choose

$$P_{BT}(0,0) = 0. \tag{A5.54}$$

Eq. (A5.51), together with boundary conditions (A5.52) - (A5.54), can be solved easily with the available softwares (e.g., Press et al., 1992). The obtained P_{BT} combined with P_{BC} makes the total pressure P, the very field needed in the MS-EVA application. A computed snapshot of P_{BT} and P is shown in Fig. A5.33 .

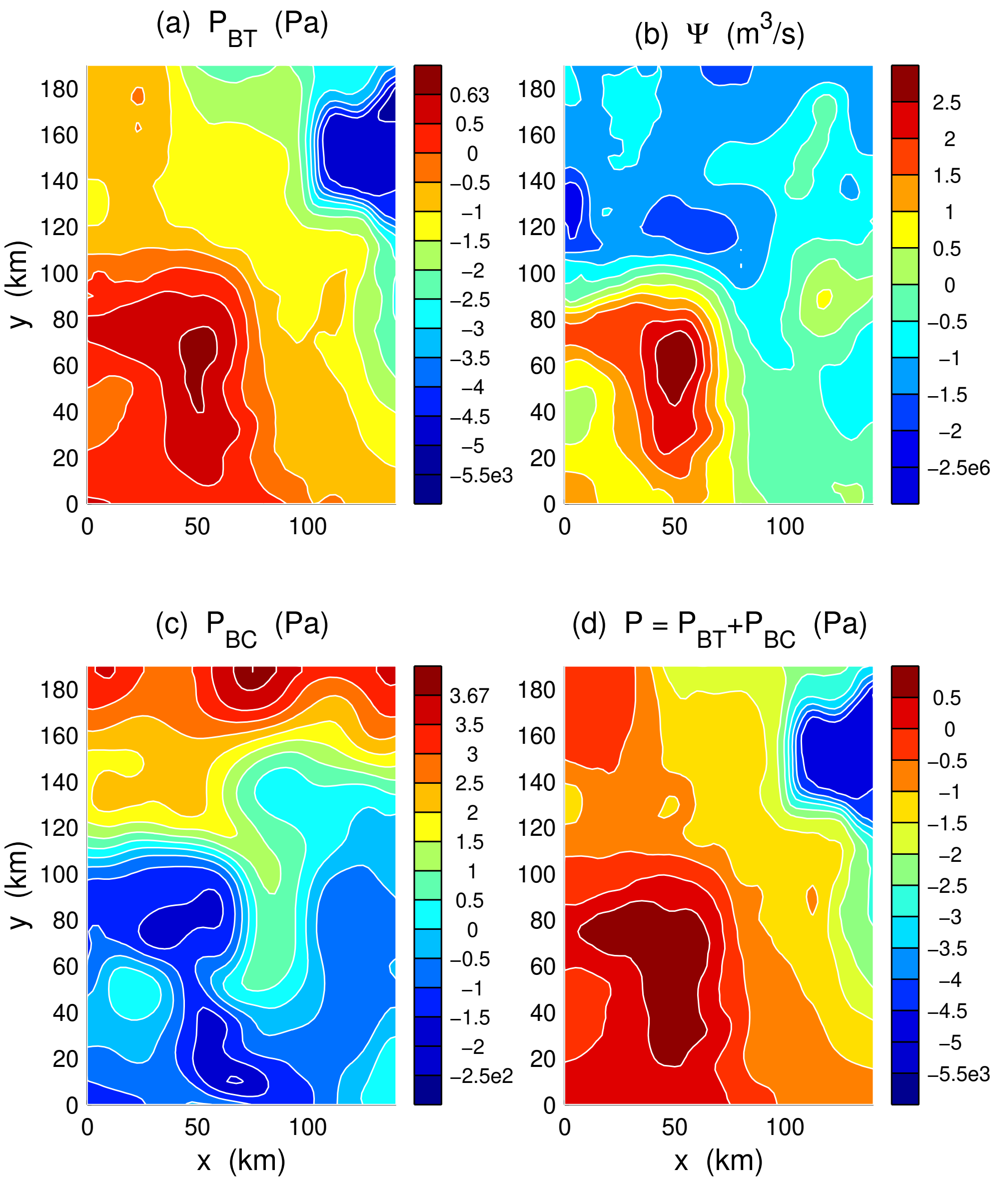

Figure A5.33: The computed barotropic pressure (a) for the eighth day of the IFF variability simulation. Also shown are the transport streamfunction (b), the baroclinic pressure at level 5 (c), and the total pressure at level 5 (d) for the same day.

Bibliography

[1] Allen, J. T., D. A. Smeed, and A. L. Chadwick, 1994: Eddies and mixing at the Iceland-Faeroes Front. *Deep-Sea Res.*, 41, No.1, 51-79.

[2] Arakawa, A., and V. R. Lamb, 1977: Computational design of the basic dynamical processes of the UCLA general circulation model. *Methods in Comp. Phys.*, 174-265.

[3] Aubry, N. et al, 1988: Dynamics of coherent structures in the wall region of a turbulence boundary layer. *J. Fluid Mech.*, 192, 115-173.

[4] Blindheim, J., 1990: Arctic Intermediate Waters in the Norwegian Sea. *Deep-Sea Res.*, 20, 1-15.

[5] Bretherton, Francis, Russ Davis, and C. B. Fandry, 1976: A technique for objective analysis and design of oceanographic experiments applied to MODE-73. *Deep-Sea Res.*, 23, 559-582.

[6] Brockett, R. W., 1970: *Finite Dimensional Linear Systems.* John Wiley and Sons, Inc., 239 pp.

[7] Brockett, Roger, 1996: Estimation and Control of Dynamic Systems, Engineering Sciences 202, Harvard University, Cambridge, MA 02138.

[8] Brockett, Roger, 1998: Stochastic Control Course Notes, Engineering Sciences 203, Harvard University, Cambridge, MA 02138.

[9] Chapman, D. C., 1985: Numerical treatment of cross-shelf open boundaries in barotropic coastal ocean model. *J. Phys. Oceanogr.*, 15, 1060-1075.

[10] Dombrowsky, Eric, and Pierre de Mey, 1992: Continuous Assimilation in an Open Domain of the Northeast Atlantic. 1. Methodology and Application to Athen A-88. *J. Geophys. Res.*, 97, No. C6, 9719-9731.

[11] Fuller, Wayne A., 1976: *Introduction to Statistical Time Series.* Wiley. 470 pp.

[12] Gandin, L. S., 1965: *Objective Analysis of Meteorological Fields.* Israel Program for Scientific Translations.

[13] Gonella, J., 1972: A rotary-component method for analyzing meteorological and oceanographic vector time series. *Deep-Sea Res.*, 19, 833-846.

[14] Gould, W. J., J. F. Read, and J. Smithers, 1987: SEASOAR profiles in the Iceland-Scotland area, May 1987: Rep. 253, Inst. of Oceanogr. Sci., Wormley, England, 50pp.

[15] Gut, 1998: *An Intermediate Course in Probability.* Springer-Verlag. 278pp.

[16] Hansen, B., and J. Meincke, 1979: Eddies and meanders in the Iceland-Faroe Ridge area. *Deep-Sea Res.*, 26A, 1067-1082.

[17] Hallock, Z. R., 1985: Variability of frontal structure in the Southern Norwegian Sea. *J. Phys. Oceanogr.*, 15, 1245-1254.

[18] Haley, Patrick J., Jr., 1995: *GRIDS User's Guide* (Ver. 3.2) HOPS documentation. Harvard University Ocean Group.

[19] Haney, R. L., 1991: On the pressure gradient force over steep topography in sigma coordinate ocean models. *J. Phys. Oceanogr.*, 21, 610-619.

[20] Holton, James R., 1992: *An Introduction to Dynamic Meteorology.* Academic Press. 496pp.

[21] Hopkins, T., 1988: The GIN Sea: Review of physical oceanography and literature from 1972. Saclant Undersea Research Center Rep. No. SR-124, 190pp.

[22] Kreyszig, Erwin, 1989: *Functional Analysis and its Applications.* John Wiely & Sons.

[23] Knudsen, M., 1899: Hydrography. *The Danish Ingolf-Expedition*, 1(2), 23-161.

[24] Lermusiaux, P. J., 1997: *Error Subspace Data Assimilation Methods for Ocean Field Estimation: Theory, Validation and Applications.* Ph.D. Thesis, Division of Engineering and Applied Sciences, Harvard University

[25] Lozano, C. J., P. J. Haley, Jr., H. G. Arango, N. Q. Sloan, and A. R. Robinson, 1994: Harvard coastal/deep water primitive equation model. Reports in Meteorology and Oceanography: Harvard Open Ocean Model Reports, 15 pp. Divsion of Engineering and Applied Sciences, Pierce Hall, Harvard University, Cambridge, MA 02138.

[26] Lozano, C. J., P. Haley, Jr., H. Arango, Q. Sloan, and A. R. Robinson, 1996: An interdiciplinary ocean prediction system: Assimilation strategies and structured data models. Reports in Meteorology and Oceanography: Harvard Open Ocean Model Reports, Division of Applied Sciences, Harvard University

[27] Meincke, J., 1978: On the distribution of low salinity intermediate waters around the Faeroes. *Dtsch. Hydrogr. Z.*, 31, 50-65.

[28] Miller, A. J., et al., 1995a: Quasigeostrophic forecasting and physical processes of Iceland-Faeroe Frontal variability. *J. Phys. Oceanogr.*, 25, 1273-1295.

[29] Miller, A. J., P.-M. Poulain, A. R. Robinson, H. G. Arango, W. G. Leslie, and A. Warn-Varnas, 1995b: Quantitative skill of quasi-geostrophic forecasts of a baroclinically unstable Iceland-Faroe front. *J. Geophys. Res.*, 100, 10 833 - 10 849.

[30] Miller, A. J., and B. D. Cornuelle, 1999: Forecasts from fits of frontal fluctuations. *Dyn. Atmos. Oceans*, Special issue in honor of Prof. A. R. Robinson.

[31] Müller, P, and G. Siedler, 1976: Consistency relations for internal waves. *Deep-Sea Res.*, 23, 613-628.

[32] Müller, T. J., J. Meincke, and G. A. Becker, 1979: The distribution of water masses on the Greenland-Scotland Ridge in August/September 1973. *Ber. Inst. Meereskde*, 62, 1-172.

[33] Niiler, P. P., S. Piacsek, L. Neuberg, and A. Warn-Varnas, 1992: Sea surface temperature variability of the Iceland-Faeroe Front. *J. Geophys. Res.*, 97, 17777-17785.

[34] Orlanski, I., 1976: A simple boundary condition for unbounded hyperbolic flows. *J. Comput. Phys.*, 41, 251-269.

[35] Preisendorfer, R., 1998: *Principal Component Analysis in Meteorology and Oceanography.* Elsevier. 418 pp.

[36] Press, W. H., S. A. Teukolsky, W. T. Vertterling, and B. P. Flannery, 1992: *Numerical Recipes in C - The Art of Scientific Computing.* 2nd Ed., Cambridge University Press. 965 pp.

[37] Read, J. F., and R. T. Pollad, 1992: Water masses in the Region of the Iceland-Faeroes Front. *J. Phys. Oceanogr.*, 22, 1365-1378.

[38] Robinson, A. R., Ed., 1983: *Eddies in Marine Science*, Springer-Verlag, 609 pp.

[39] Robinson, A. R., 1992: Shipboard prediction with a regional forecast model. *Oceanography*, 5, 42-48.

[40] Robinson, A. R., H. G. Arango, A. J. Miller, A. Warn-Varnas, P.-M. Poulain, and W. G. Leslie, 1996: Real-time operational forecasting on shipboard of the Iceland-Faeroe Frontal variability. *Bulletin of the American Meterological Society*, 243-259.

[41] Robinson, A. R., M. A. Spall, and N. Pinardi, 1988: Gulf Stream simulations and the dynamics of ring and meander processes. *J. Phys. Oceanogr.*, 18, 1811-1853.

[42] Scott, J. C., and N. M. Lane, 1990: Frontal boundaries and eddies on the Iceland-Faeroes Ridge. In: *Ocean Variability and Acoustic Propagation*, J. Potter and A. Warn-Varnas (eds.), 449-461.

[43] Shapiro, R., 1971: The use of linear filtering as a parameterization of atmospheric diffusion. *J. Atmos. Sci.*, 28, 523-531.

[44] Sloan, N. Q., 1996: *Dynamics of a Shelf-Slope Front: Process Studies and Data-Driven Simulations.* Division of Applied Sciences. Ph.D. Thesis, Harvard University.

[45] Smart, J. H., 1984: Spatial variability of major frontal systems in the North Atlantic-Norwegian Sea area: 1980-81. *J. Phys. Oceanogr.*, 14, 185-192.

[46] von Storch, Hans, Claude Frankignoul, 1998: Empirical Modal Decomposition in Coastal Oceanography. In *The Sea*, Vol. 10, K. H. Brink and A. R. Robinson, eds., Jonh Wiley & Sons, Inc., 419-455.

[47] Willebrand, J., and J. Meincke, 1980: Statistical analysis of fluctuations in the Iceland-Scotland frontal zone. *Deep-Sea Res.*, 27A, 1047-1066.

[48] Willems, R. C., 1994: Experiment evaluates ocean models and data assimilation in the Gulf Stream. *Trans. Amer. Geophys. Union*, 75, 34.

Chapter 6

Application of MS-EVA to the IFF variability study

6.1 Introduction

In this chapter, we will perform an MS-EVA diagnosis for the IFF variability forecast conducted previously. The successfully validated MS-EVA, together with the conspicuous meandering event captured on August 22, 1993, is expected to shed light on the physics which is characteristic of the complex frontal processes in the IFF region.

The dynamical study of the IFF variability has long been of interest among physical oceanographers. The results to date, in summary, show that the variability could be either driven by external forcing or due to some intrinsic reason. For the former, the existence of the Iceland-Faeroe Ridge was once of concern (e.g. Maskell et al., 1992) but its role is still to be investigated. The fluctuation of atmospheric pressure or wind stress has also been speculated to be a cause, based on the observation that intermittent events generally happen on two time scales, 10 days and 2-5 days, which have apparent correspondences in the atmospheric motion (see Willebrand and Meincke, 1980, and the references therein). In 1975, this speculation motivated a field experiment, the MONA project (Monitoring Overflow in the North Atlantic). Ironically, the coherence between the atmospheric fluctuation and the current is not as significant as expected. The available potential energy conversion calculated by Willebrand

and Meincke is at least one order of magnitude larger than that from the wind effect. This fact is also seen in the analysis of Allen et al. (1994), which is based on the data set collected from *Charles Darwin* Cruise 51, and is consistent with their quantitative estimation with the simple two-layer model of Killworth et al. (1984). The intermittent events thus seem to be forced intrinsically by frontal instabilities, particularly baroclinic instabilities by Willebrand and Meincke (1980), rather than by external forcing.

The intrinsic mechanism assertion has been further supported by early EVA investigations. In 1995, Miller et al. made the first real-time forecast of the IFF variability with a quasi-geostrophic (QG) model of the Harvard Ocean Prediction System (HOPS), driven by data acquired during the 1992 SACLANTCEN-Harvard survey (19-26 October). The forecast output was then diagnosed with the QG-EVA developed by Pinardi and Robinson (1986). They found, from their EVA maps, that there is clearly a significant energy conversion, and at mid-depths the conversion is from APE to KE. The converted energy is removed by the vertical pressure work, forcing the motion at the top and bottom boundaries. The whole scenario is in nice agreement with the process identified in an unstable Eady model, which, as they argue, must be an expression of baroclinic instability. In contrast, the Reynolds stress effect, which extracts energy from the mean velocity shear, is found to be inconsequential in their case.

Recall that our purpose of this chapter is to understand the dynamics of the IFF variability. Particularly, we want to know what gives rise to the deep-sock meandering intrusion observed in the PE forecast conducted before. The previous results, though supported by different diagnoses, are problem-specific in themselves. We are not sure at this moment whether they are applicable to our study or not. Besides, it is questionable whether the process identifiers used before can faithfully represent the IFF variability, which occurs on multiple scales and is highly localized in nature. These concerns motivate us to diagnose the forecast output with our newly developed MS-EVA, in the hope of gaining an understanding of the observed event in an objective and relatively quantitative approach. The diagnosis begins with a scale analysis of the IFF variability, and the results are used to determine the time and scale window bounds needed for the process decomposition. The application of MS-EVA is set up in §6.3.1, which is followed by a description of results that are related to the deep-sock meandering. Processes are identified, analyzed, and further verified in this section. In §6.4, we test what we have obtained against parametric change. This chapter is concluded in §6.5.

6.2 IFF variability and scale window bound determination

6.2.1 Time scales

Time variability determines time scale windowing. We study the IFF time variability through wavelet spectral analysis. By a wavelet spectrum (time), or simply spectrum, we mean the "energy"[1] of a time series unfolded on the time-scale level plane. To perform the spectral analysis, fifteen series have been extracted from five points within the model domain. These points, labeled as 1 through 5 on the temperature map of Fig. 6.1, are located at places representative of the frontal activities under study. In terms of the series extracted, point 3 is to a large extent similar to point 4, for all depths available in the water column, while point 1 bears some resemblance to point 2. (But features are not as significant as point 2, probably due to the proximity to the western open boundary.) It is thus enough to show the results for points 2, 4, and 5. These signals are 1024 in length (j_2 is hence 10), with a time interval $\Delta t = 1800$ s. (They are obtained by sampling the filtered forecast output. See §6.3.1.) The whole time span is thence 1024×0.5 hour $= 21.3$ day, from day 0 to day 21.3. Recall the deep-sock meander is observed on day 8. We limit our discussion to the period from day 2 through day 10. The initialization stage (before day 2) and the period after day 10 are not under consideration.

Look at the temperature signals first. They are similar to the density evolution (not shown) for these locations on the wavelet spectrum, but with features much enhanced. Graphed in Fig. 6.2 are the natural logarithm of energy as a function of time and scale level, for point 2 at depths 25 m, 125 m, and 300 m, which are typical of the upper, middle, and deep layers as named in the preceding chapter. We plot the logarithm of energy instead of energy itself in order to distinguish locations of interest and make feature identification easier. For comparison, attached to each spectrum is the corresponding time series.

From Fig. 6.2, the scale windowing is apparent. A gap exists between the duration scale

[1]Here the term "energy" refers to the square of the wavelet transform coefficients with the basis built in Chapter 2. The factor 2^{j_2} is not multiplied for simplicity.

and smaller scales. The three spikes, which are located at days 2, 4 and 5, mark the instances when data are assimilated. (Recall the zigzag data assimilation is ramped up from day 4 to day 5.) Clearly, the data assimilation affects processes with scale level higher than or equal to 5 (approximately $21.3/2^5 \approx 0.7$ day), but its influence does not have significance outside the sub-mesoscale window. For the series at 25 m, there is an obvious peak at $j = 2$ from beginning to day 10 and beyond, which corresponds approximately to a scale of $21.3/2^2 = 5.3$ days. This peak is seen on the spectra for all the depths in the water column, albeit a little blurred for deep-layer signals (e.g., the 300-m level). On about day 5.3, a new scale emerges on the spectrum in the upper and middle layers, which has a scale level $j = 3$ ($\sim$2.7 days). Correspondingly the magnitude on scale level $j = 2$ decreases, implying some energy transfer at this juncture. This phenomenon is also seen in Fig. 6.3, the spectra for point 4, and is particularly significant at the 300-m depth. In that case, the peak shifts from $j = 2$ (5.3 days) to $j = 3$ (2.7 days), then dominates the spectrum until day 10.7, though after day 8 it clearly becomes weakened. The whole scenario is reminiscent of the meandering event we captured in the forecast, and the identified scales (2.7 days-5.3 days) are in agreement with the estimates by Willebrand and Meincke (1980) and Hansen and Meincke (1979) (2-5 days), but a little larger than Allen's observation (2-4 days).

While the spectrum for temperature directs attention to potential energy transfer, the spectrum for velocity may unravel the multiscale kinetic energy information. Figs. 6.4-6.6 show a selection of spectra for the fields u and v at points 2, 4, and 5, respectively. A conspicuous feature on these maps is the maxima of the variability (above the duration scale level) at j=1 and j=3 (or sometimes j=4), which corresponds to a scale of 10.7 days, and 2.7 days (or 1.3 days). (See Figs. 6.4a,b, 6.5a,b,c, and 6.6a,c. Similar phenomena can also be found on the temperature spectrum Fig. 6.3b.) This double peak structure has been documented in literature. For example, Willebrand and Meincke (1980) identified a 2-5-day event and a 10-day fluctuation, and Hansen and Meincke (1979) found that the eddies vary on a scale of 2-5 days and 8-11 days. Their results agree with our estimation very well.

Different from that of temperature, the velocity field is very energetic in the sub-mesoscale window. This feature is especially conspicuous on the spectrum of v at point 5 (Fig. 6.6), where an apparent peak exist at $j = 6$ (0.33 day) virtually all the time within the duration.

The corresponding one-third-day process is also evident on the time series attached below.

Another difference between temperature and velocity is that, in a velocity spectrum, the energy maximum at $j = 3$ generally exists before day 5.3. It is not a new emergence at that time, though it does get enhanced during the meandering period, as is shown in Figs. 6.4a,b, Figs. 6.5a,b, and Figs. 6.6b,c. The new emergence happens at depth 300 m, for u at point 4 (Fig. 6.4c) and v at point 2 (Fig. 6.5c). In Fig. 6.4c, a significant part of energy goes to $j = 4$ during day 6.7-9.3, making it an apparent peak. This process spans right over the intrusion event, lasting for about 2.7 days, but it is much weakened after day 8, the day when the meandering intrusion is observed.

Consequently, the time variability of the IFF observed in the 1993 cruise occurs mainly on two scales: One is of 10.3 days, and the other includes a range from 1.3 days to 5.3 days. While the former process is more or less uniformly distributed over the event span in the region of concern, the latter generally varies from time to time and from location to location. During day 2 through day 10, a peak has been identified at $j = 2$ (5.3 days) and $j = 3$ (2.7 days), and sometimes $j = 4$ (1.3 days), on the spectra for both temperature and velocity. In general, within this range the variability is characterized by a combination of processes on all the three scales.

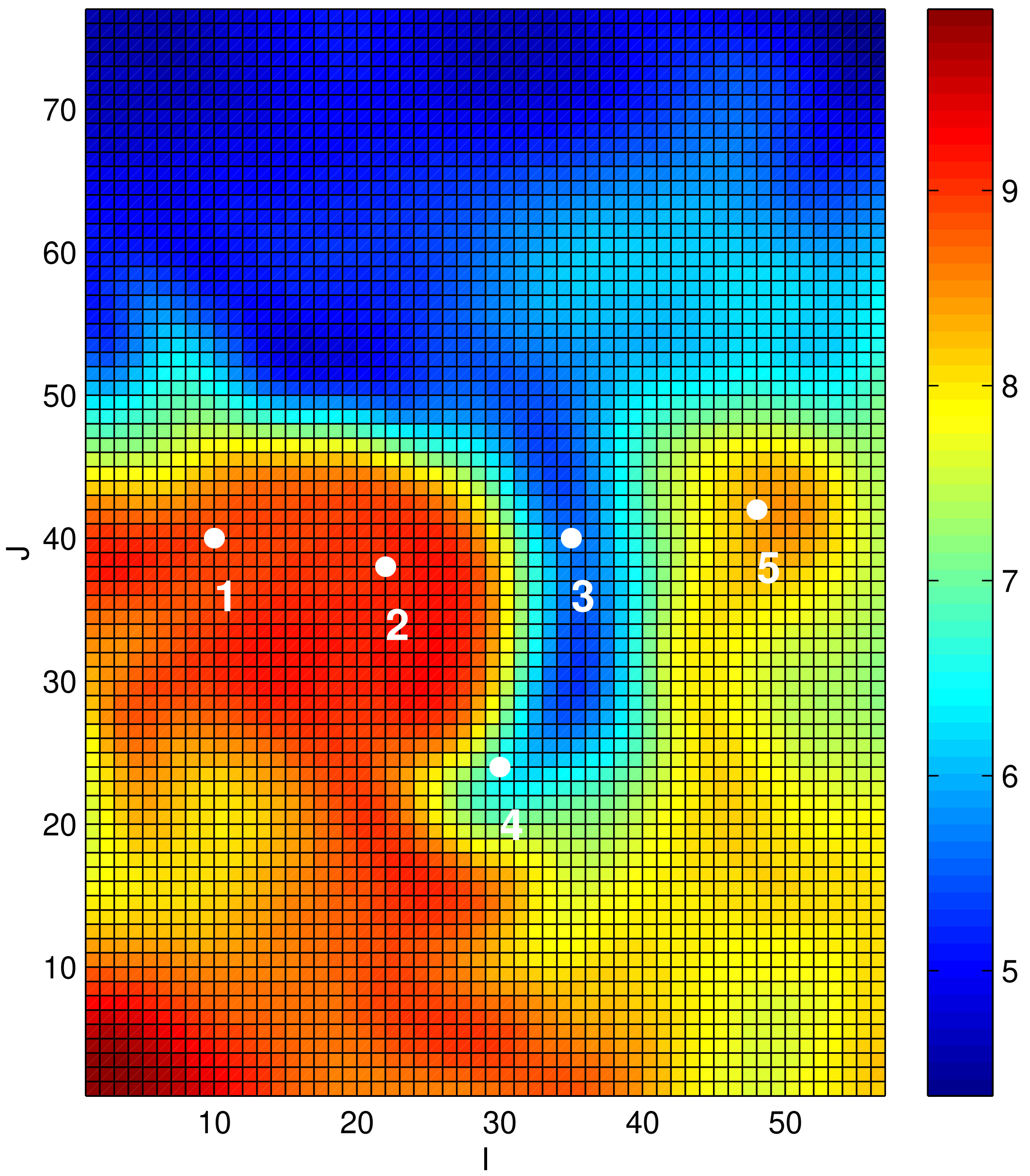

Figure 6.1: Locations of the point time series for spectral analysis overlaid on the forecast temperature (in ^{o}C) map for day 8 at depth 25 m. The coordinates are grid indices I and J, and in (I, J) pair, these points are located at

$$\mathrm{point1}:(10,40),\quad \mathrm{point2}:(22,38),\quad \mathrm{point3}:(35,40),\quad \mathrm{point4}:(30,24),\quad \mathrm{point5}:(48,42).$$

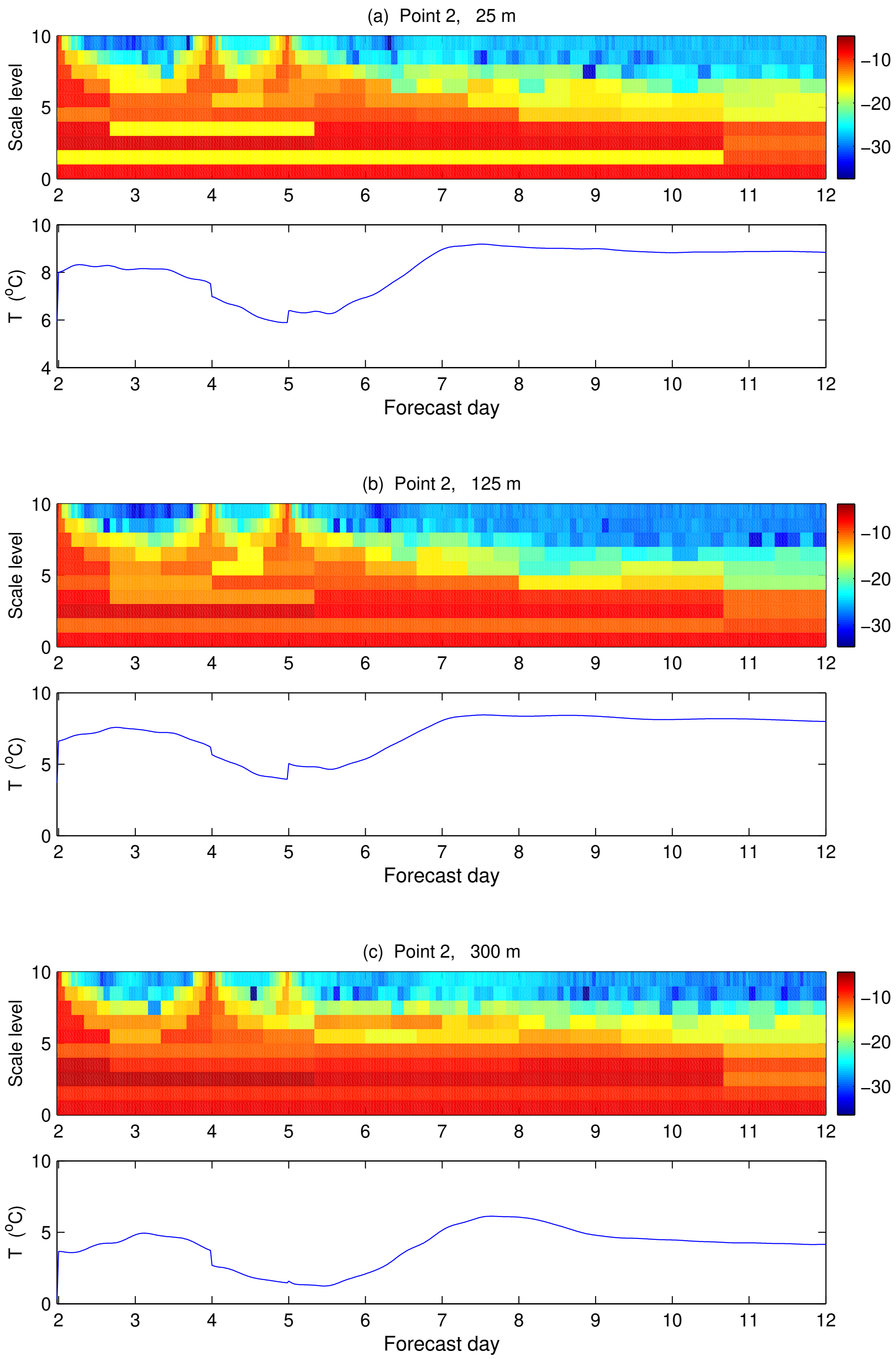

Figure 6.2: Time spectrums for the temperature series at point 2 for depths 25 m (top), 125 m (middle), and 300 m (bottom). The independent variable is the logarithm of energy instead of energy itself. Attached to each spectrum is the corresponding time signal. Notice the three spikes arising from data assimilation on days 2, 4, and 5.

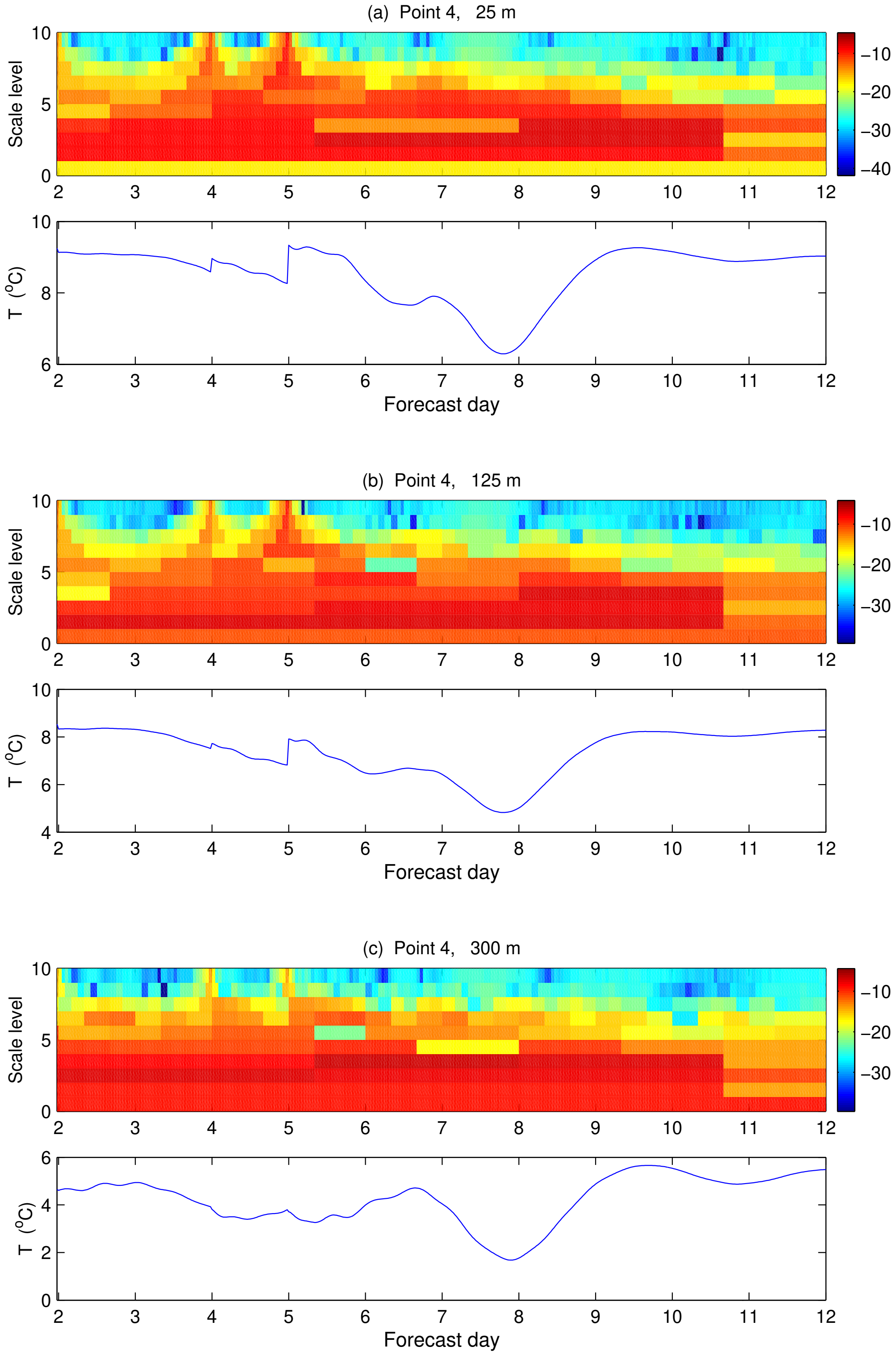

Figure 6.3: Same as Fig. 6.2, except for point 4.

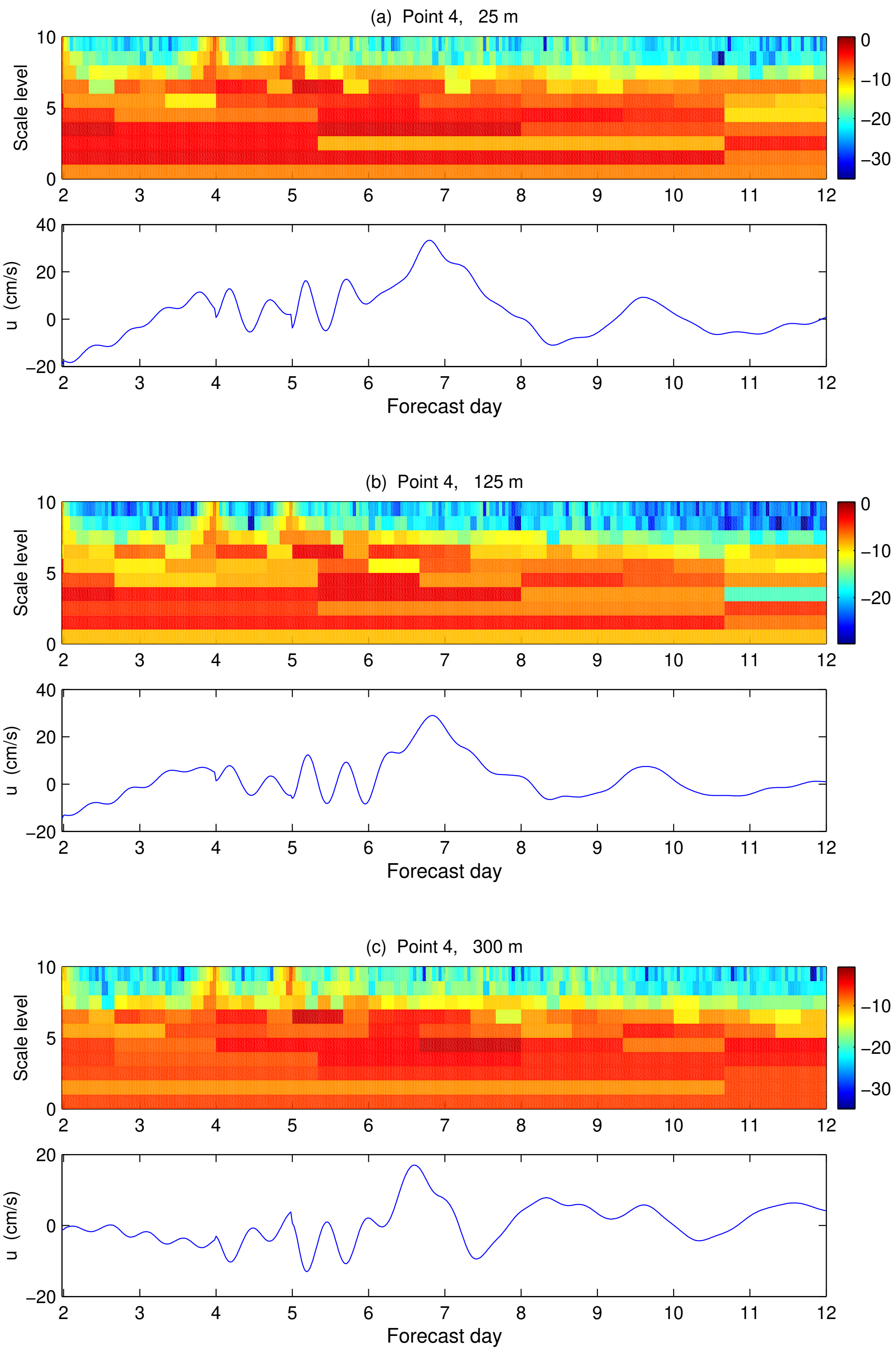

Figure 6.4: Same as Fig. 6.3, except for u.

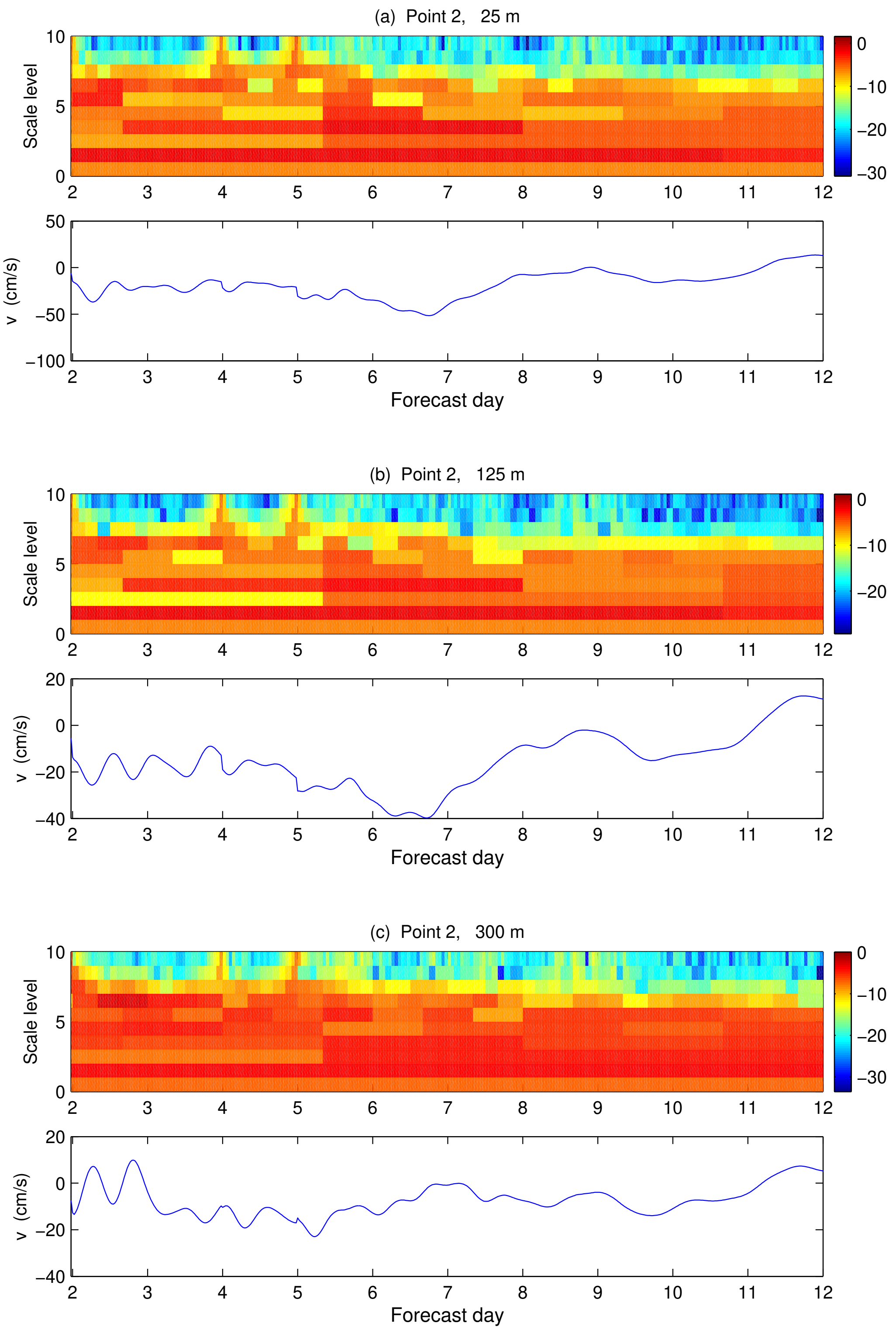

Figure 6.5: Same as Fig. 6.2, except for v.

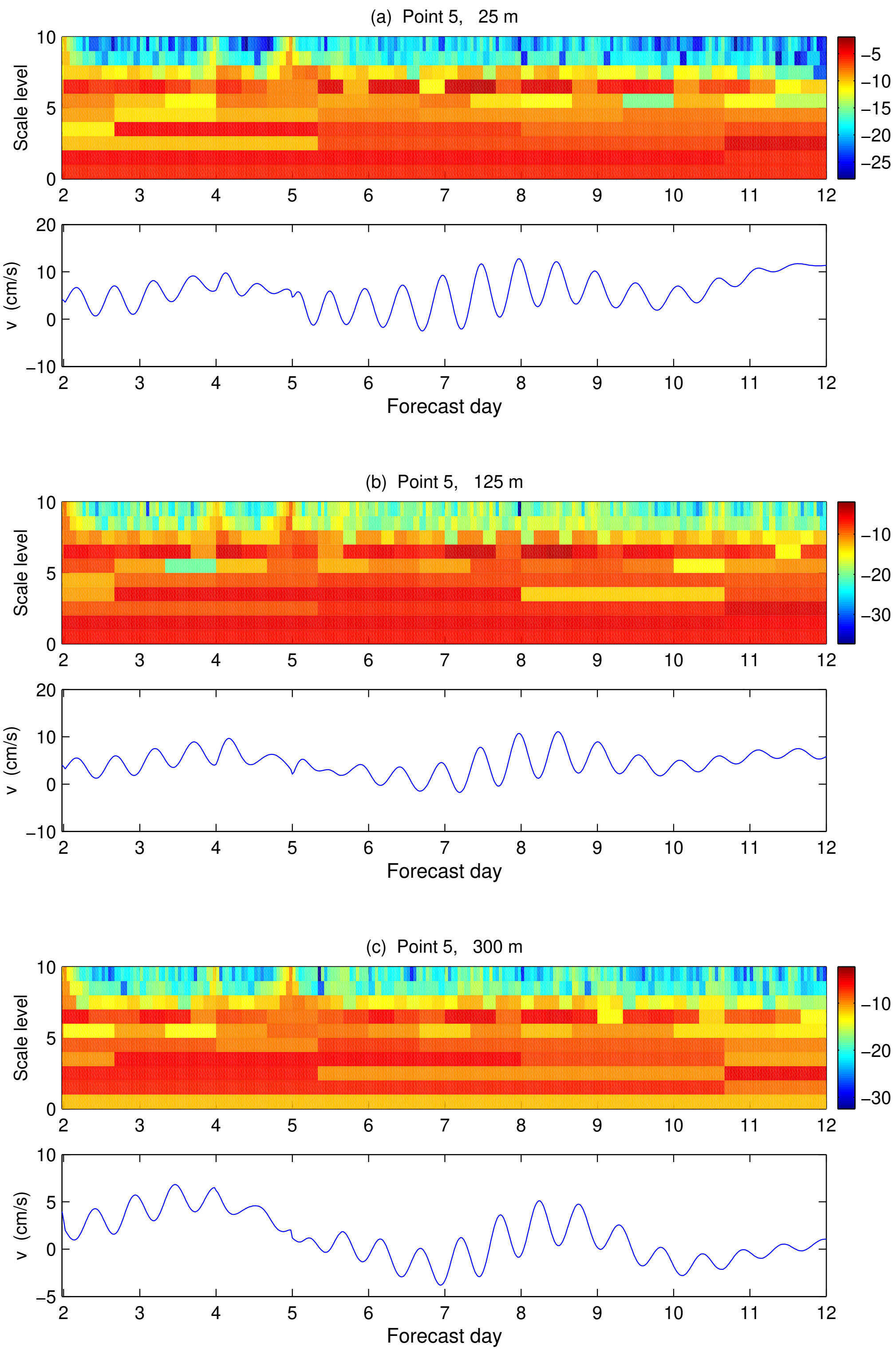

Figure 6.6: Same as Fig. 6.5, except for point 5.

6.2.2 Spatial scales

The spatial spectrum differs from its temporal counterpart in that more than just one dimension needs to be considered. Here we use the 2-D basis developed in §2.7 to examine the horizontal variability of the Iceland-Faeroe Front. We continue to use j to denote the scale level. But in order to avoid confusion with the usage adopted previously for the time scale analysis, a superscript "sp" is added so as to make it appear like j^{sp}. Remember that the model domain is a 140 km $\times$ 190 km rectangle which is discretized into 56×76 meshes (57×77 grid lines). We first interpolate the data on a $64{\times}64$ square, which implies a maximal scale level $j_2^{\mathrm{sp}} = 12$ (because $2^{12/2} = 64$), then re-map the results back to the original grid. Given a scale level j^{sp}, the scale then ranges from $140/\sqrt{2^{j^{\mathrm{sp}}}}$ km to $190/\sqrt{2^{j^{\mathrm{sp}}}}$ km. For simplicity, we pick the average, $165/\sqrt{2^{j^{\mathrm{sp}}}}$ km, for our description. Recall the wavelet transform has a resolution problem. In order for the energy obtained to be comparable point by point on the horizontal plane, we replace the the wavelet transform for level j^{sp} by the multiscale window transform with bounds j^{sp} and $j^{\mathrm{sp}}+1$ (cf., the 2-D multi-resolution analysis). As in the preceding subsection, it is the natural logarithm of energy instead of energy itself that is graphed for the spectrum.

We examine the variability for the eighth forecast day only. The deep-sock meander has been fully developed as of that time. Shown in Fig. 6.7 is the spectrum for the forecast temperature at depth 125 m. A first observation is that the variability is concentrated in an area in the proximity of the front. Away from the axis (>40-50 km), the basin scale ($j^{\mathrm{sp}} = 0$) dominates the spectrum. A second observation is that significantly energetic spots can still be identified at a scale level as high as $j^{\mathrm{sp}} = 8$ ($\sim$10 km). What we are interested lies in between these two extremes. By examining the maps for every scale level, there is an apparent peak at $j^{\mathrm{sp}} = 3$ in the west (I=1-25, J=10-50), which corresponds to a scale of 60 km, and another peak at j^{sp}=4 ($\sim$40 km) in the east (I=20-50, J=15-45). More peaks can be identified at higher levels. Around point (20,40), for instance, there is one at j^{sp}=6 (20 km). At the same scale level, local maxima, though not globally maximized in energy, can also be identified in regions in the east.

The temperature variability therefore occurs generally on two spatial scales, though these scales may vary from location to location. To see this scenario more clearly, the logarithm of energy is unfolded on the I-$j^{\rm sp}$ plane for the zonal section $J = 35$, and the plot is drawn beneath the spectrum maps of Fig. 6.7. On the section map, above the basin scale level is the variability. The maximum at $j^{\rm sp}$=3-4 is clearly seen from I=10-27, and 35-27. A second maximum exists at $j^{\rm sp}$=6 from I=15 to I=23. These two maxima form for the frontal variability the double-peak structure we have described above. Similar structure is also seen around I=45 on the same section.

In comparison to Fig. 6.7, the spectra for u and v are different in the following two ways. Firstly, the energy is no longer concentrated on the basin scale, even in areas far away from the frontal axis. Secondly, u and v series are much more energetic than that of temperature. The energy is significant on virtually all the scales over the spectrum. The spectral pattern is complicated, and it is not our intention to give a characterization for the whole model domain. We focus only on the central region where the cold tongue intrusion takes place.

Figure 6.8 shows the spectrum for the zonal velocity of day 8 at vertical level 5 (125 m). Energy maxima happen mainly at $j^{\rm sp}$ = 3, 5, 6, and are identifiable at $j^{\rm sp}$ = 7 in the intrusion region. An apparent gap exists between $j^{\rm sp}$ = 3 and $j^{\rm sp}$ = 5 around (25, 40), reminiscent of the double-peak structure we have observed on the temperature spectrum. If the J=35 section map (Fig. 6.8, bottom) is examined, this structure becomes especially clear. The two maxima run from I=10 across the intrusion area all the way to the right until I=35, where the pattern undergoes some change. Over that area, two new peaks appear at $j^{\rm sp}$=4 and $j^{\rm sp}$=6.

A similar scenario happens on the spectrum for v. In Fig. 6.9, the scale level $j^{\rm sp}$=4 divides the spectrum into two halves. Energy is concentrated at $j^{\rm sp}$=3 on one side, and at $j^{\rm sp}$=5 and 6 on the other. This structure is also evident on the J=35 section map attached below between I=25 and I=40. But from I=36 - 47, energy becomes accumulated on $j^{\rm sp}$=3 only.

To summarize, the IFF variability in the intrusion region occurs mainly at two spatial scale levels: One is $j^{\rm sp}$ = 5 or 6, and another is $j^{\rm sp}$ = 3 for u and v and $j^{\rm sp}$ = 3 or 4 for the temperature. The spatial scales are approximately 20-30 km, and 60 km (40-60 km for temperature), respectively. This double-peak structure agrees well with the observation by Allen et al. (1994), but the gap obtained is smaller than their estimates, which give 15-17 km and 60-70 km for the two scales.

6.2.3 Scale window bound determination

With the spectral analysis, we are now able to determine the window bounds for the multiscale decomposition. The decomposition is not arbitrary. It depends on the process of interest. In this chapter we feel particularly interested in the formation of the sudden meandering on day 8, and accordingly the bounds should be set up in accordance with this event. In the following we first examine the time windows, and then switch to the spatial variability for the local averaging purpose.

On the time spectrum, the highest scale level is 10, as the signals are output at $1024 = 2^{10}$ time points. The value of j_2 hence must be 10. The setting of j_1 is also easy. We have identified two ranges of scales, one peaked at $j = 1$ (10.3 days), another varying from $j = 2$ to $j = 4$ (1.3-5.3 days). Level j_1 may be then set to be 5 to take in the event $j = 4$ (1.3 days). Note here we have used the fact that, in a multi-resolution analysis, the function space formed with a scaling basis with level j contains features with scale levels up to j but not including j.

For another bound index, j_0, the setting is a little problematic. On the spectrum, clearly $j = 1$ is not what we want for the short-term intrusion event. It should be put into the "large-scale" window, though its length is typical of a synoptic scale. But what about the other peak, which may occur either on $j = 2$, or on $j = 3$ and $j = 4$? A natural and reasonable choice is to put them together to make a "meso-scale" window. In that case, we have a $j_0 = 2$. The problem is, however, $j_0 = 2$ corresponds to a scale of 5.3 days, which might be a little too long if we want to focus just on the deep-sock intrusion. In order to give it an illustration, we draw in Fig. 6.10 a sequence of plots for the 125-m (Level 5) temperature from day 2 to day 10. Observe that the meandering intrusion begins to emerge roughly on day 6, and has been fully developed as of day 8. So it is of a scale less than 5 days, and $j = 2$ should be separated from the window of interest.

But from the time spectrum $j = 2$ has been identified as a level with significant energy, so what is it if it does not account for the meandering intrusion? Look at Fig. 6.10 again. In the western part, beginning day 2, the frontal axis is getting more and more depressed until it is straightened toward the southeast on day 5. After that, it starts to crest, and a pattern similar to that of day 2 resumes on day 7. This whole cycle takes about 5 days to complete, in agreement to the observed peak at $j = 2$ on the spectrum. Therefore, if only the meandered intrusion is focused on, this scale (i.e. the one corresponding to $j = 2$) should be ruled out, otherwise the MS-EVA analysis obtained will not yield the correct information. We will get back to this issue later in §6.4.

In addition to the time scale windows, we also need to determine the bounds for the spatial variation in order to set up the local averages. As clarified in the previous subsection, the highest spatial scale level is $j_2^{\mathrm{sp}} = 12$. For the remaining two parameters, j_0^{sp} and j_1^{sp}, they are determined by the spectral structure. As in the temporal case, we have also identified two peaks in the intrusion region on the space spectrum. One is at $j^{\mathrm{sp}} = 3$, and another at $j^{\mathrm{sp}} = 5$ or 6. The latter corresponds to a scale of 20-30 km, which we claim to be the right window for the deep-sock meander (cf. Fig. 6.10). So we may choose $j_0^{\mathrm{sp}} = 5$ or 4, and $j_1^{\mathrm{sp}} = 7$. In order to include more smaller features in the meso-scale window, j_1^{sp} may also be relaxed a little to 8, or even 9.

Another peak, $j^{\mathrm{sp}} = 3$, which has a scale of 60 km for u and v, is believed not relevant for the deep-sock meandering. By an inspection of the temperature sequence, it could correspond to the 5.3-day oscillation as in the time spectrum. But since on a temperature spectrum, there is also a $j^{\mathrm{sp}} = 4$ peak somewhere around the meander, things are more complex here. As we are not attempting to interpret processes other than the intrusion, this identified scale is left for future work.

In a brief summary, the time window bounds for the deep-sock intrusion study may be set as $j_0 = 3$, $j_1 = 5$, and $j_2 = 10$, though there is still some ambiguity in making the parametric combination. For the space windows, a reasonable setting is $j_0^{\mathrm{sp}} = 5$, $j_1^{\mathrm{sp}} = 7$, $j_2^{\mathrm{sp}} = 12$. But other choices, such as $j_0^{\mathrm{sp}} = 4$, $j_1^{\mathrm{sp}} = 8$ or 9, are not ruled out. A most appropriate choice will not be set until these candidates have been tested against sensitivity, which is deferred to §6.4.

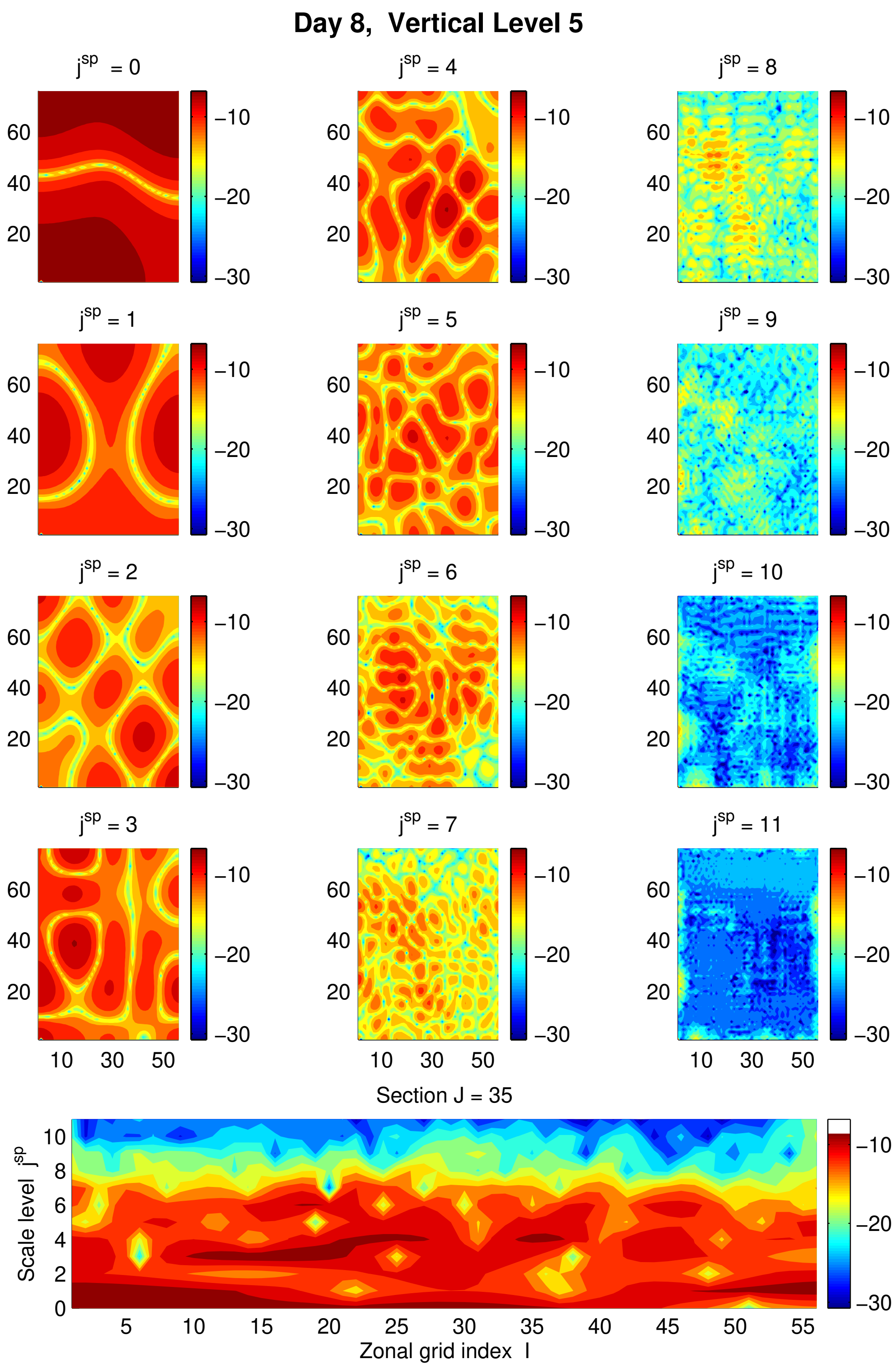

Figure 6.7: Space spectrum for the 125-m temperature on day 8. Each map corresponds to one particular scale level as indicated in the subplot title. The x and y axes for these maps are grid indices I and J, respectively. Graphed in the maps are the logarithm of energy instead of energy itself. Shown at the bottom is the spectrum along I on section J=35.

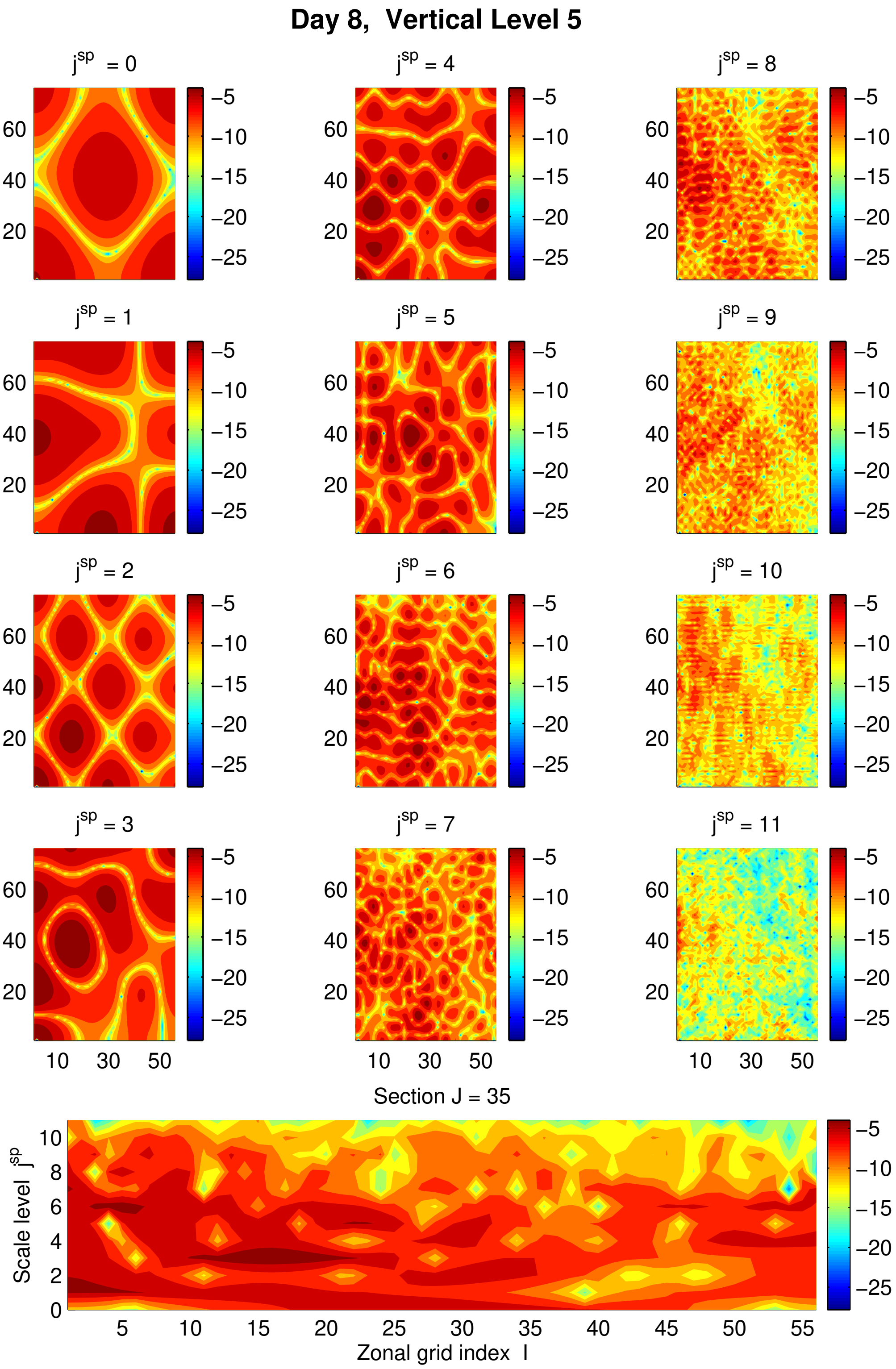

Figure 6.8: Same as Fig. 6.7, except for u.

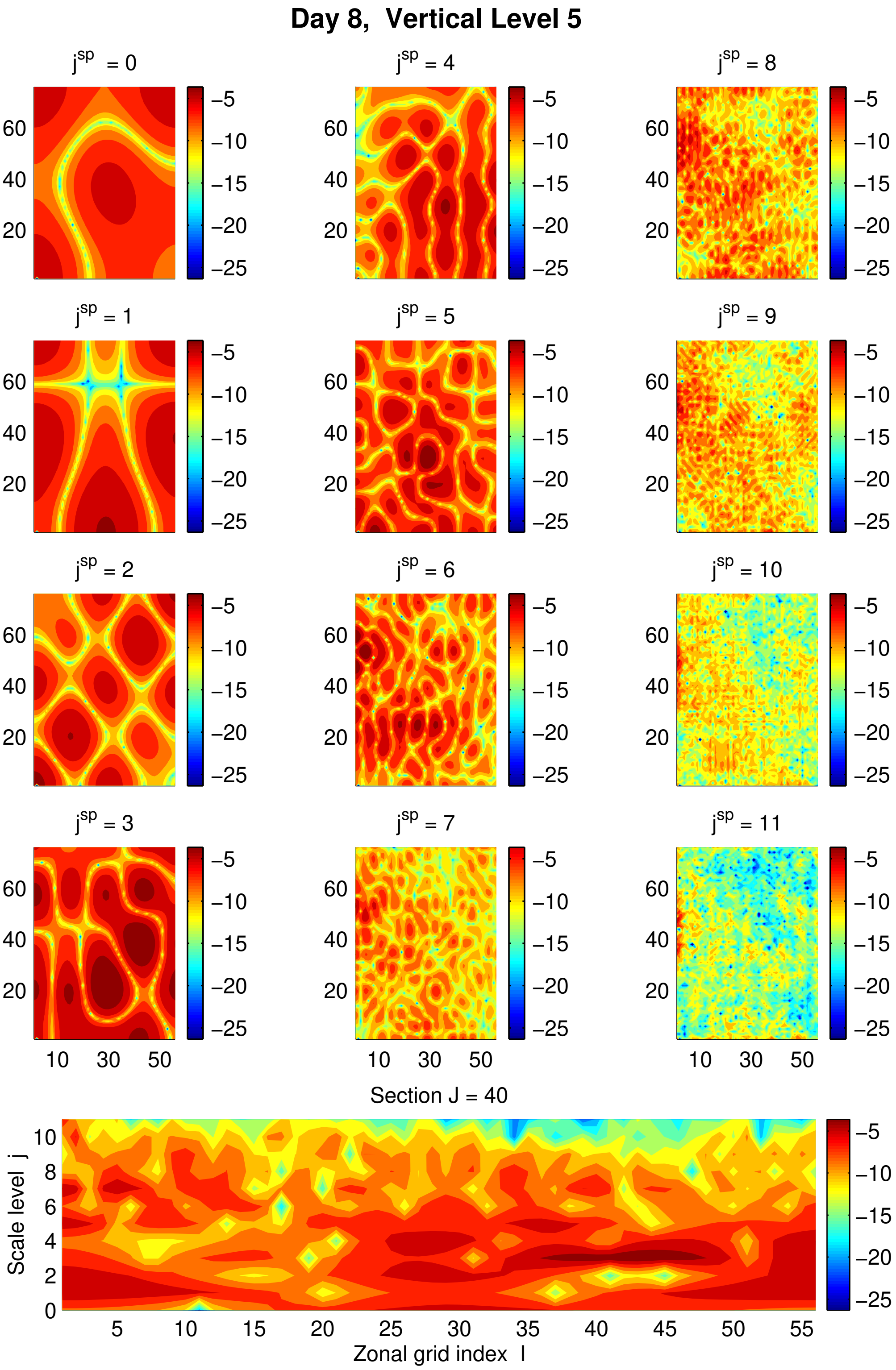

Figure 6.9: Same as Fig. 6.7, except for v.

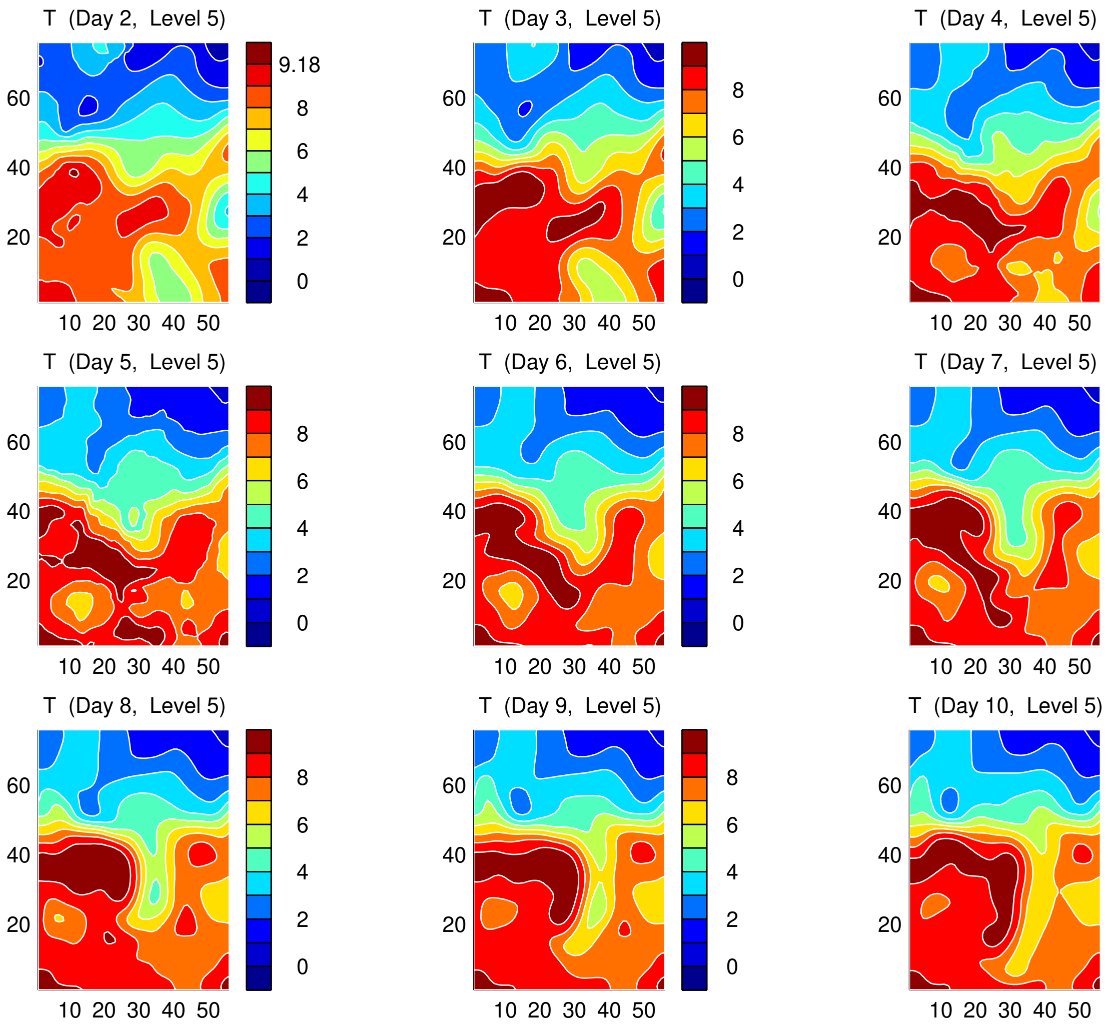

Figure 6.10: A time sequence of the forecast middle-depth temperature (125 m) on day 2 through day 10. The x and y axes are the grid indices I and J, respectively. The units for temperature are degrees Celsius.

6.3 MS-EVA study of the IFF variability

We now apply the MS-EVA to study the IFF deep-sock meandering. With the scale window bounds in hand, the application set-up becomes straightforward. What one needs to do is to get the MS-EVA ready and configure it with the forecast. The remaining task is then to visualize and read the outputs. In §6.3.2, energetics on are described on three vertical levels which are typical of three different layers in the water column. The purpose of this description is to find the corresponding dynamical process behind the observed meandering event, and the result is further substantiated in the subsection that follows (§6.3.3). In the last part of this section, we show that the energy gained from the identified process is indeed used to fuel the growth of the meandering cold intrusion.

6.3.1 MS-EVA set-up

The MS-EVA set-up begins with the determination of the time and space window bounds, which was already done in the preceding section. The remaining issues concern (1) the stationary density profile construction, (2) the hybrid vertical coordinate-to-z-coordinate interpolation, and (3) the time direction sampling.

We first need to build a background density $\bar{\rho}$, which is invariant in time and x and y. This is needed for both the pressure computation (cf. §A5.7) and the multiscale potential energy analysis. In the forecast, density anomaly ρ is computed from temperature T and salinity S using the 1980 UNESCO International Equation of State (IES'80) for seawater (see Fofonoff, 1985). We already have for T and S each typical vertical distributions both south and north of the front, as shown in Fig. 5.6. With the IES'80, the corresponding density distributions across the front are obtained accordingly. We take the arithmetic average of the two, then interpolate with cubic splines the obtained values, denoted as $\bar{\rho}^*$, onto 1024 equi-distant z points. The resulting $\bar{\rho}^*(z)$ is inputted into a low-pass filter, which is a scaling transform followed by a synthesis with a scale level $j = 2$. The output is taken as the stationary density $\bar{\rho} = \bar{\rho}(z)$. In Fig. 6.11, $\bar{\rho}$ (left), $s = -\frac{d\bar{\rho}}{dz}$ (right, solid) and $\frac{d \log s}{dz}$ (right, dashed) are plotted as functions of z. The profiles $s = s(z)$ is related to the Brunt-Väisälä frequency which is needed in calculating the buoyancy conversion, and $\frac{d \log s}{dz}$ is needed for the transfer term TS_{AM} in the

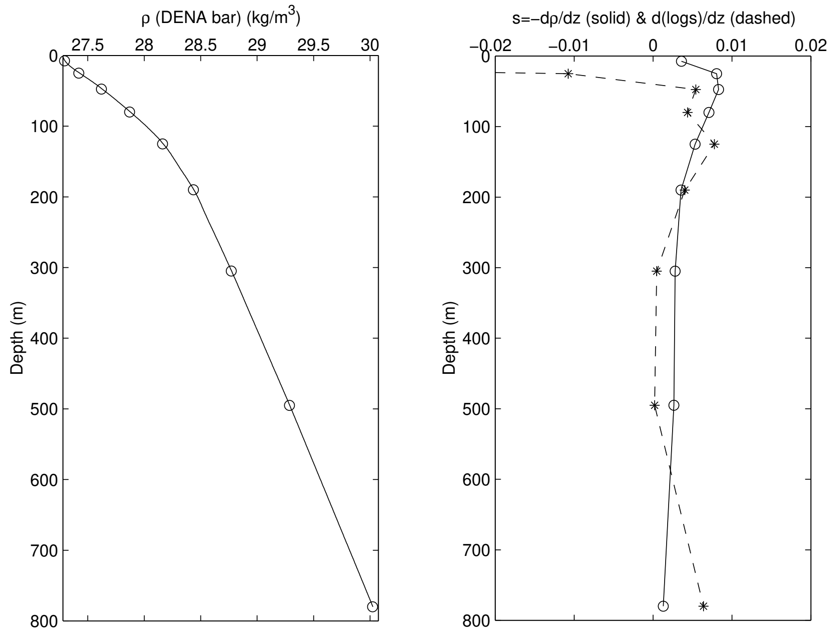

Figure 6.11: The stationary density anomaly and its derived properties, $s = -\frac{d\bar{\rho}}{dz}$ and $\frac{d\log s}{dz}$, as functions of z. The circles (and stars) mark the nine z-levels where the MS-EVA is to be applied.

APE equation. The circles and stars mark the nine horizontal levels where the MS-EVA is to be applied.

We also need to transform the hybrid vertical coordinate of the forecast model into a z-coordinate, as the MS-EVA is developed on horizontal levels. (A curved coordinate would introduce unphysical terms in the energetics.) The transformation is achieved through some appropriate interpolation. We have tried both linear and cubic-spline schemes, and found that the linear interpolation handles this case better (because of the few levels in the vertical direction). For simplicity, only nine levels are considered. They are located at the following depths: 7.5 m, 25 m, 47.5 m, 80 m, 125 m, 190 m, 300 m, 500 m, 780 m.

The third issue, time sampling, is determined by computational concerns. It is impractical to deal with huge files for energetics because of memory limits and computational efficiency. On an ULTRA-10 Sun microsystem, an IFF forecast with a length more than 1024 points in time will generate a file using too much storage and make the computation intolerably expensive. To maximize the yield within the machine capacity, we set the series length to be 1024 (j_2 is hence $\log_2 1024 = 10$). Recall that the time step adopted in the simulation is

180 s (refer to Table 5.2). If the HOPS output were directly taken as the MS-EVA input, the series would span only 1024×180 s $= 2.13$ days, which is apparently too short for our purpose. To solve this problem, we sample the available sequences every ten time steps. This yields a time span of 21.3 days. But, again, this sampling cannot be performed directly on the HOPS output. The modes with time scales smaller than ten step sizes must be filtered out or aliasing could arise. The filtering is done with the scaling transform introduced before. To be specific, given a data series with 1024×16 time points, we perform on it, again, a scaling transform followed by a synthesis with scale level $j = 10$. The output is the filtered series, and what we need to do is to sample the first 10240 time points (every 10 steps) to obtain the MS-EVA-ready data file.

Once the above issues are resolved, the MS-EVA set-up is completed. For reference, the related parameters adopted in this application are tabulated in Table 6.1. Notice again the nine "vertical levels" in this chapter are not those used in the forecast model in Chapter 5 any more. When they are referred to, what they mean are the z-levels listed in Table 6.1.

6.3.2 MS-EVA diagnosis of the IFF meandering intrusion

An application of the MS-EVA is bound to open a Pandora's box. The feature-rich energy maps reveal to us a colorful set of fingerprints underlining the complex dynamics. For simplicity, only the meso-scale energetics are studied here (but with interactions with other windows included), as it is (by construction) the meso-scale window that is essential for elucidating the formation of the deep-sock intrusion. In this section, we describe the standard experiment with the time and space window bounds being $j_0 = 3$, $j_1 = 5$, $j_0^{\mathrm{sp}} = 5$, and $j_1^{\mathrm{sp}} = 9$, as justified in the preceding section. The sensitivity of the MS-EVA terms to these parameters will be examined later.

Figures 6.12 and 6.13 are the 125-m meso-scale (in time) MS-EVA distributions for the eighth forecast day, a day when the cold intrusion is observed. In each figure, the first eight contour maps (a-h) show the energetic terms of a meso-scale energy equation (Eq. 3.72 or 3.63), while on the right-bottom corner is the total APE/KE transfer from the large-scale window. Terms appearing in Eq. 3.72 (Eq. 3.63) but missing in Fig. 6.12a-h (Fig. 6.13a-h) are not significant in comparison to their counterparts in the figure. The horizontal and vertical

Table 6.1: Parameters for the application of MS-EVA.

Parameters	Value
Time window bounds	
j_0, j_1, j_2	3, 5, 10
Space window bounds	
j_0^{sp}, j_1^{sp}, j_2^{sp}	5, 9, 10
$\bar{\rho}(z)$, s, $\frac{d\log s}{dz}$	See Fig. 6.11
Grid	$57 \times 77 \times 9$
Time stepsize Δt	1800 s
Horizontal grid spacing Δx, Δy	2.5 km
Vertical grid spacing Δz	
level 1 (at depth 7.5 m)	15 m
level 2 (25.0 m)	20 m
level 3 (47.5 m)	25 m
level 4 (80.0 m)	40 m
level 5 (125.0 m)	50 m
level 6 (190.0 m)	80 m
level 7 (300.0 m)	140 m
level 8 (500.0 m)	260 m
level 9 (780.0 m)	300 m

diffusions are left for future work and not shown here, either. The BC and BT_H graphed respectively in Figs. 6.12i and 6.13i are of particular interest. Their spatial averages have been justified to be indicators of the two types of instabilities (cf. Chapter 4). The other instability-related quantity, BT_V, is very small compared to BT_H and hence is ignored in the pictorial presentation.

A general observation about Figs. 6.12 and 6.13 is that the meso-scale event happens mainly in a small region where the deep-sock meander is observed. Outside this region the density and the current are much less energetic (almost negligible). Another observation is that these quantities oscillate a lot on the horizontal plane. High and low centers usually appear in pairs. The oscillation amplitudes can be easily identified from maps with features further decomposed by spatial scales. Shown in Figs. 6.14 and 6.15 are such maps, but with only the meso-scale (in space) features retained. They correspond to Figs. 6.12 and 6.13, respectively, with a meso-scale window (in space) synthesis applied in the horizontal direction. These maps tell the localities of the processes that are energetic on the meso-scale window, both in space and in time.

Shown above is an example of the original meso-scale energetic terms. These maps, though calculated directly from the MS-EVA equations, are usually not helpful in process identification. The large-amplitude oscillation, as shown in Figs. 6.14 and 6.15, tends to disguise the features which are essential to the process, though it is not impossible in some special cases to distinguish from them spots of interest. (For example, one may see a positive baroclinic transfer and a negative buoyancy conversion on a larger scale from the distributions of BC and b_n^M.) Recall what we have done in the highly idealized examples in Chapter 4. It is the residual of the oscillation that counts. In the following we then focus only on the energetic residuals, which are obtained through local averaging, i.e., a spatially large-scale synthesis (with j_0^{sp} pre-set). The local averaging basically eliminates the oscillation with levels $j^{\mathrm{sp}} \geq j_0^{\mathrm{sp}}$, while allowing the residual to vary at a spatial scale just a little larger than the size of the deep-sock meander. *From now on to the end of this chapter, all terms referred to are supposed to have been locally averaged, though for convenience the same symbols as before will be used throughout.* No further clarification will be made unless otherwise needed.

The 125-m level

We first look at the 125-m meso-scale potential energetics. Figs. 6.16-6.21 show the locally averaged APE distributions for days 5 through 10. The layouts for these figures are the same as that of Fig. 6.12, with the first eight maps from the APE equation, and the last one being BC, the total transfer from the large-scale window. As shown in Chapter 4, BC is the baroclinic instability indicator. We begin the description with this quantity.

From day 5 to day 9, obviously there is a solitary positive spot on each of the BC maps. Recall that day 5 is the starting point of the F5 forecast after the model takes in all the zigzag data, and day 8 the day when the cold intrusion is observed (refer to Chapter 5). We have speculated that some process could be happening during this period. The BC maps shown here strongly support this speculation. On day 5 (Fig. 6.16i), the hotspot is centered around (20, 30), with an orientation toward the southeast. One day later, the location and orientation are still the same, but the strength is significantly enhanced and the influence area enlarged (Fig. 6.17i). On day 7, while its magnitude continues to grow, its alignment is suddenly switched, from northwest-southeast to southwest-northeast (Fig. 6.18i). This sudden change of orientation is followed by a chain of morphogenesis and migration, and as of day 8, the hotspot has moved to the south, with the influence region shrunk, marking the maturing of the meander (Fig. 6.19i). After that day, the strength of the hotspot diminishes, and eventually disappears from our focus (Fig. 6.20i-6.21i).

The migration and deformation of the BC hotspot correspond exactly to the formation and evolution of the deep-sock intrusion. It is thus safe to claim that we have captured the right dynamics for the event in study. The positive value of BC, as we have shown before, clearly indicates that some interesting process is going on in the region of concern. Potential energy is transferred from the large-scale window to fuel the perturbation growth, and the transfer is getting stronger and stronger until day 8, when it reaches its maximum and goes downhill afterwards. On day 10 there is also such a positive BC center, but the position is away from where we focus. Apparently, the intrusion event has been terminated as of that time.

In Figs. 6.16 through 6.21, the maps other than that of BC present the distributions for the meso-scale APE terms. The meaning for these symbols is referred to Table 3.2. For clarity, we have omitted those terms, such as TS_{AM}, which are not significant in comparison to the ones shown here. The vertical and horizontal diffusions are not considered in this context, either.

On the balance of the available potential energetics, generally all terms are important, but in many cases, we may identify the time rate of change of A_n^M ($\hat{\delta}_n A_n^M = \dot{A}_n^M$), the buoyancy conversion (b_n^M), and the APE transfer due to the horizontal density gradient ($T_{A^M,\partial_h\rho} = T_{A^M,\partial_x\rho} + T_{A^M,\partial_y\rho}$), as larger in absolute value than the others. Among these three terms, $\dot{A}_n^M$ is the result, and the transfer and buoyancy conversion are the two causes. For the transfer, the most important part is BC, which is from the interaction analysis and has been described above. We hence look at the evolution of b_n^M only.

Generally speaking, the buoyancy conversion evolves in the same way as BC does, except that now the hotspot is a negative center, which indicates a conversion of energy from APE to KE. From Figs. 6.16e-Fig. 6.19e, the conversion is getting stronger and stronger, until the meander matures on day 8. This increasing trend (in absolute value) agrees well with our previous BC observation. In fact, the place of conversion coincides roughly with the place of transfer. On day 7 and day 8 (Figs. 6.18e and 6.19e), the maximal conversion (maximum in absolute value) happens approximately at the highest point on the BC map, though on days 5 and 6 (Figs. 6.16e and 6.17e)), it is located a little northwest. After day 8, the conversion becomes less relevant to the meandering event.

Above is a brief characterization of the APE energetics. Now look at their kinetic counterparts. Figures 6.22-6.27 show the locally averaged (in space) meso-scale (in time) kinetic energetics for the 125-m level from days 5 to 10. Terms missing from the KE equation are not significant in comparison to those shown here.

In contrast to the potential energetics, no hotspot such as the one seen on the BC map is identified on the level-5 KE maps, except for the buoyancy term, which is merely b_n^M with an opposite sign. For the barotropic indicator, BT_H, the high and low centers usually appear in pairs, and, within a pair, the negative center has an absolute value larger than the positive center has (e.g., Figs 6.24i and 6.25i). This generally excludes the possibility of barotropic instability, as the local average of BT_H over a region that covers the two centers will be less than zero. In fact, these distributions, among many others, provide a good example of self-similarity. On a larger scale, similar centers also appear in pairs, though with a different size (figures not shown).

But on day 6, there does exist a positive center on the BT_H map. It is located, approximately, between I=10-30, J=30-50. Weak as it might be, the magnitude ($1 \times 10^{-8}\ m^2/s^3$) is comparable to the one on the BC map for the same day ($3 \times 10^{-8}\ m^2/s^3$). From the balance it should result mainly from $T_{K^M,\zeta}$. One may thus expect something happening there due to the evolution of vorticity.

In a brief summary, on the energetic maps for the mid-depth (125 m), there is an apparent center of large-to-mesoscale potential energy transfer. Its location and evolution, both in strength and in geometry, correspond well to the formation of the deep-sock meander. On the buoyancy conversion maps, there is also a center corresponding to the one on the BC maps, but with a negative sign. In contrast to the potential energetics, the KE maps generally do not exhibit a feature as conspicuous as the one on the APE maps, except on day 6, when a weak center of positive KE transfer is identified. All these features observed so far are for the 125-m level. They don't contain information about vertical variation. In order to see how these features vary with depth, we next examine two other levels, one above the 125-m depth and another below. We first look at the one above, which is characteristic of the surface layer dynamics.

The 25-m level

Figures 6.28-6.35 show the potential and kinetic energetics for the 25-m level (vertical level 2) on day 5 through day 8. In Figs. 6.28-6.31, the solitary BC hotspot still exists, but it is located a little south of the location identified on the level-5 maps, and besides, no indication of migration and re-orientation is observed. The isolation of a negative buoyancy conversion center is not evidenced until day 6.

On the kinetic energy maps, the distribution and evolution of BT_H is similar to that of the 125-m level. A distinct feature is that, on day 6, there appears a solitary parallelogram between I=10-30, J=30-50 where it is positively valued (Fig. 6.33). In terms of BT_H magnitude, this parallelogram is stronger than the one observed in Fig. 6.23. Clearly, BT_H is mainly from the transfer due to the rotating fluid, as the distribution of $T_{K^M,\zeta}$ is very similar to BT_H. Another observation is that there is also a parallelogram on the map of $\Delta_z Q_{K^M}$, located just a little bit left to the BT_H center (Fig. 6.33c). This is a negative center, with a maximal absolute value comparable to that of BT_H (3.5×10^{-8} m^2/s^3 versus 5×10^{-8} m^2/s^3). It is thence reasonable to conjecture that a major part of energy transferred from the large-scale KE in the surface layer is moved downward in the water column through advection.

Different from that of the 125-m level, one observes a solitary spot on the $\Delta_z Q_{PM}$ maps in Figs. 6.34 and 6.35. This hotspot is located at the intrusion area and seemingly evolves in pace with the buoyancy conversion center, except with a much larger strength. From the Eady model energetics, this is a good indicator for a baroclinic instability. But at this moment, we are not sure whether it is from the mid-depth buoyancy conversion. In other words, its physical implication is still not clear.

The 300-m level

We have seen in the energetics for both level 2 (25 m) and level 5 (125 m) that a conspicuous feature is the significant transfer of potential energy from the large-scale window to the meso-scale window, and an apparent conversion from meso-scale potential energy to meso-scale kinetic energy. One may conjecture there is a baroclinic instability happening in the related region. But the corresponding signatures are not significant on the maps of vertical pressure work rate $\Delta_z Q_{PM}$. Recall that in the Eady case, $\Delta_z Q_{PM}$ carries the energy converted from APE toward the two vertical boundaries. One hence expects a net gain of energy in the surface and bottom layers and a net loss at the middle depths on the maps for $\Delta_z Q_{PM}$. In Figs. 6.34 and 6.35, this seems to be the case, but on the level-5 energetic maps, the correlation between $\Delta_z Q_{PM}$ and b_n^M in the intrusion region is poor. We therefore should not expect an Eady-like instability scenario there. In fact, if the total transfer from the large-scale window, $BC + BT_H + BT_V$, is plotted, one cannot find a sole positive center as we did for the BC distributions at the 25-m and 125-m levels, except that on day 6, a hotspot does exist, but at the same place as the one identified from the BT_H map (see Figs. 6.36 and 6.37). On days 7 and 8, the total transfer will be negative if averaged over a larger scale. By what we have shown in §4.3.4, the system should be stable in this case.

We hence cannot say that there is a baroclinic instability in the surface and middle layers, though both the BC distributions and buoyancy conversion suggest so. But, what about the deep layers, i.e., the layers below 125-m in the water column? An examination of the 300-m energetics is expected to give this question a satisfactory answer.

Shown in Figs. 6.38-6.45 are the level-7 (300 m) potential and kinetic energetics for days 5 through 8. An immediate observation is that the APE transfer, BC (Figs. 6.38i-6.41i), is significantly larger than BT_H, the KE transfer (Figs. 6.42i-6.45i). As a result, the distribution of the total energy transferred from the large-scale window will not change much just because of the addition of BT. Fig. 6.46 confirms this assertion, and reveals clearly a positive transfer center around the deep-sock intrusion area. The strength of the transfer increases from a maximum of 1.25×10^{-7} m^2/s^3 on day 5, to a maximum of 1.5×10^{-7} m^2/s^3 on day 6, then goes down a little to 0.9×10^{-7} m^2/s^3 on day 7. The influence radius keeps increasing until day 7, the day just before the meander matures. When time reaches day 8, the process switches to something else.

The scenario shown in Fig. 6.46 reveals clearly that the system around the cold intrusion area is unstable, and from Figs. 6.38i-6.41i, the instability is baroclinic. This assertion is not only evidenced in the buoyancy conversion (Figs. 6.38e-6.41e), as we have seen before for the level-2 and level-5 cases, but also supported by the distribution of the vertical pressure work $\Delta_z Q_{PM}$. One may find that, in Figs. 6.42f-6.45f, there is a negative center whose location roughly agrees with the hotspot on the $-b_n^M$ distribution. Its evolution also correlates highly to the evolution of b_n^M. Both of them grow in (absolute) magnitude, from day 5 to day 7, and then move downhill until the intrusion tongue forms. Apparently, a part of the energy converted from the APE is being carried through $\Delta_z Q_{PM}$ away from level-7, driving the motions above and below. This scenario is exactly what we have seen for an unstable Eady model, and therefore our previous assertion of baroclinic instability is convincingly verified.

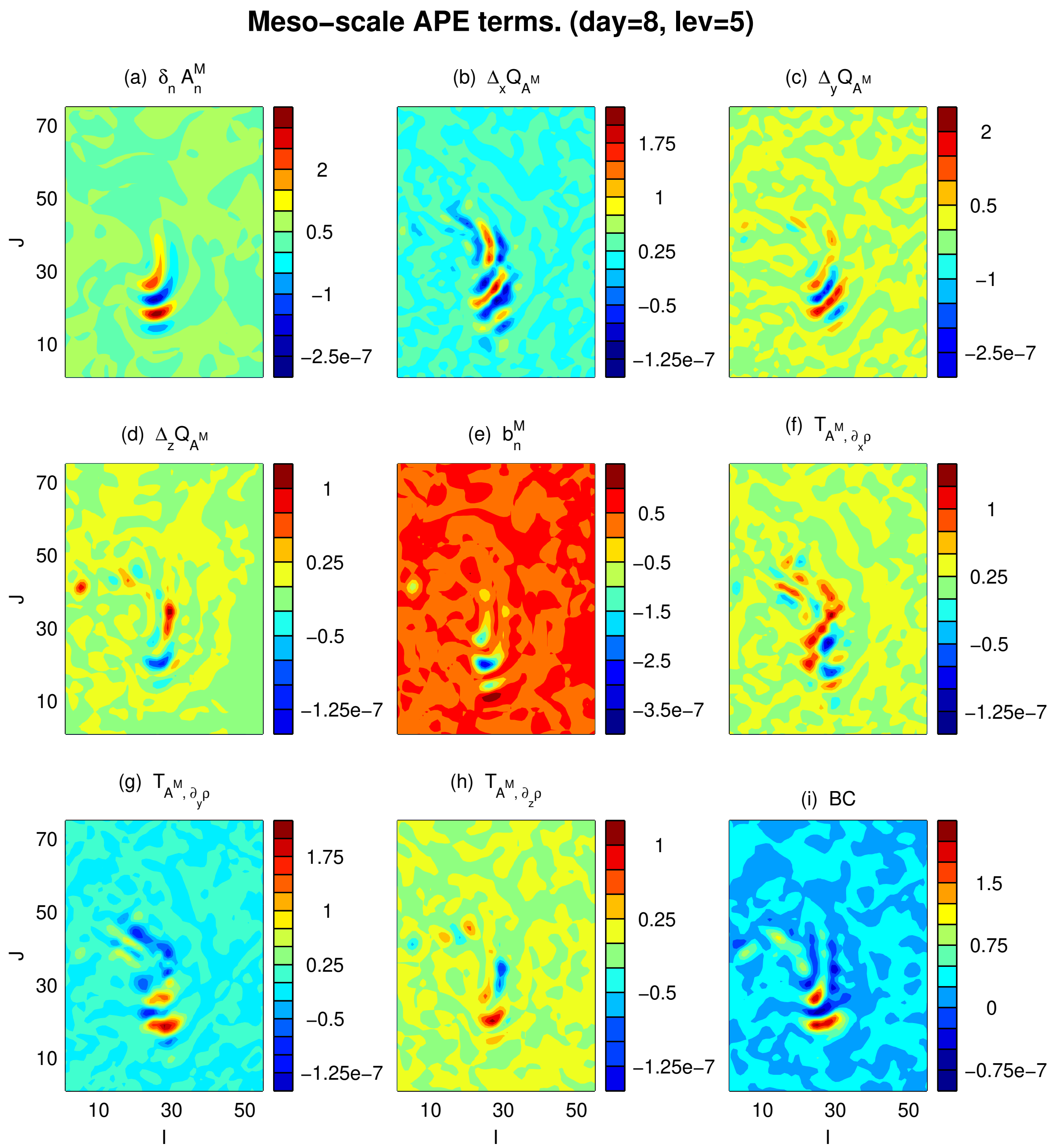

Figure 6.12: Meso-scale (in time) potential energetics (m^2s^{-3}) for day 8 at depth 125 m. The first eight subplots [(a) through (h)] show terms from the meso-scale APE equation (cf. Table 3.2), and the last one, BC, is the total APE transferred from the large-scale window.

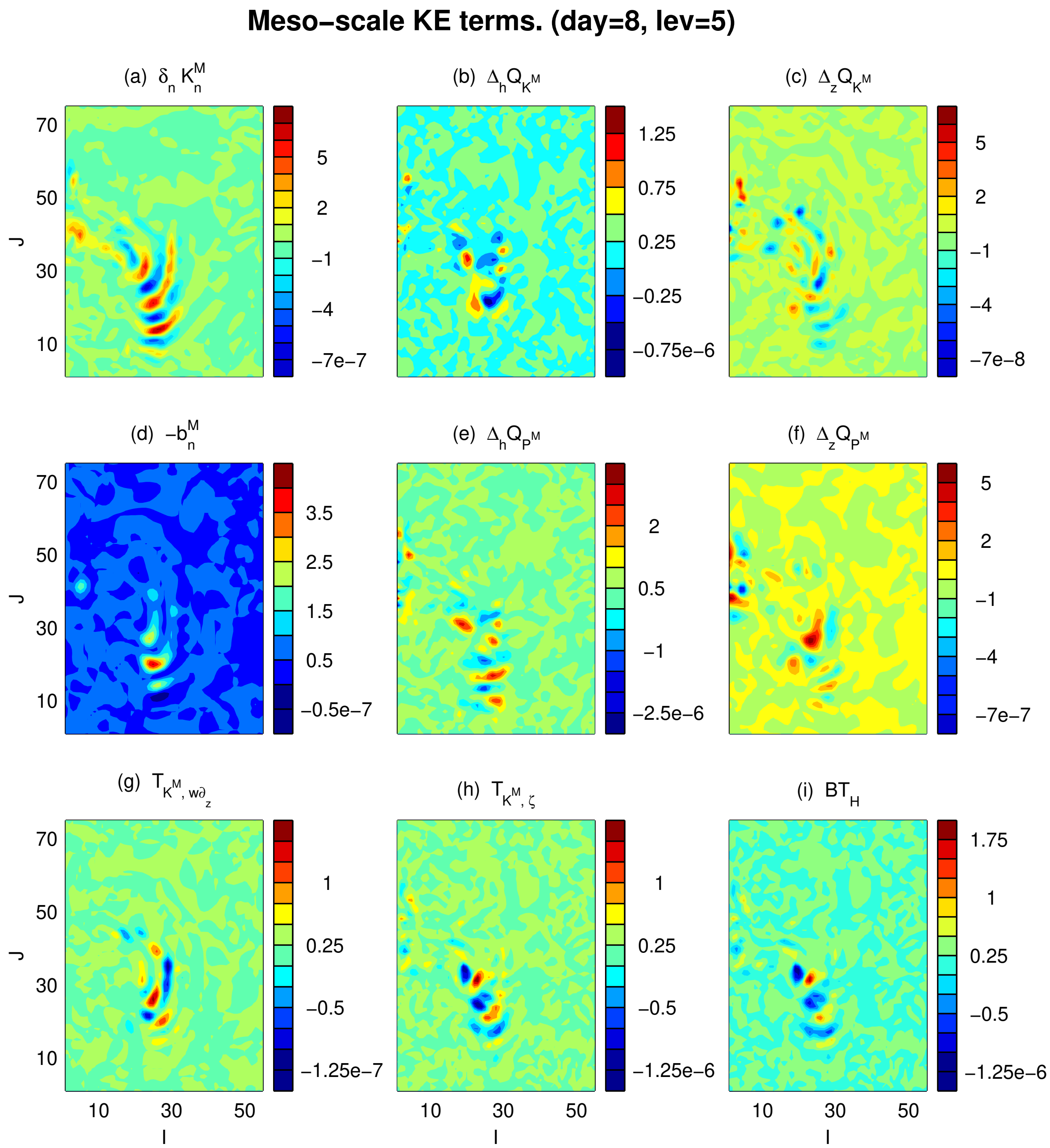

Figure 6.13: Same as Fig. 6.12, but for kinetic energetics. Shown in subplot (i) is BT_H, the part of transfer $T_{K^M,\zeta} + T_{K^M,w\partial_z}$ contributed from the large-scale window.

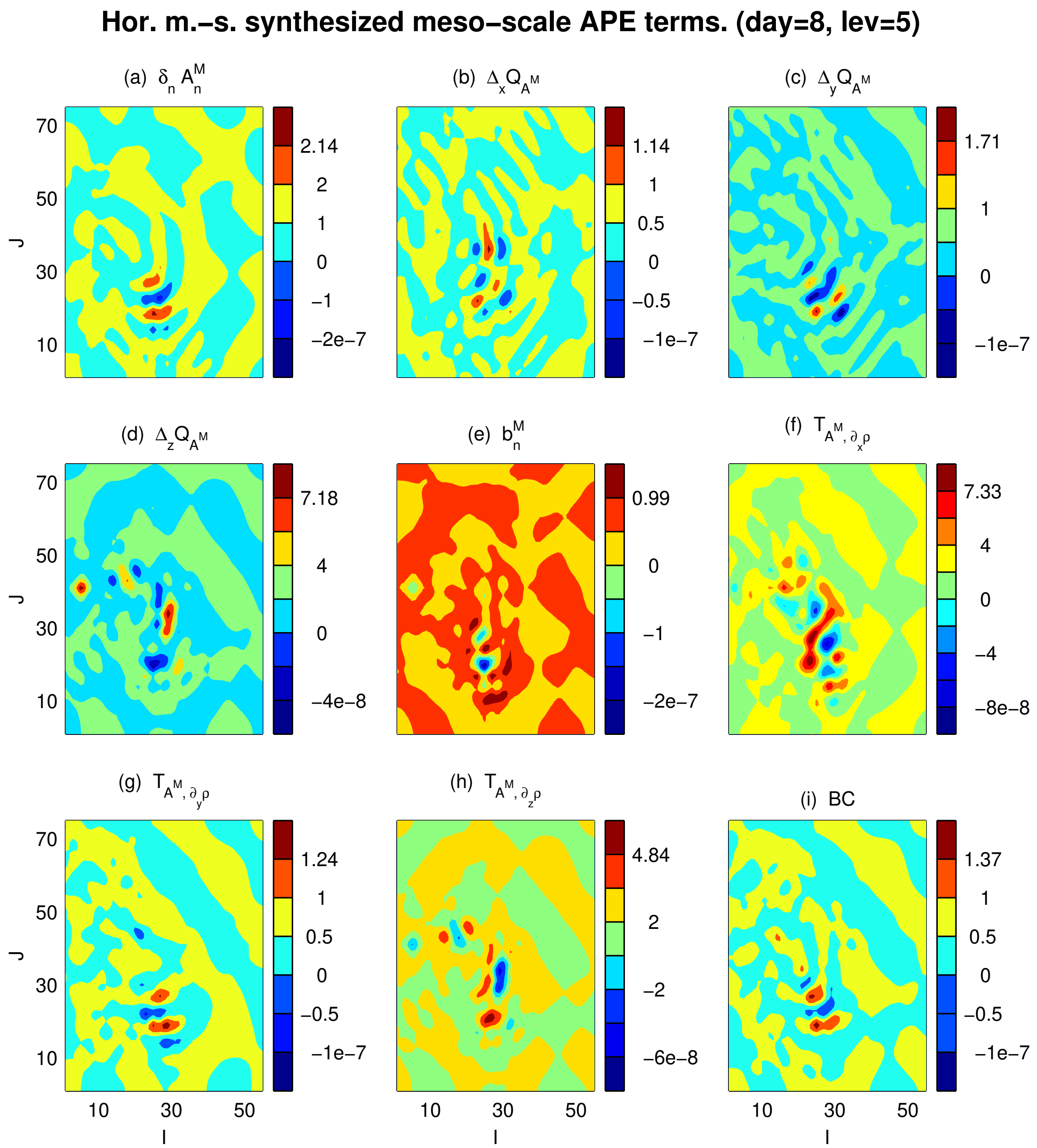

Figure 6.14: Same as Fig. 6.12, but now all the fields are horizontally meso-scale synthesized.

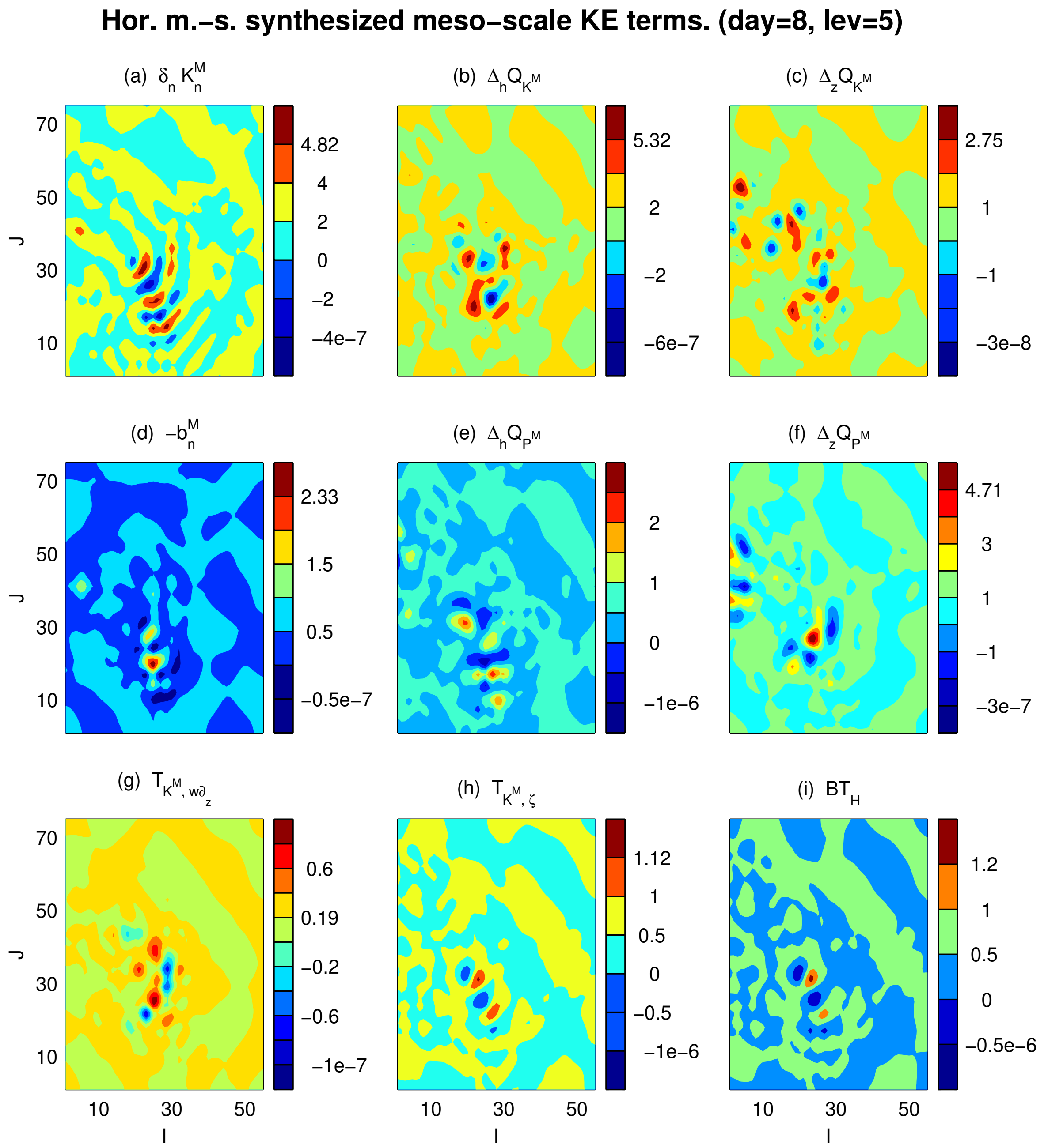

Figure 6.15: Same as Fig. 6.13, but now all the fields are horizontally meso-scale synthesized.

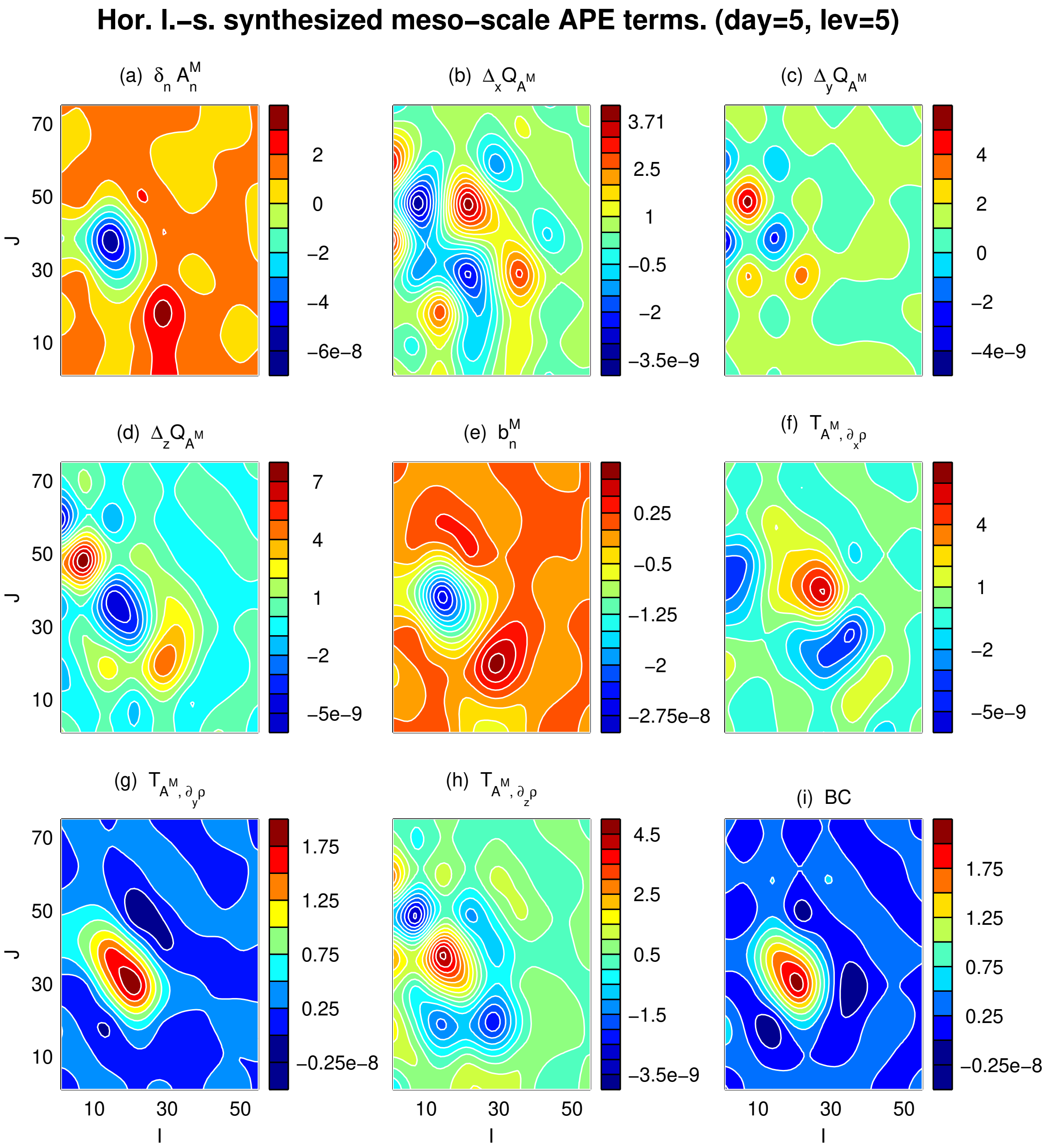

Figure 6.16: Locally averaged meso-scale potential energetics (in m^2s^{-3}) for day 5 at depth 125 m. The first eight subplots [(a) through (h)] show terms from the meso-scale APE equation (cf. Table 3.2), and the last one, BC, is the total APE transferred from the large-scale window. All terms are supposed to be locally averaged, which is a large-scale synthesis in the horizontal direction with $j_0^{sp} = 5$.

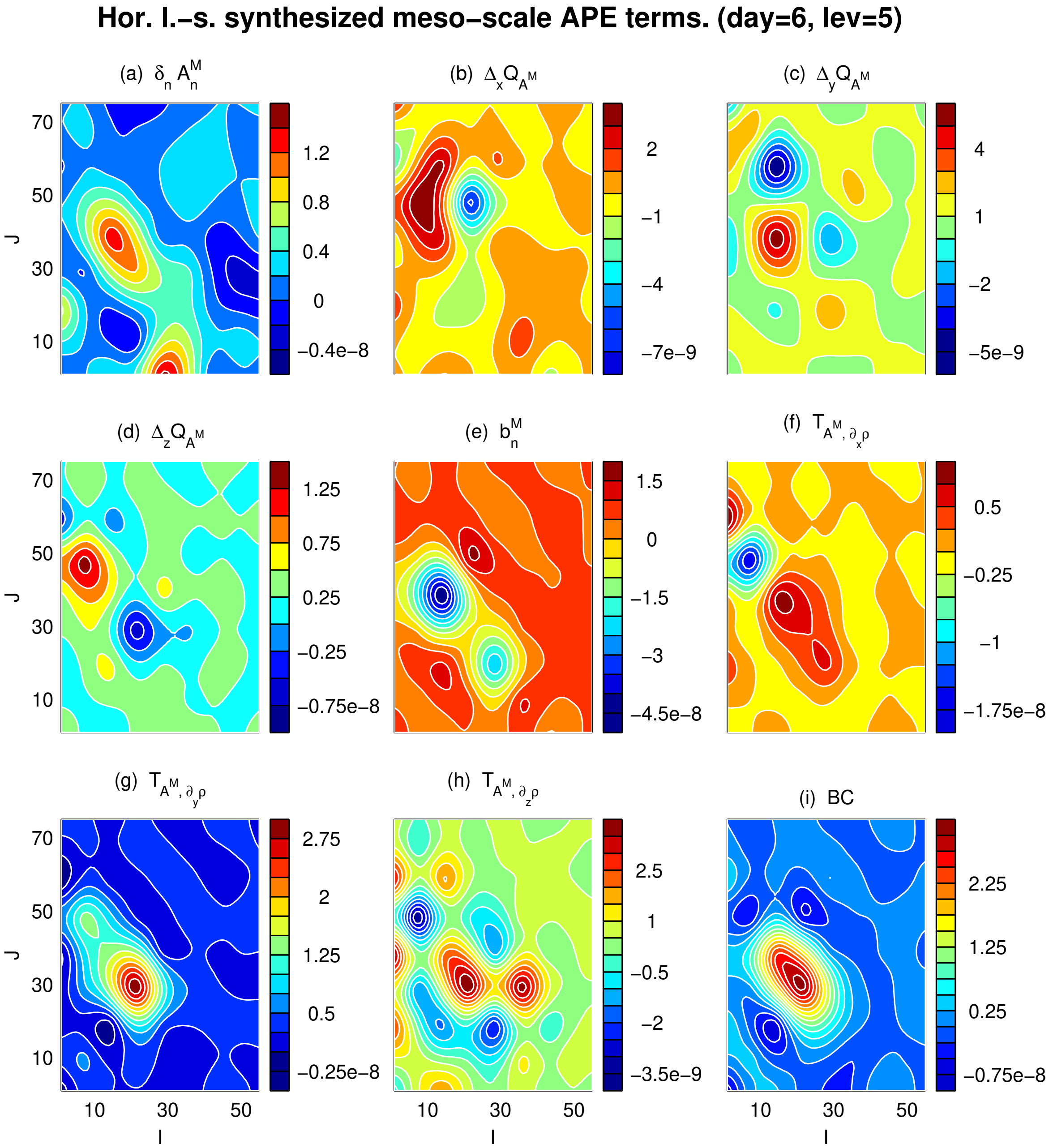

Figure 6.17: Same as Fig. 6.16, but for day 6.

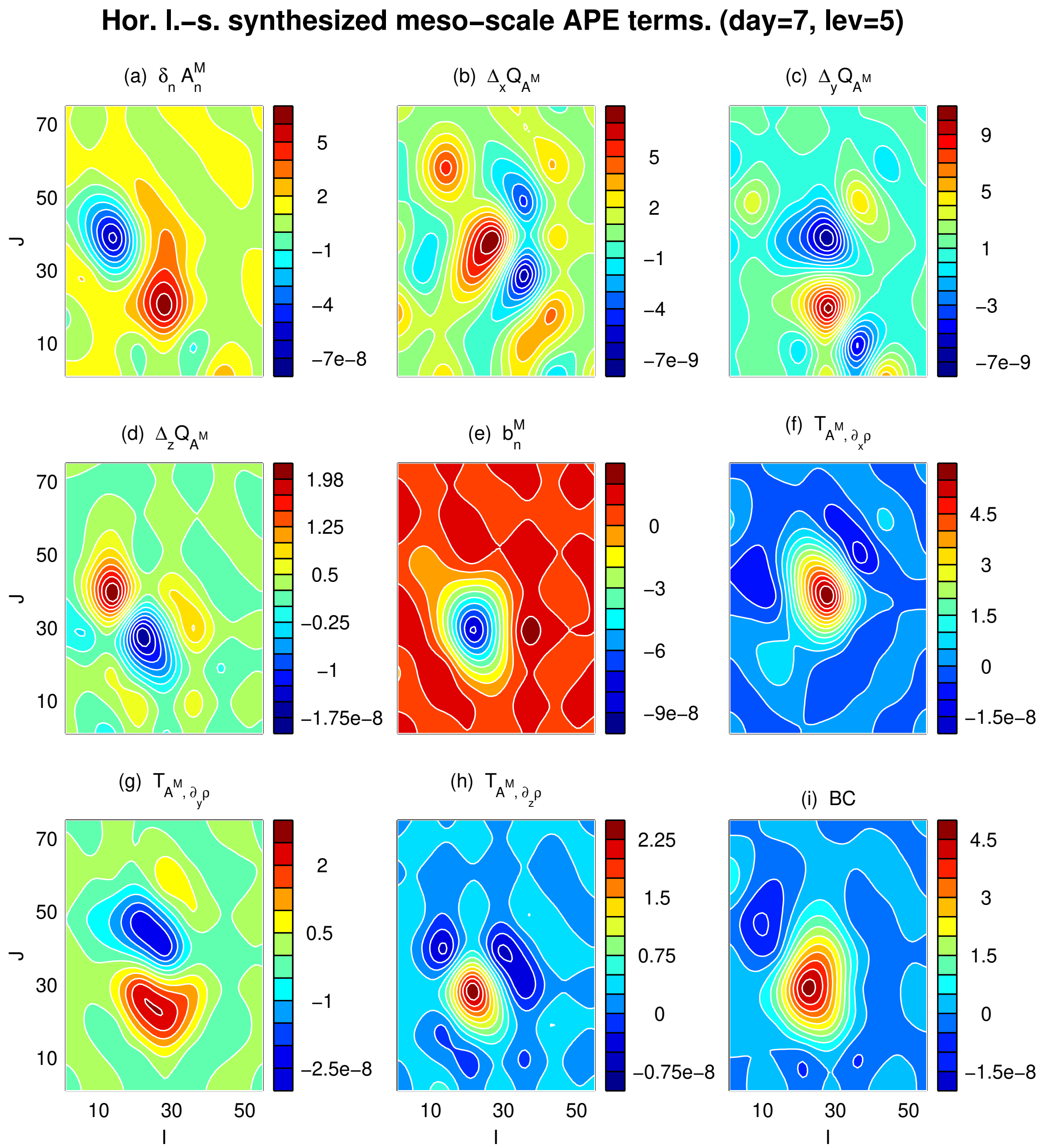

Figure 6.18: Same as Fig. 6.16, but for day 7.

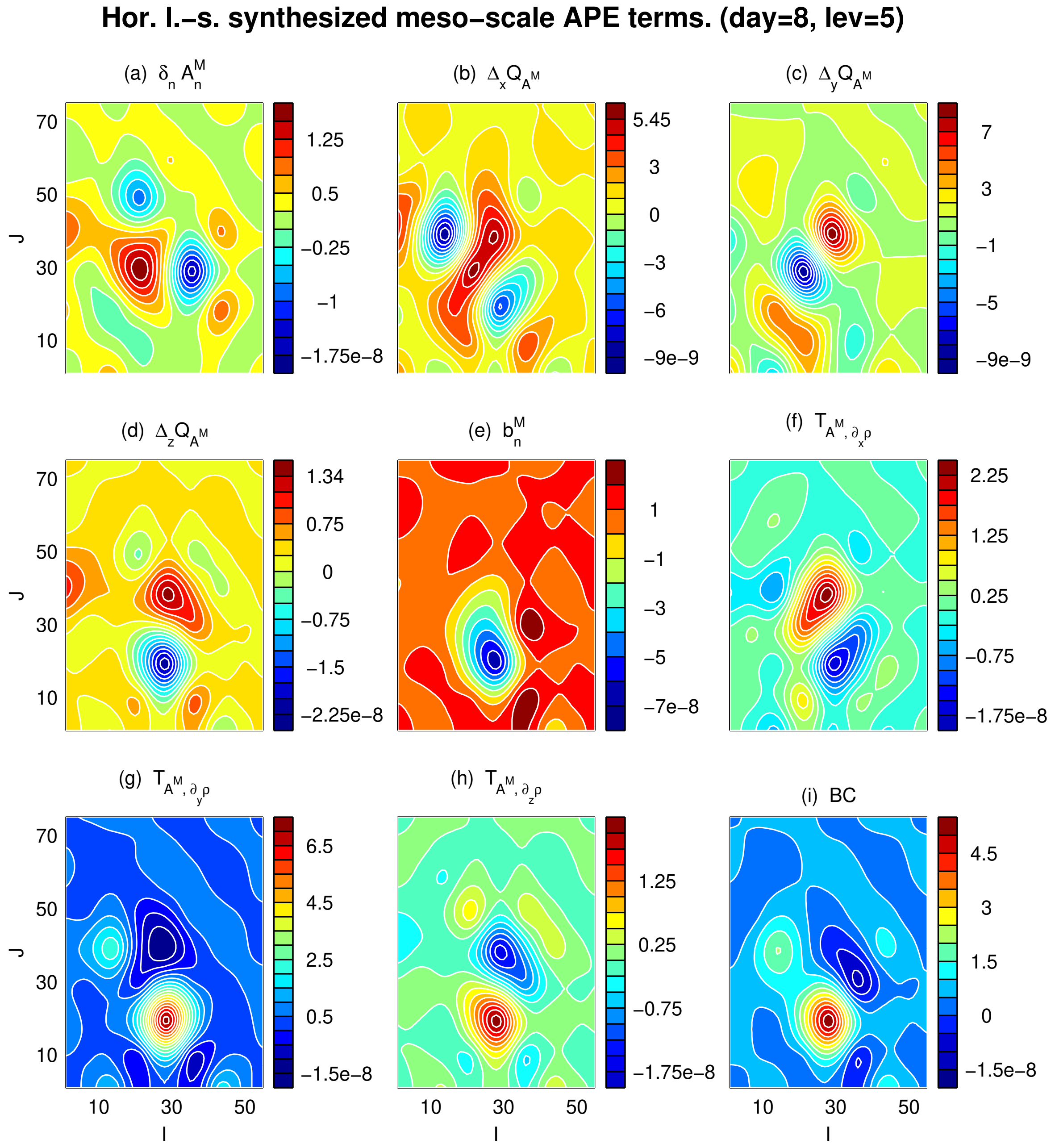

Figure 6.19: Same as Fig. 6.16, but for day 8.

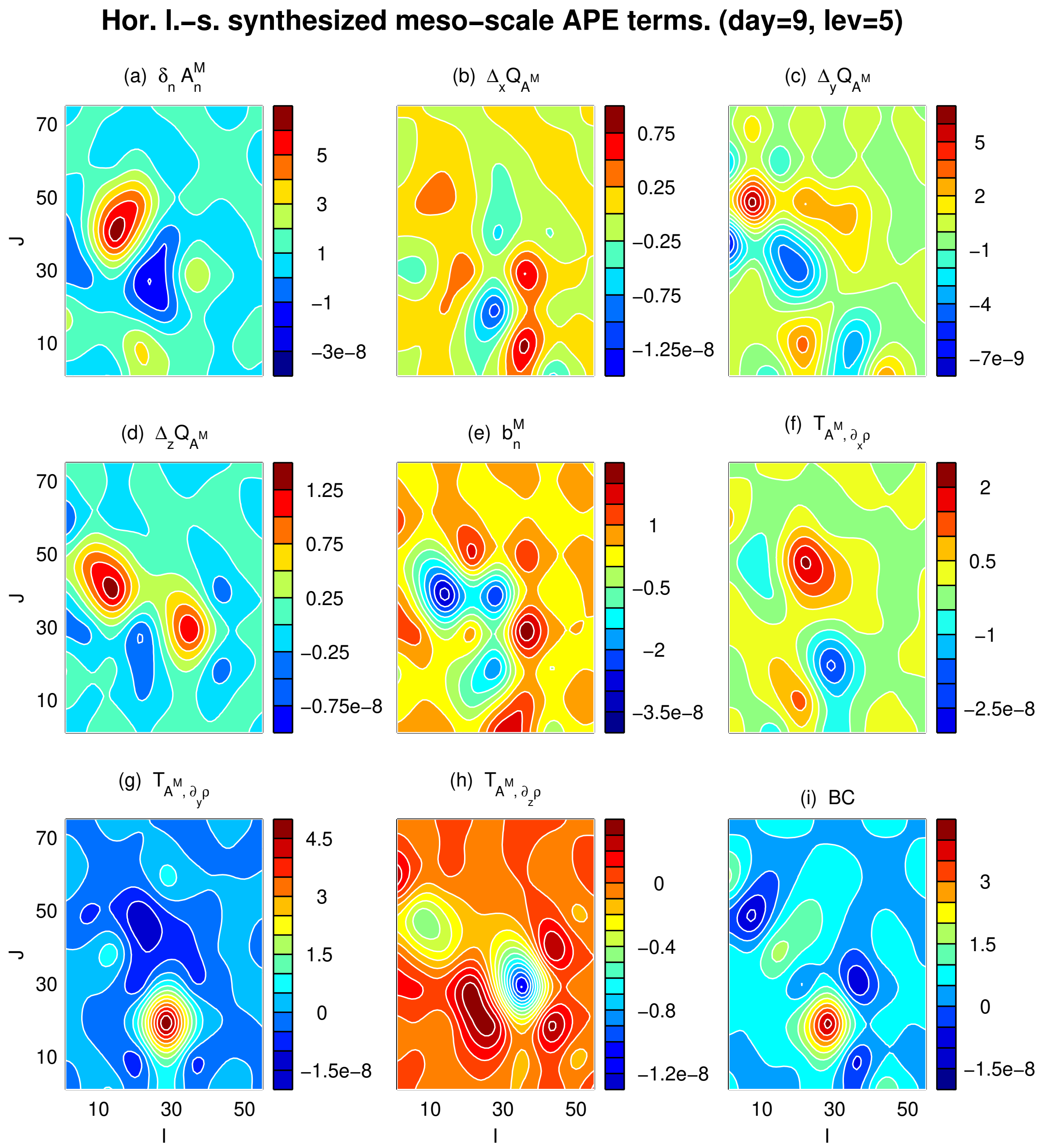

Figure 6.20: Same as Fig. 6.16, but for day 9.

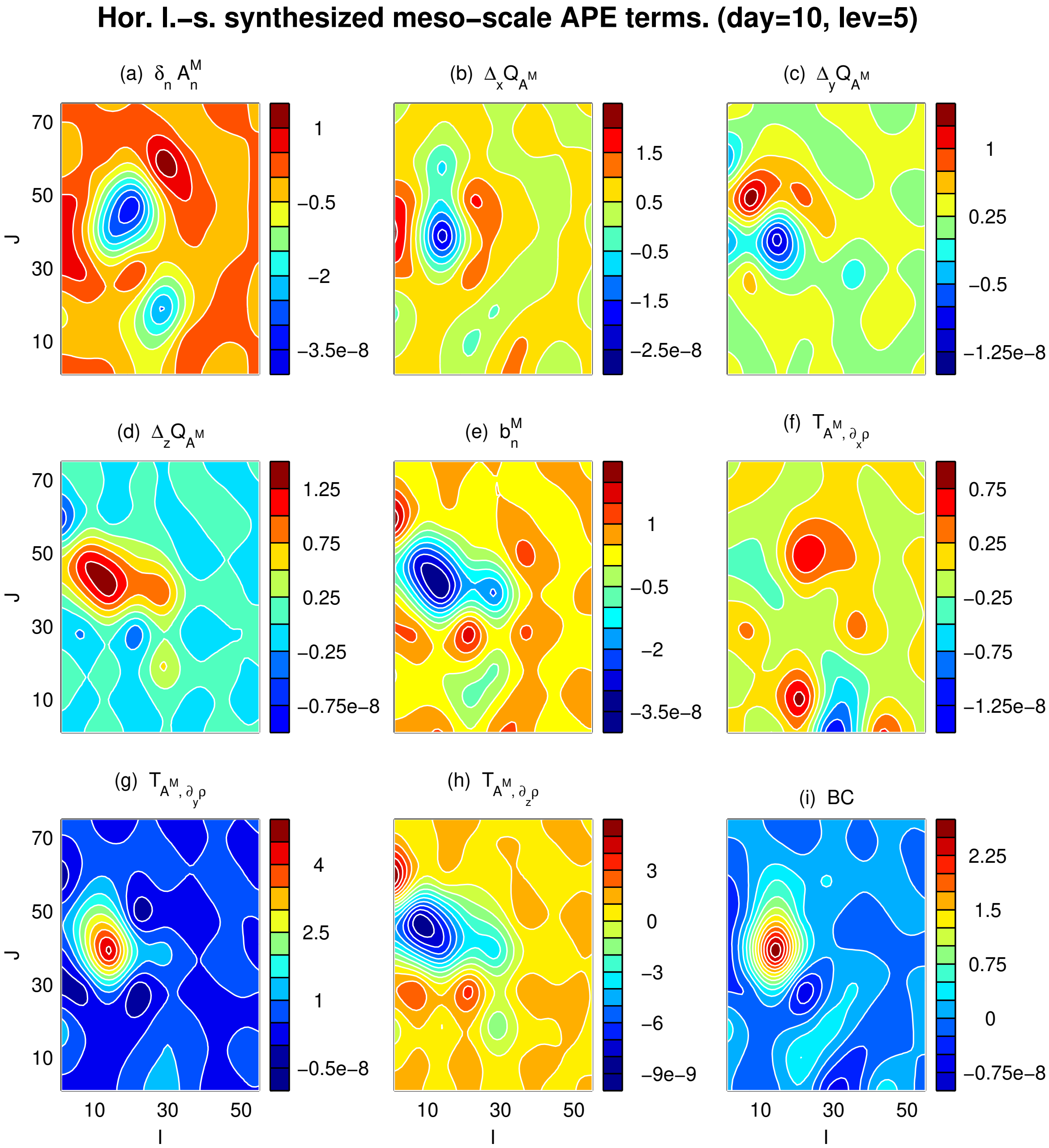

Figure 6.21: Same as Fig. 6.16, but for day 10.

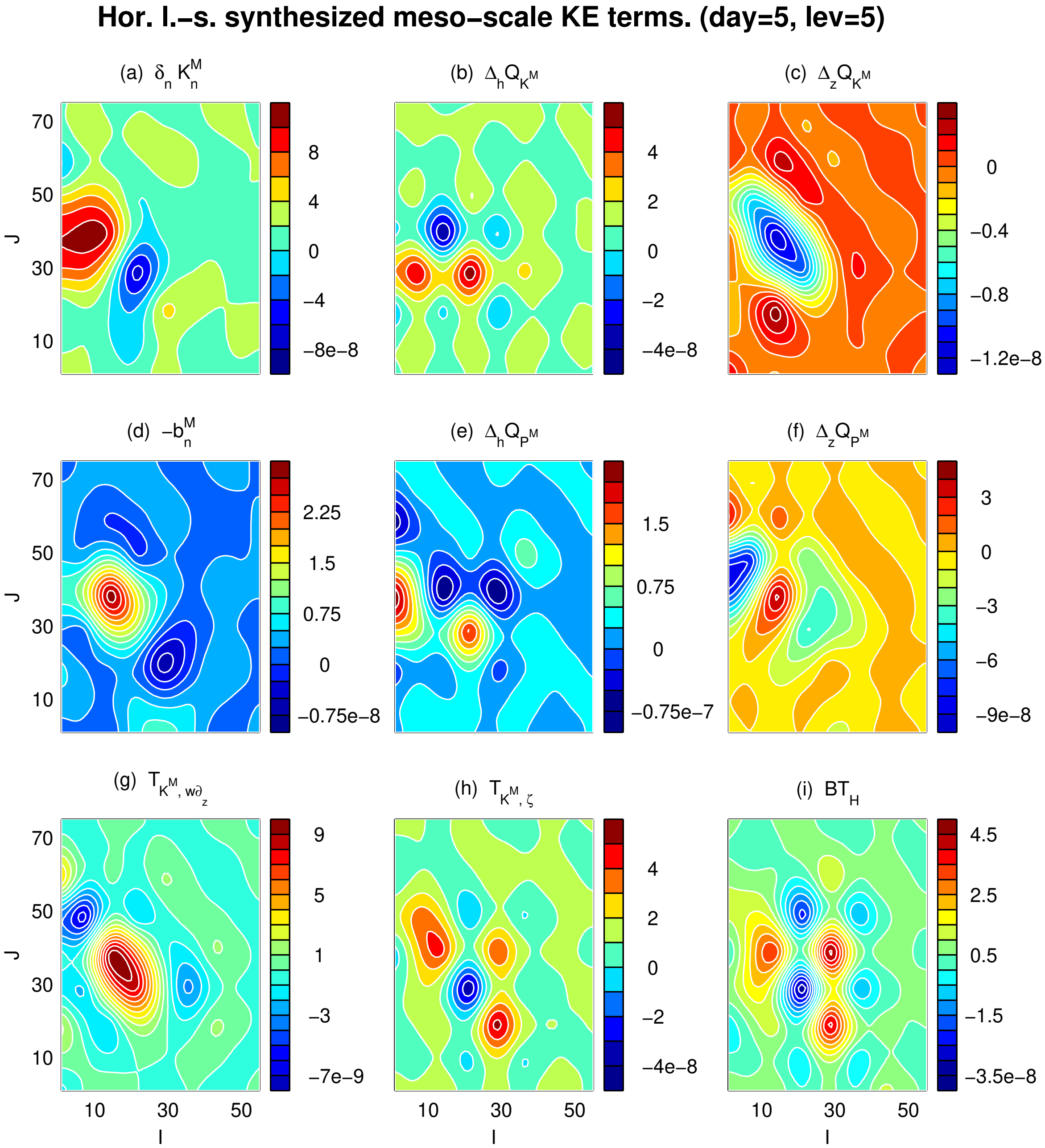

Figure 6.22: Locally averaged meso-scale kinetic energetics (in m^2s^{-3}) for day 5 at depth 125 m. The first eight subplots [(a) through (h)] show terms from the meso-scale KE equation (cf. Table 3.2), and the last one, BT_H, is related to the barotropic instability (cf. §4.3.4). All terms are supposed to be locally averaged, which is a large-scale synthesis in the horizontal direction with $j_0^{sp} = 5$.

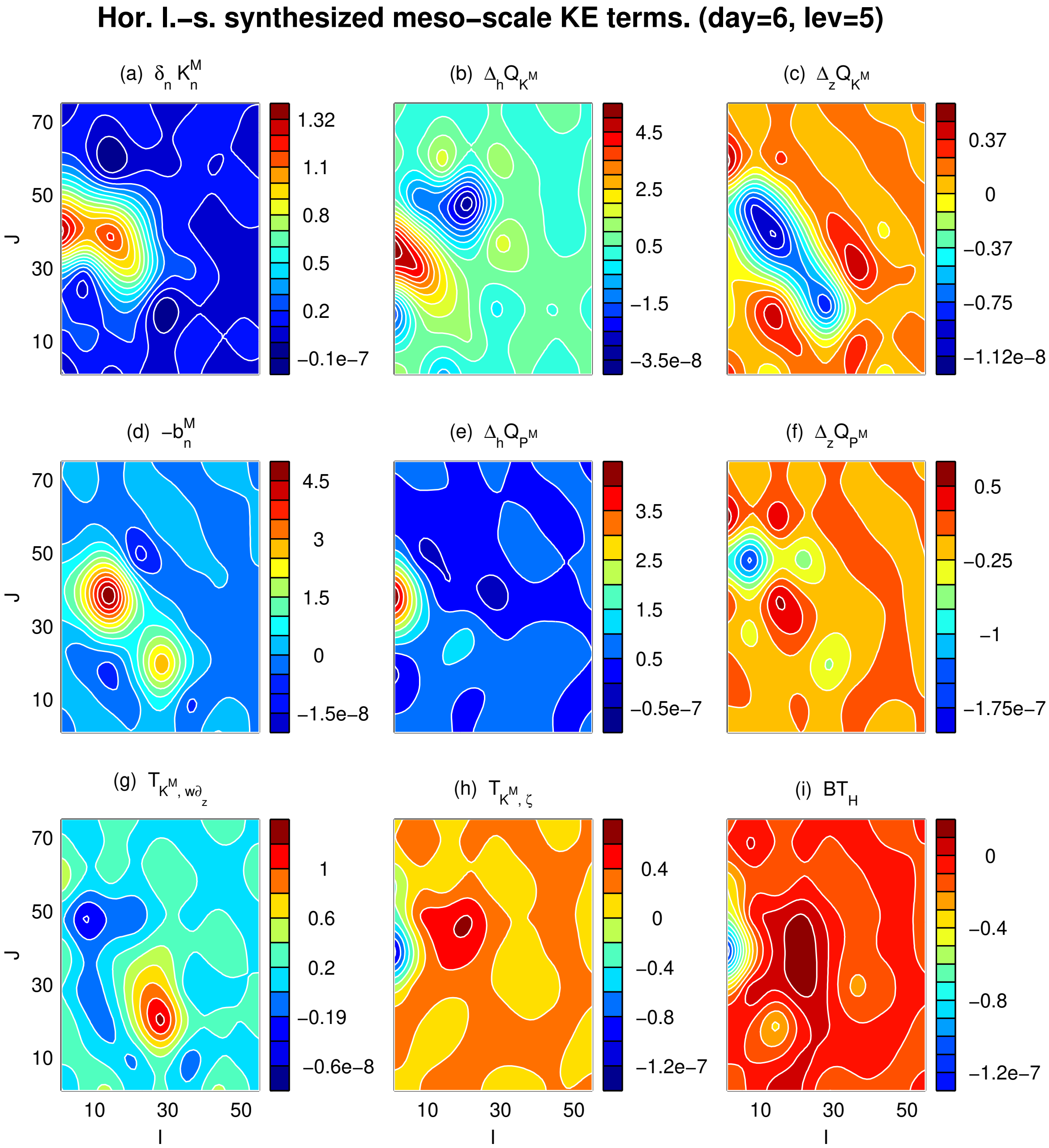

Figure 6.23: Same as Fig. 6.22, but for day 6.

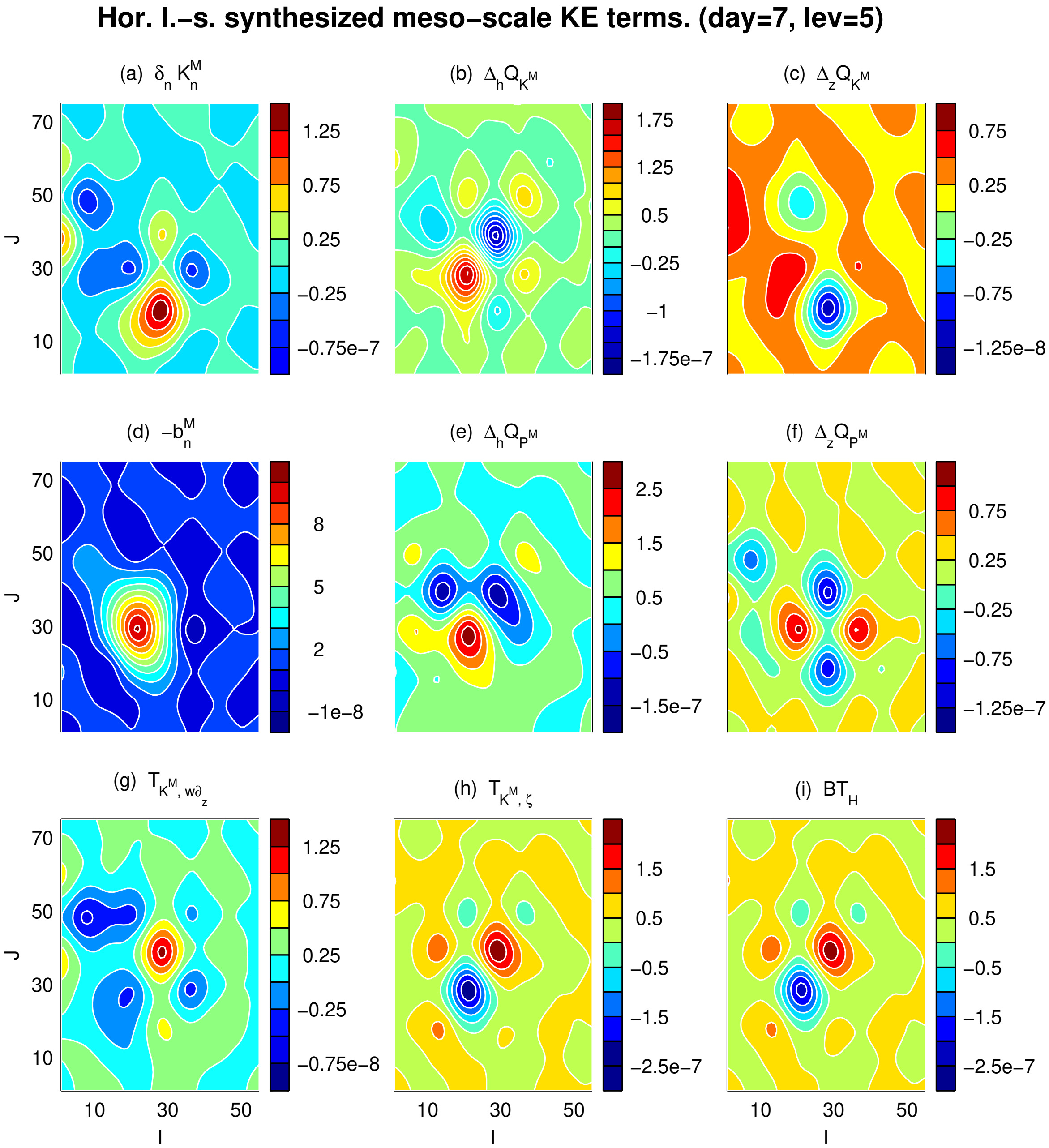

Figure 6.24: Same as Fig. 6.22, but for day 7.

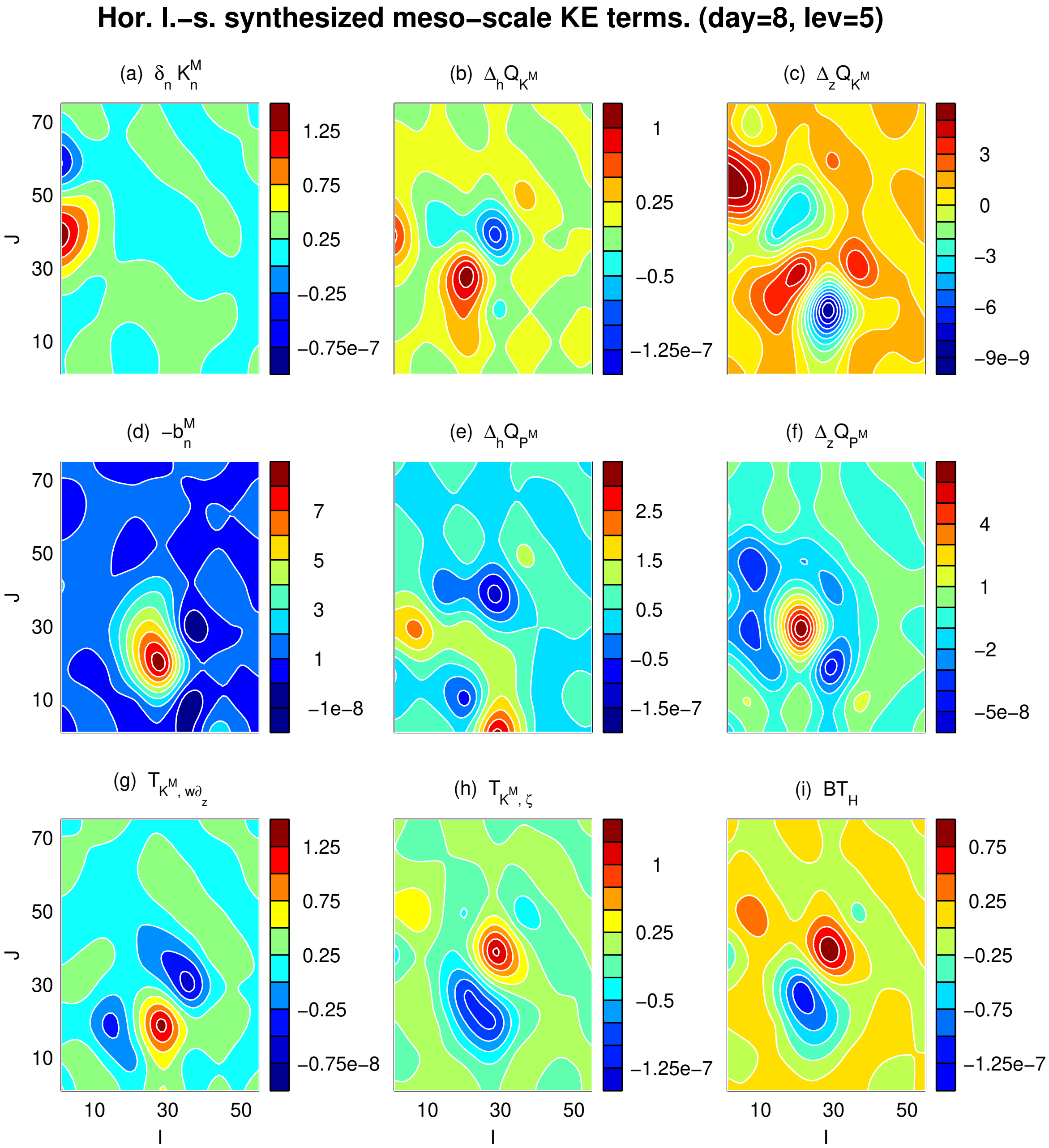

Figure 6.25: Same as Fig. 6.22, but for day 8.

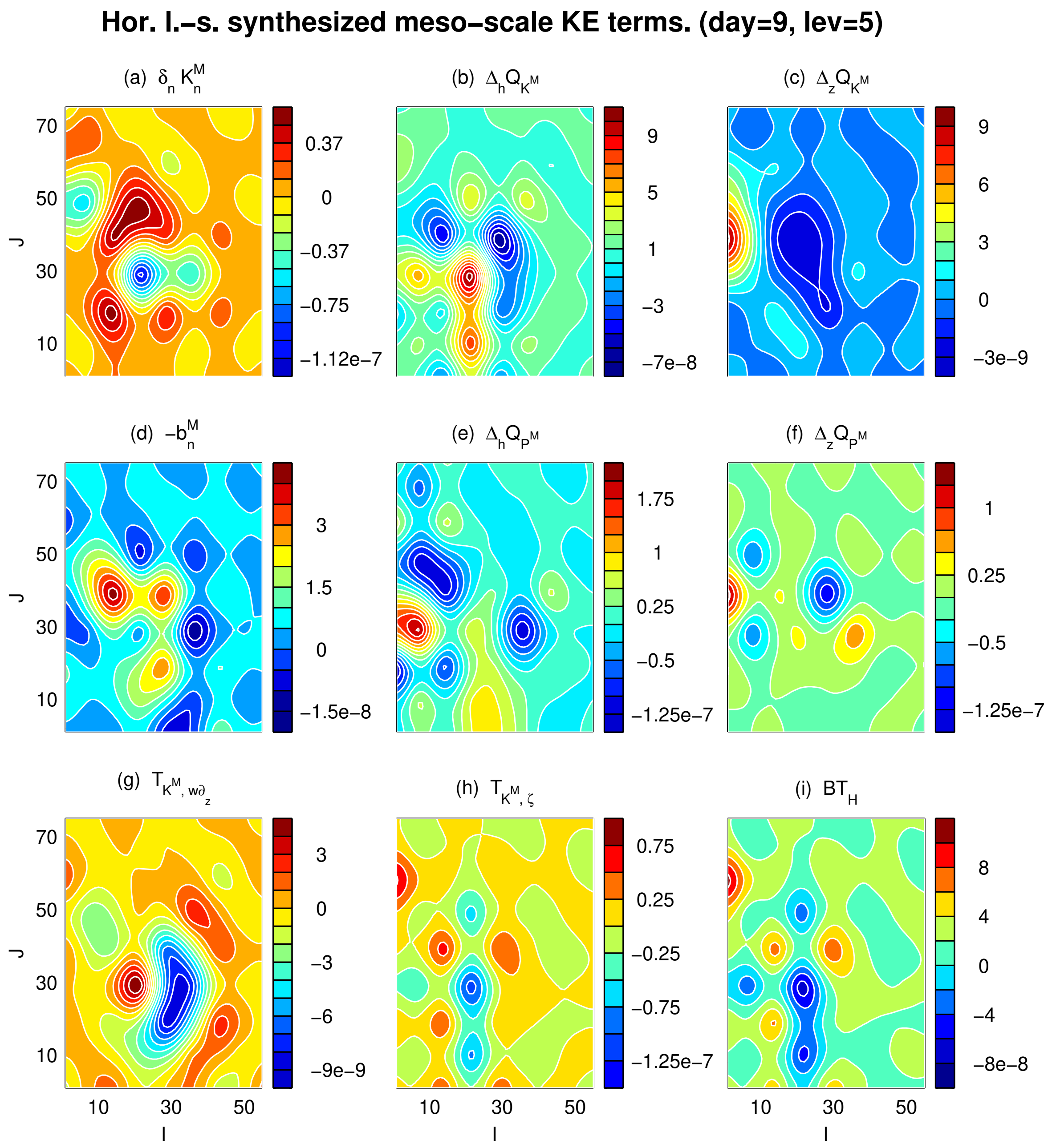

Figure 6.26: Same as Fig. 6.22, but for day 9.

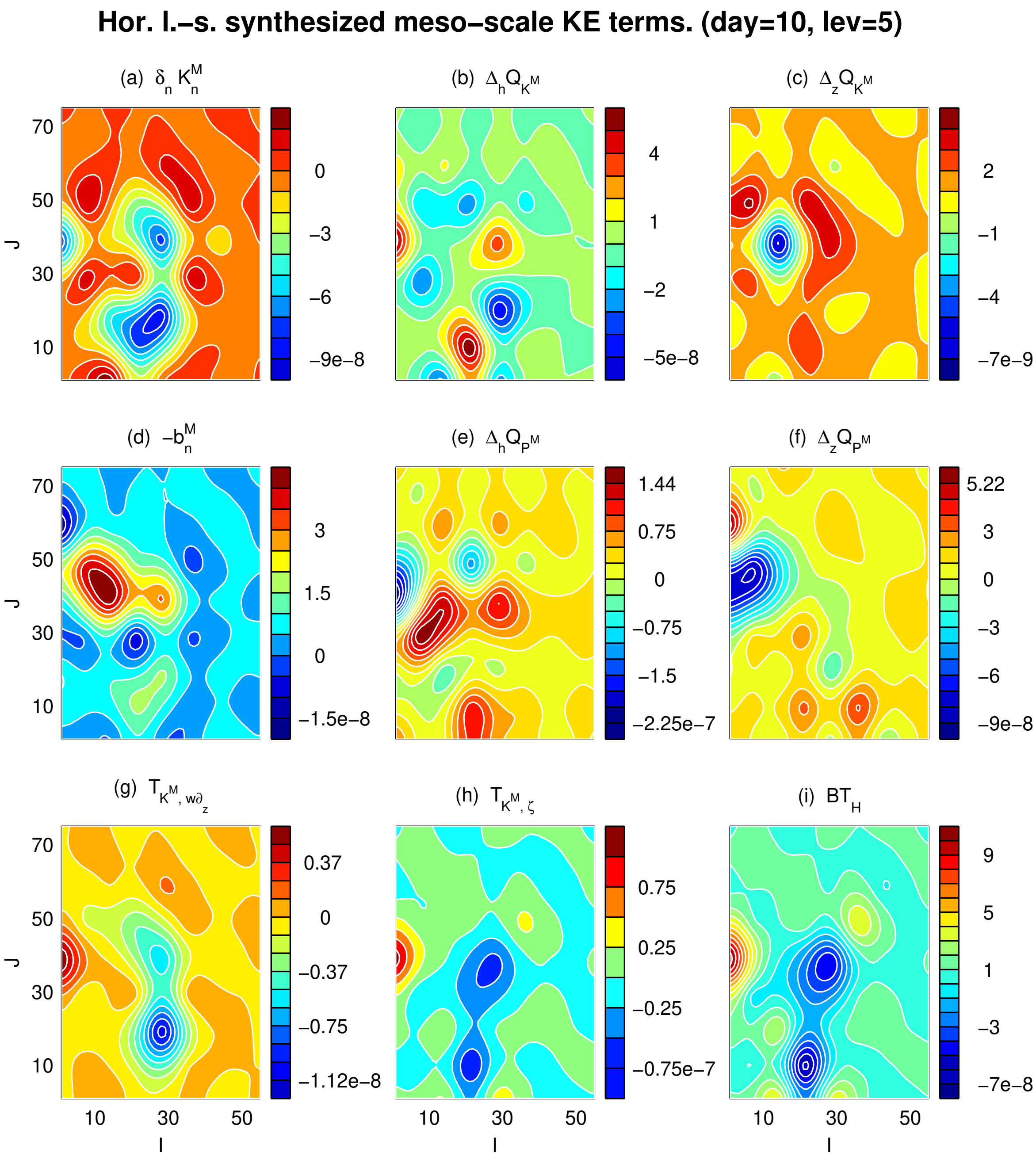

Figure 6.27: Same as Fig. 6.22, but for day 10.

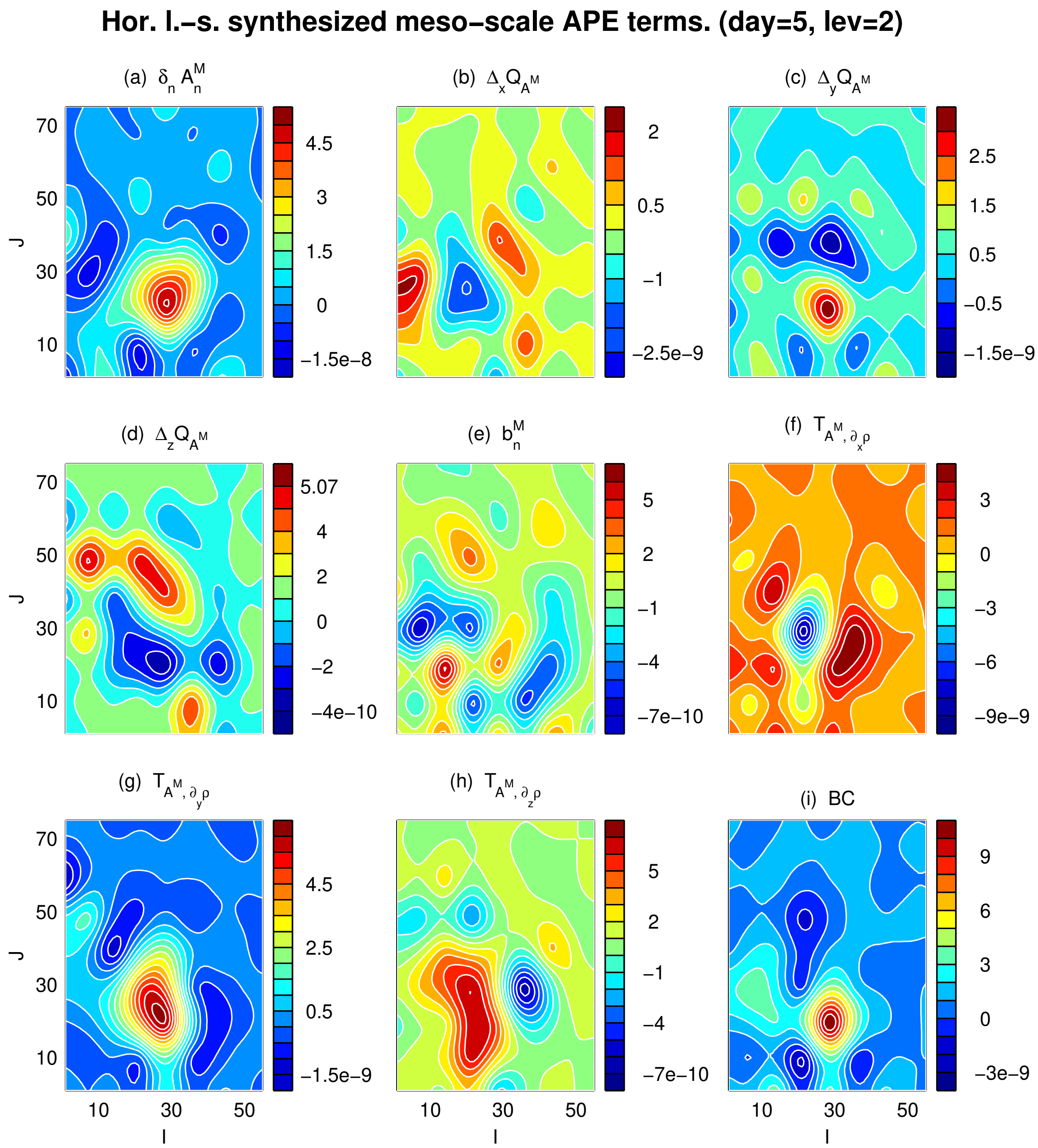

Figure 6.28: Same as Fig. 6.16, but for depth 25 m. Terms which are not significant are masked up.

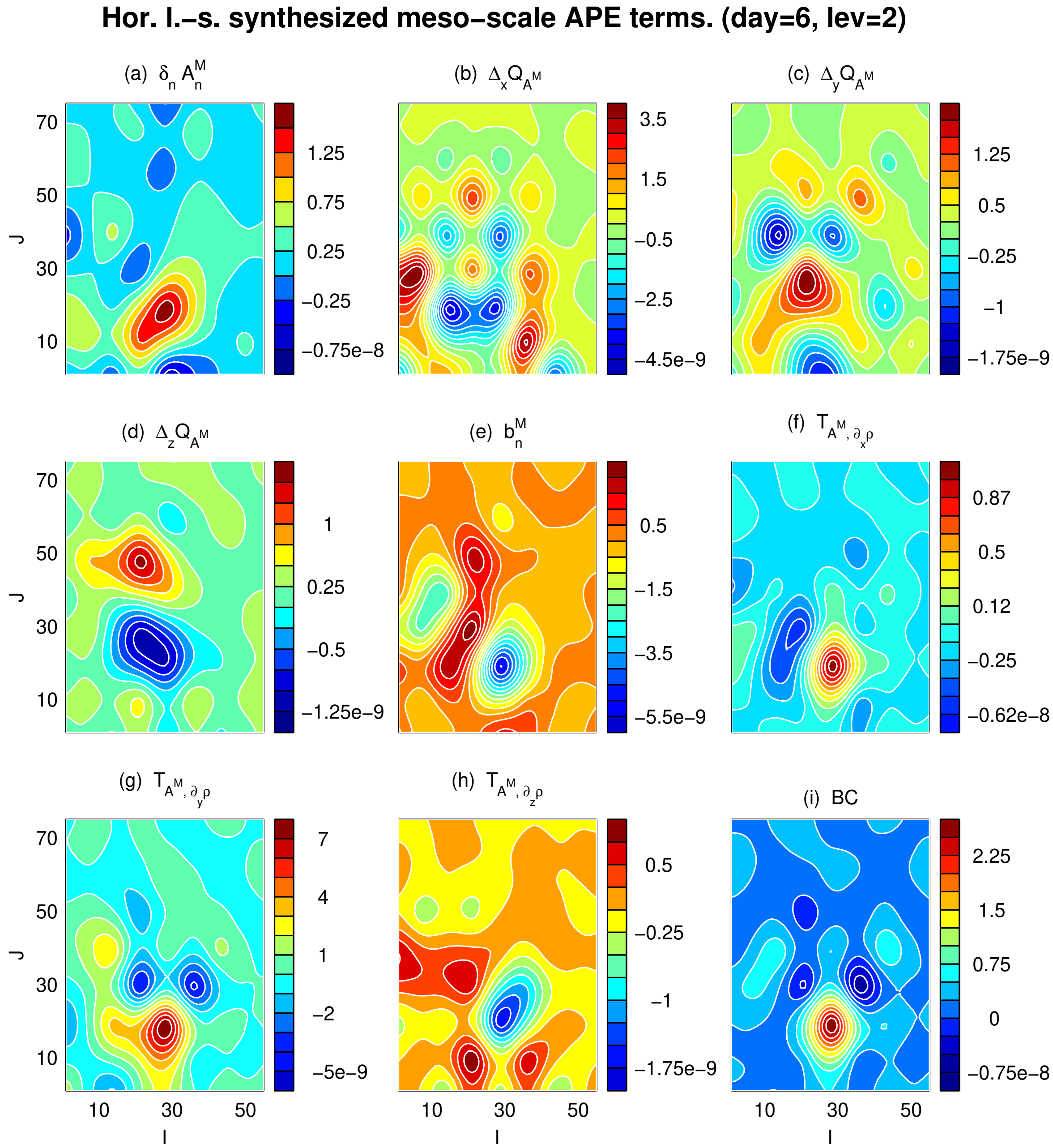

Figure 6.29: Same as Fig. 6.28, but for day 6.

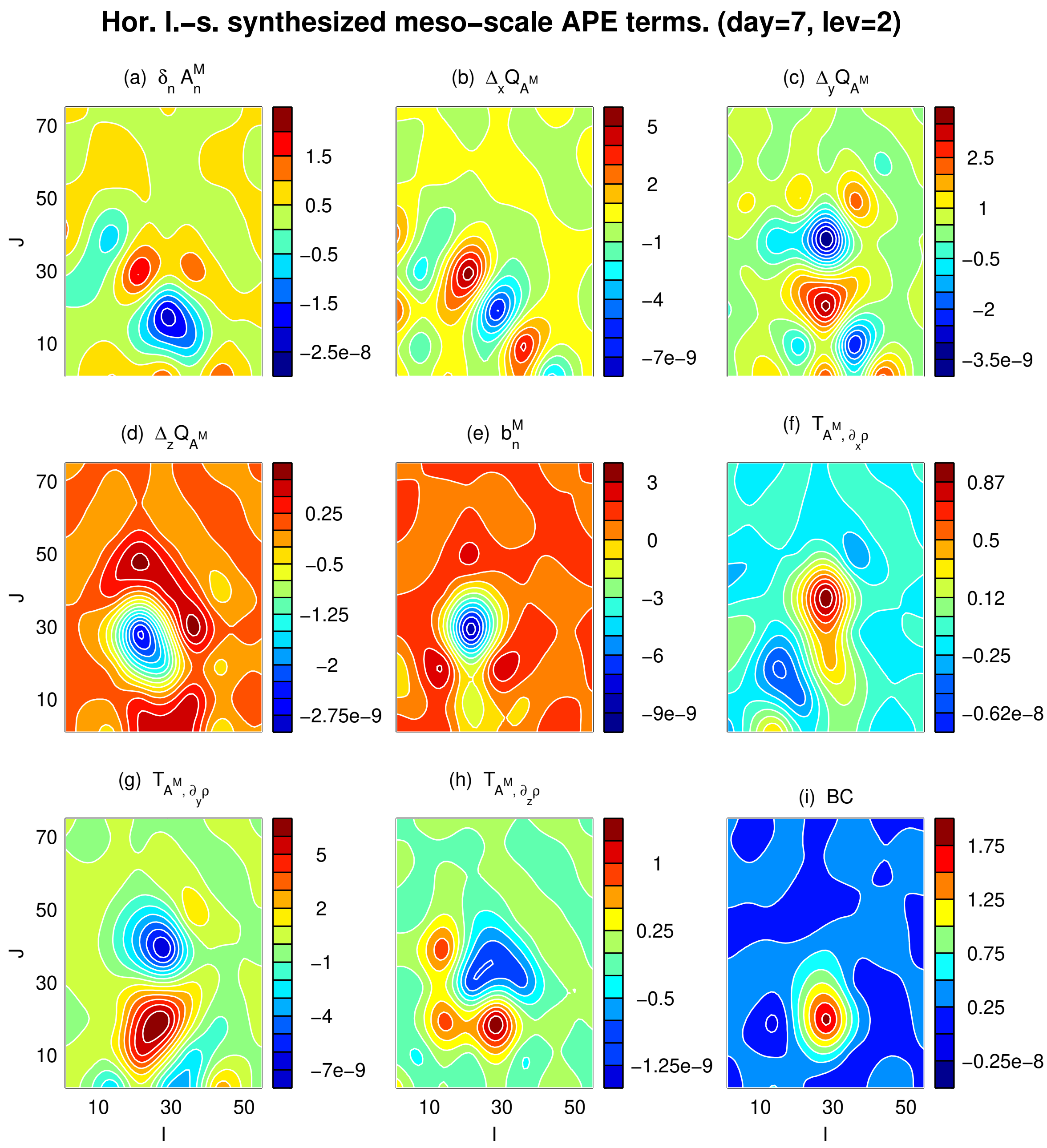

Figure 6.30: Same as Fig. 6.28, but for day 7.

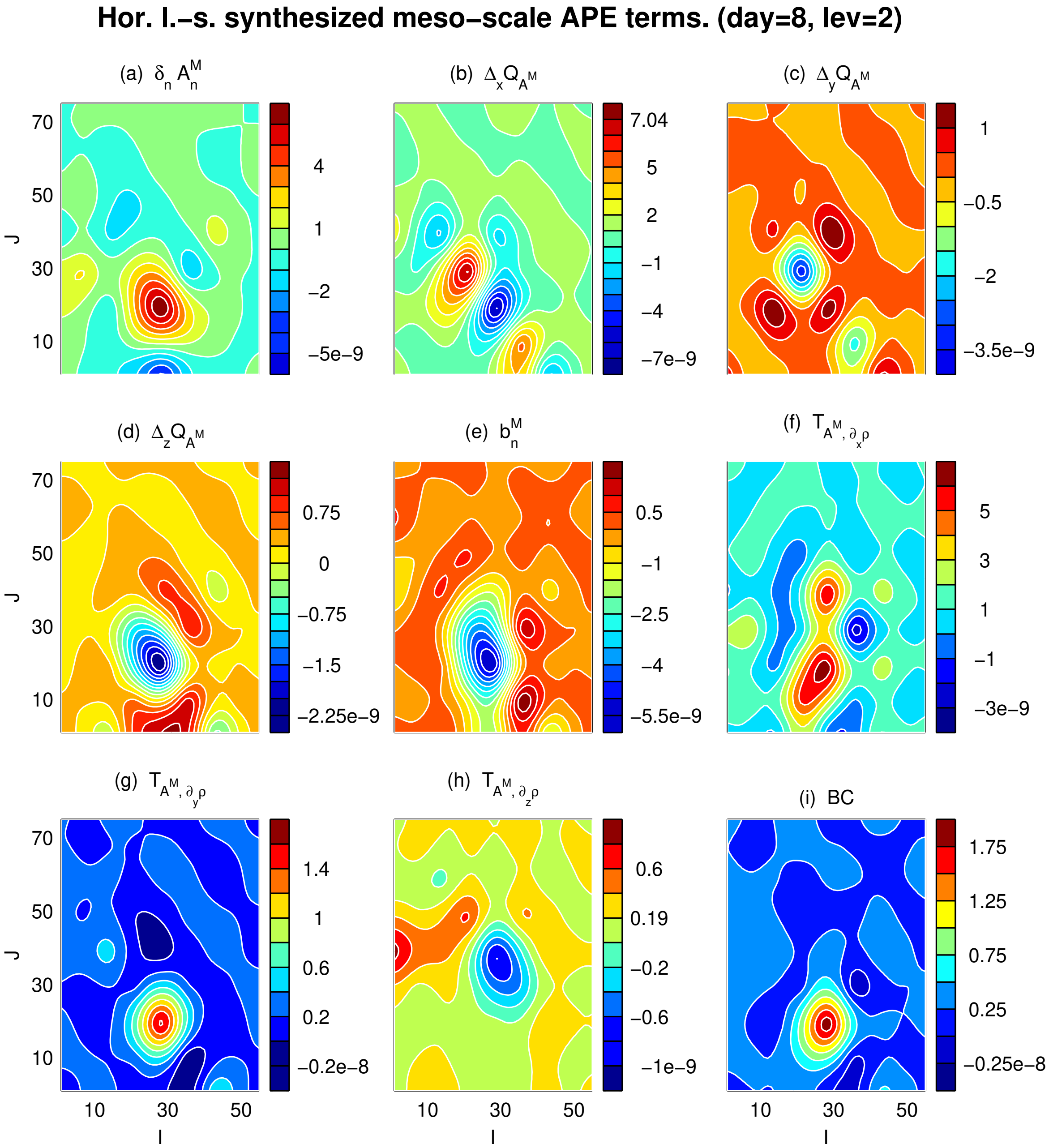

Figure 6.31: Same as Fig. 6.28, but for day 8.

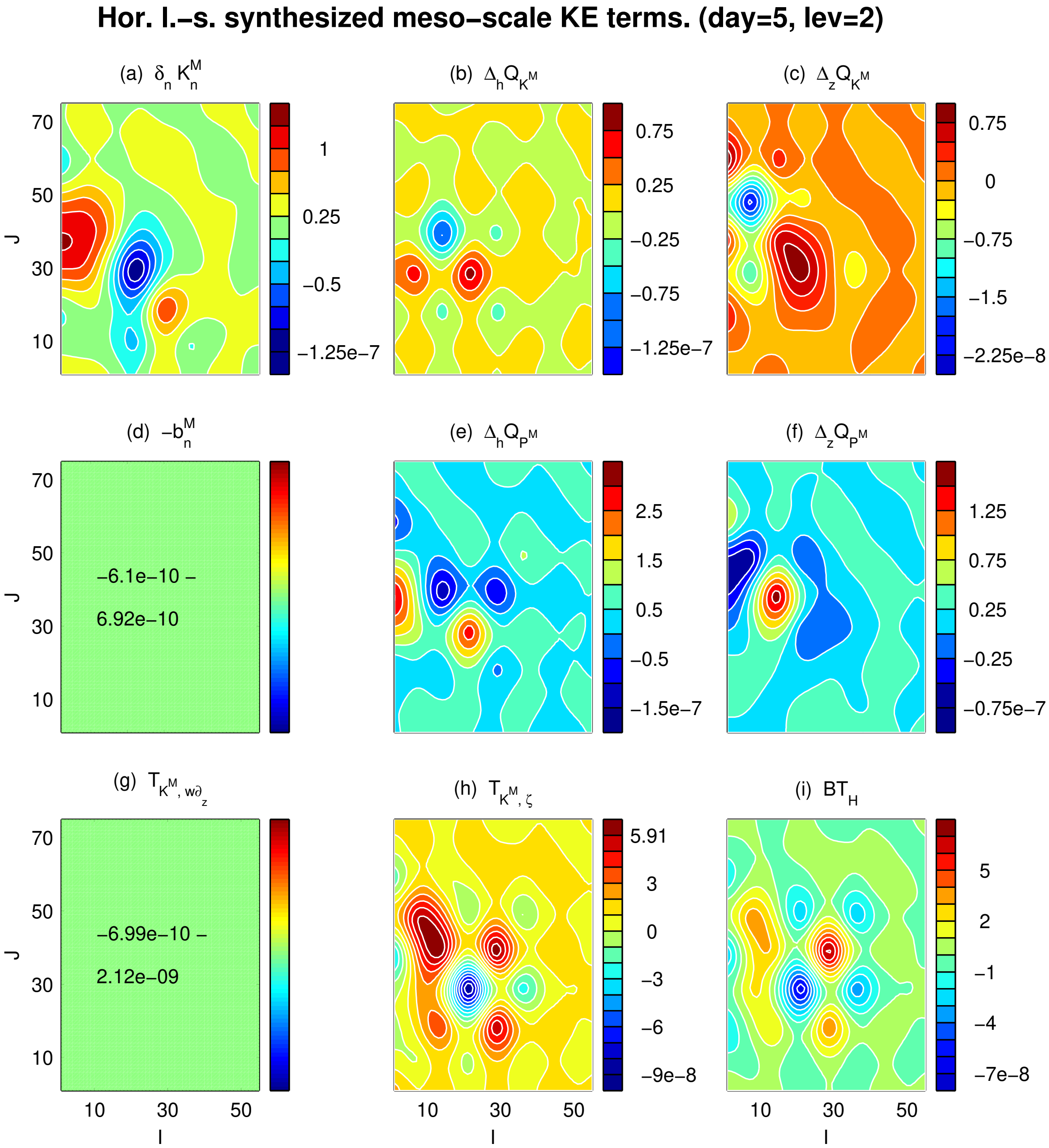

Figure 6.32: Same as Fig. 6.22, but for depth 25 m.

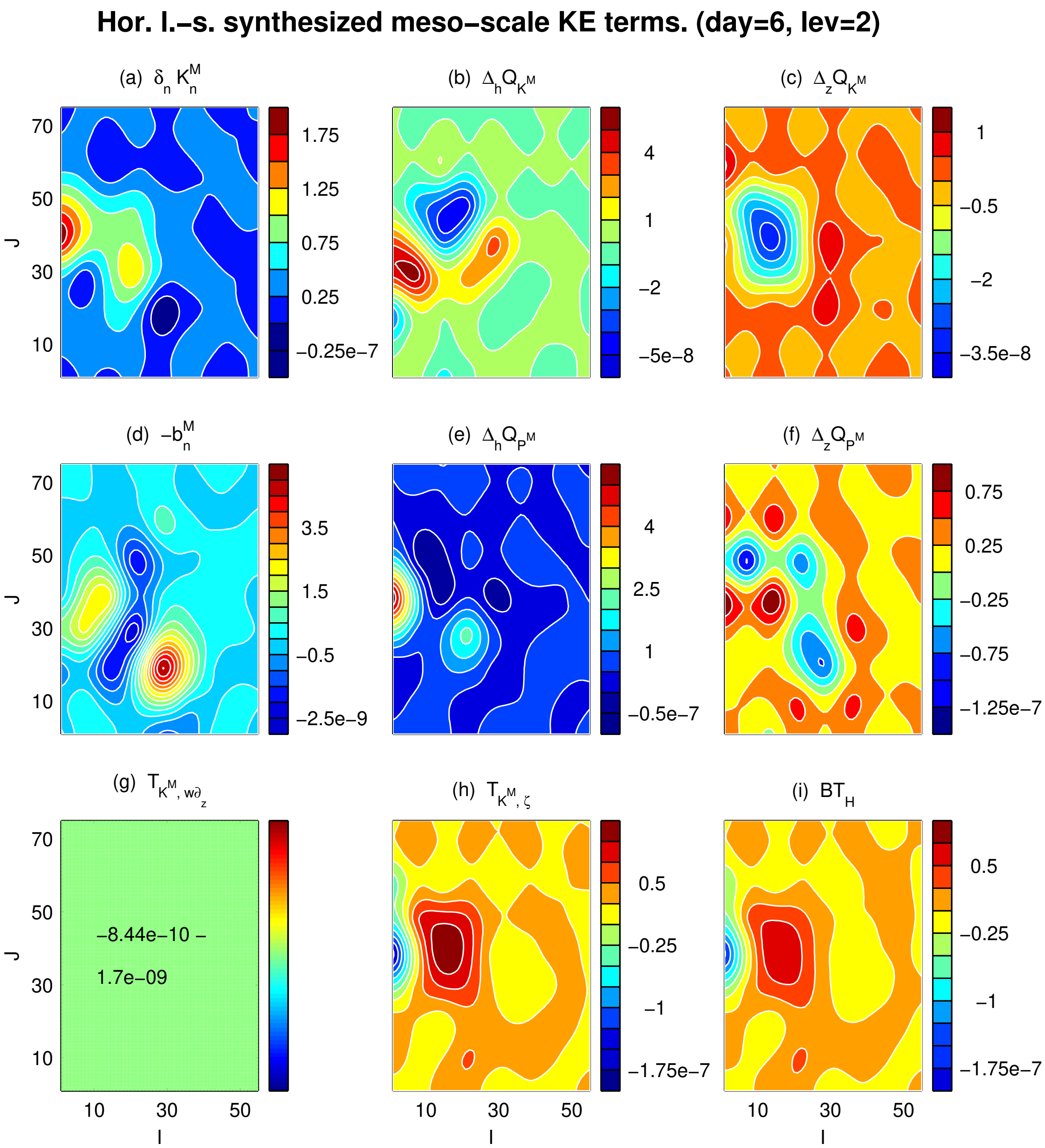

Figure 6.33: Same as Fig. 6.32, but for day 6.

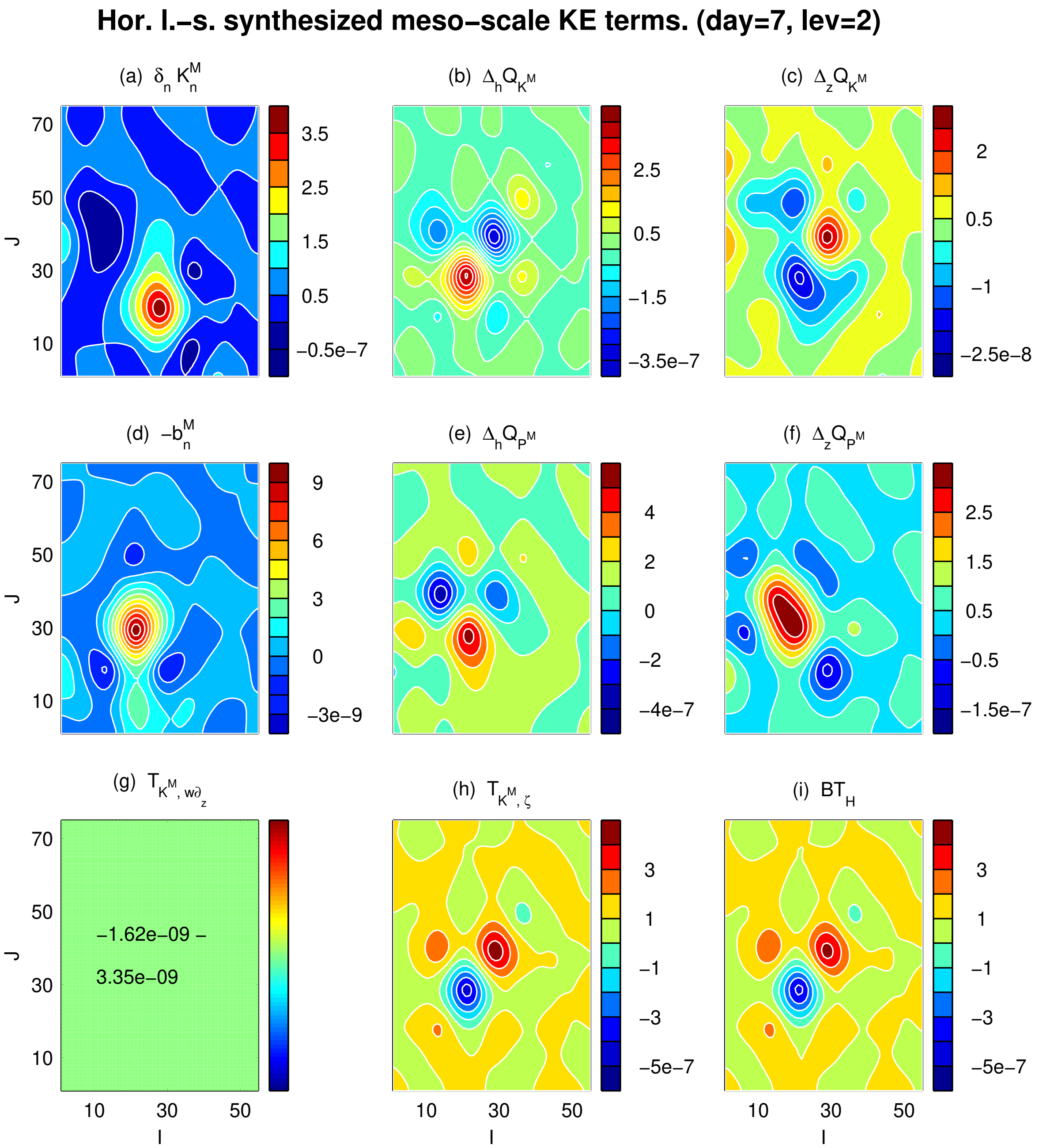

Figure 6.34: Same as Fig. 6.32, but for day 7.

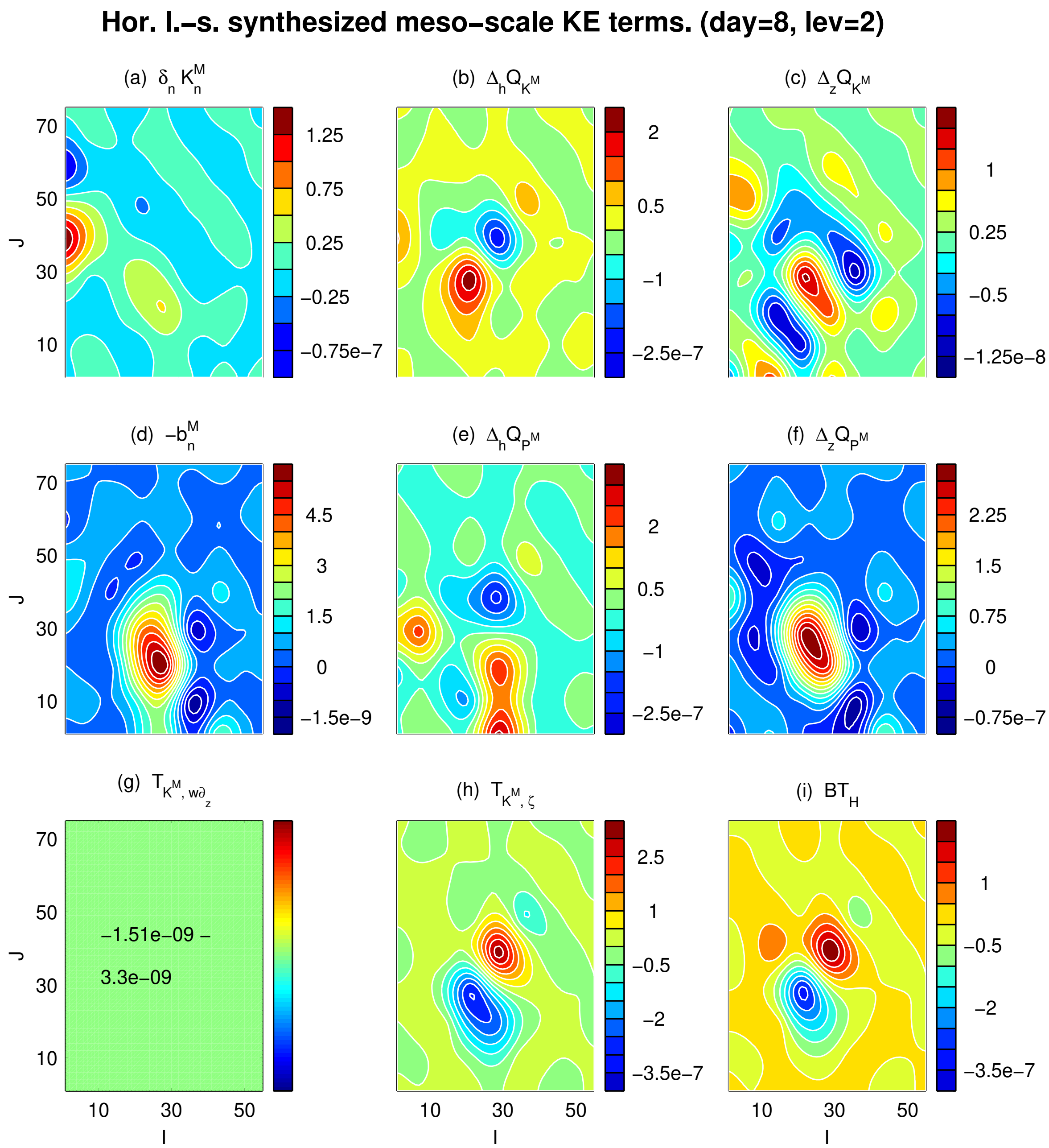

Figure 6.35: Same as Fig. 6.32, but for day 8.

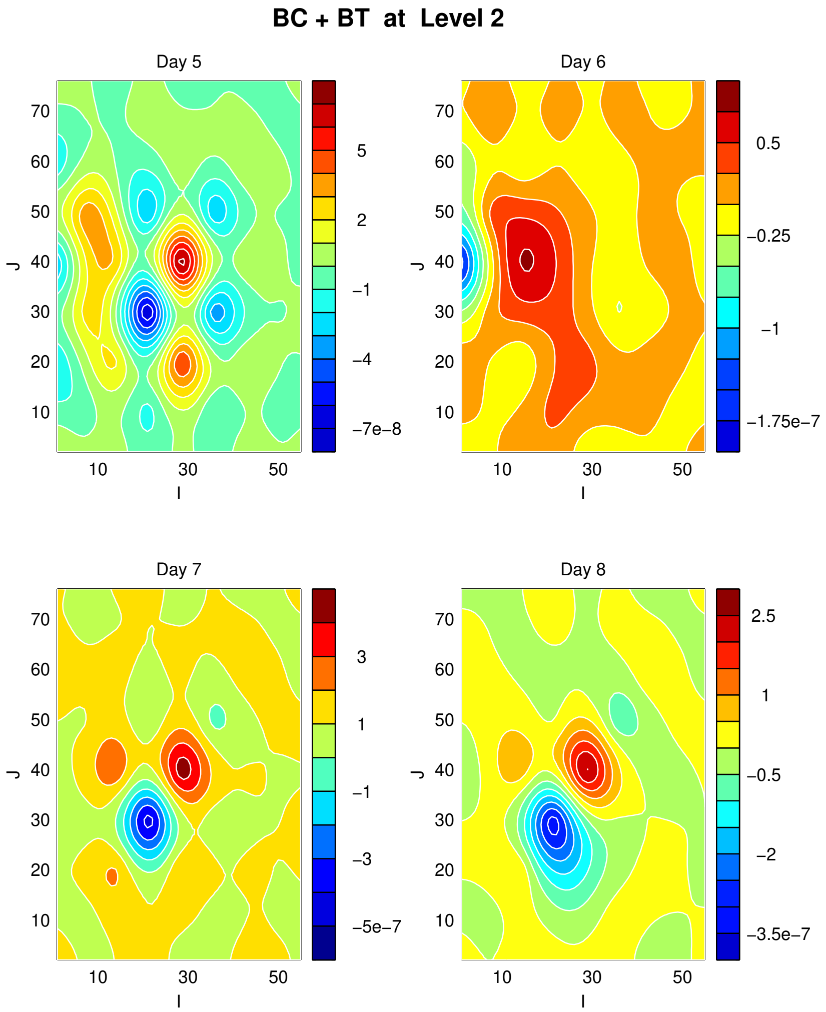

Figure 6.36: Total meso-scale energy transfer from the large-scale window at depth 25 m.

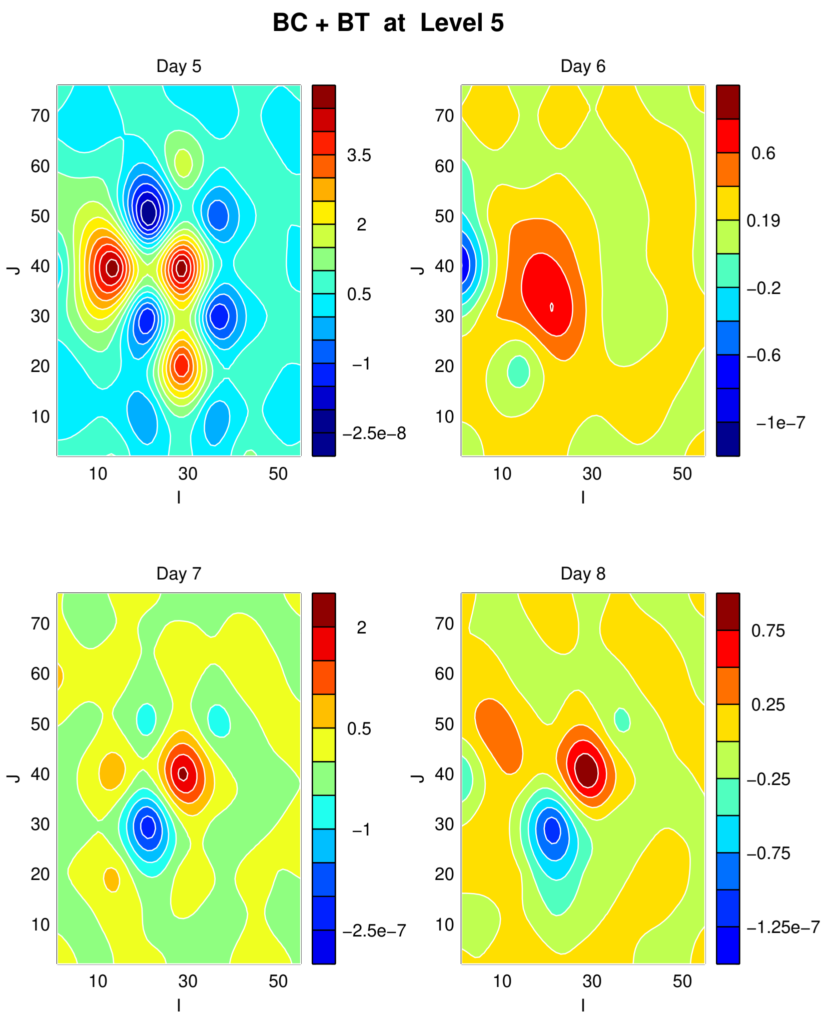

Figure 6.37: Total meso-scale energy transfer from the large-scale window at depth 125 m.

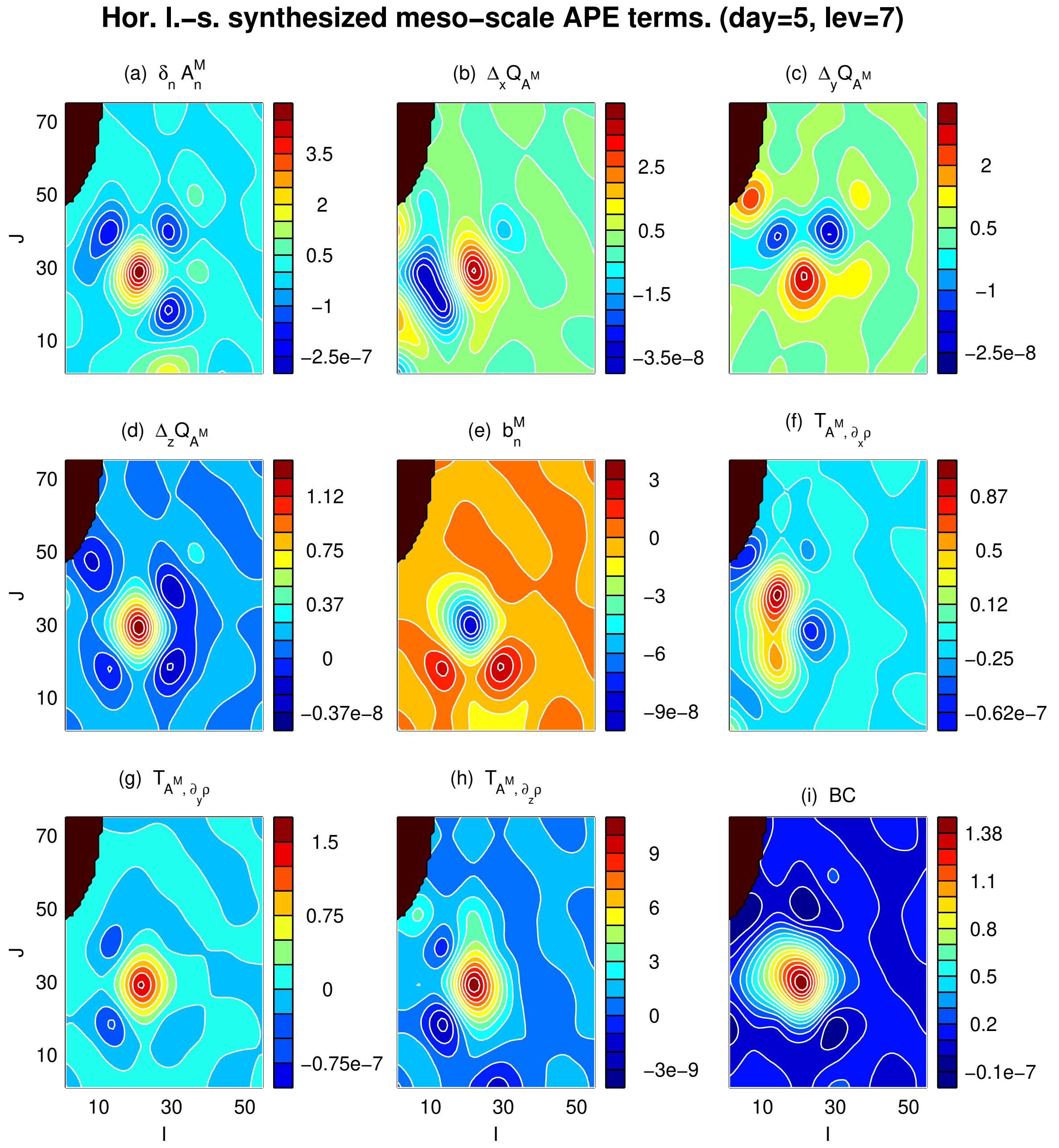

Figure 6.38: Same as Fig. 6.16, but for depth 300 m.

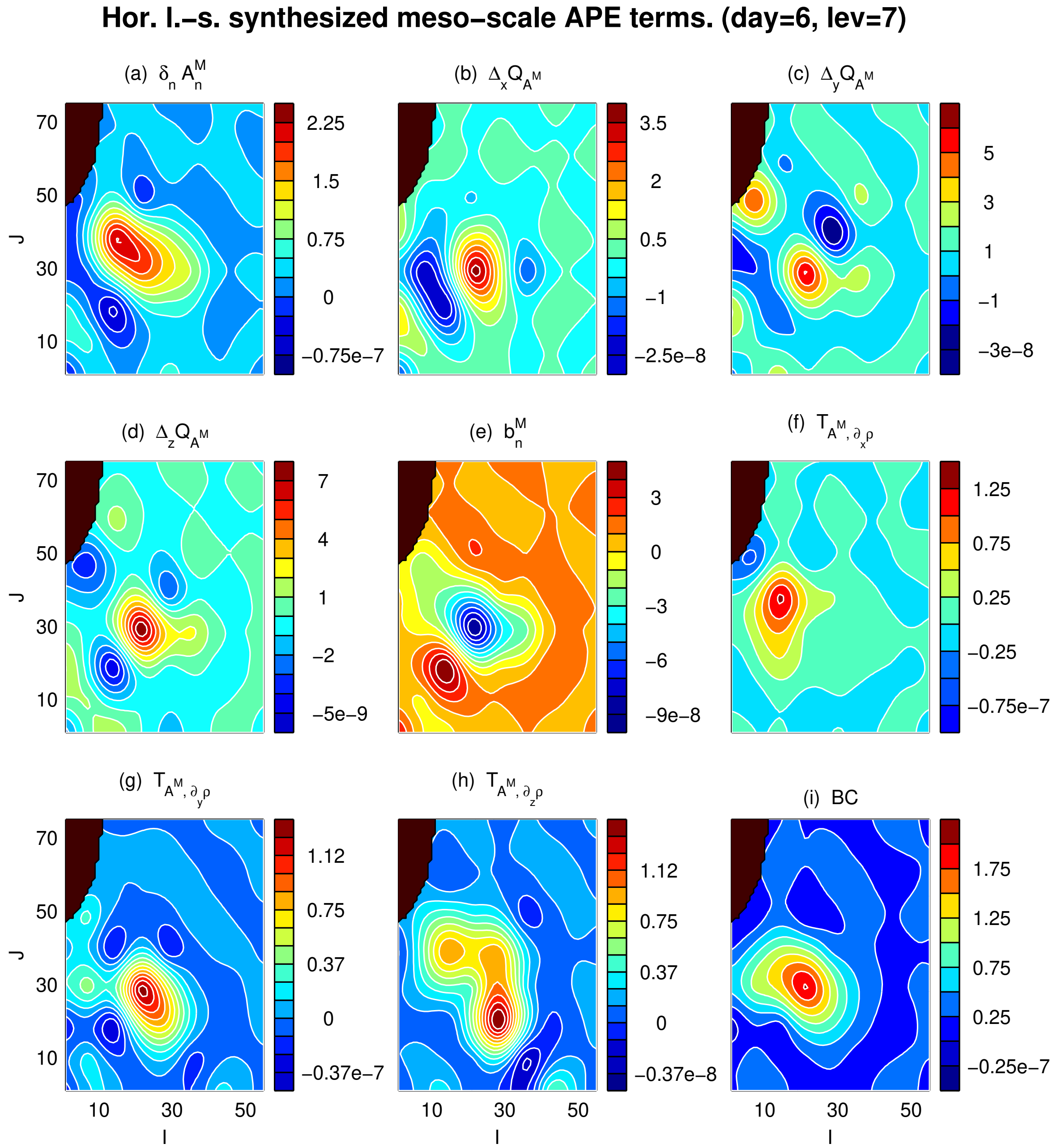

Figure 6.39: Same as Fig. 6.38, but for day 6.

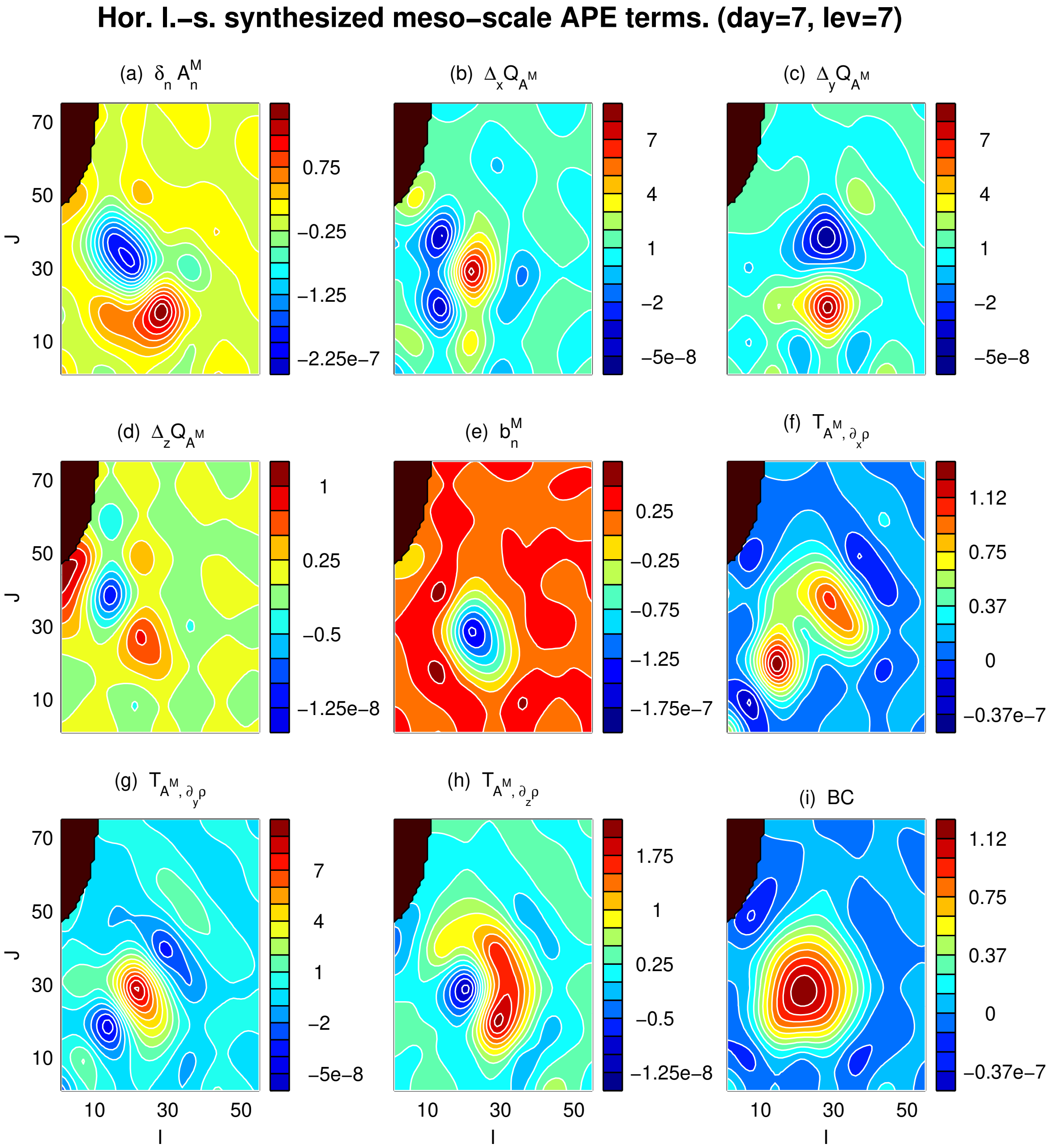

Figure 6.40: Same as Fig. 6.38, but for day 7.

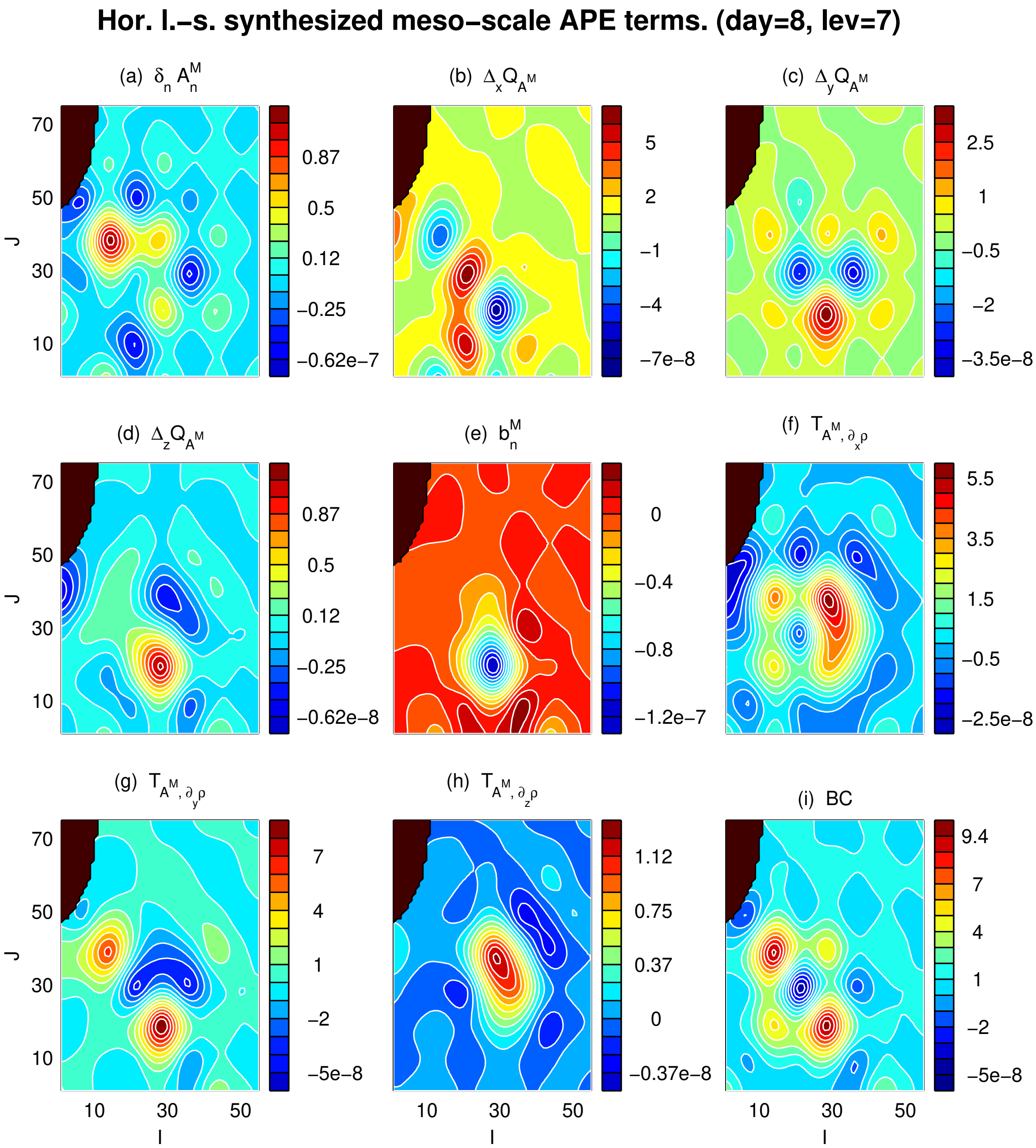

Figure 6.41: Same as Fig. 6.38, but for day 8.

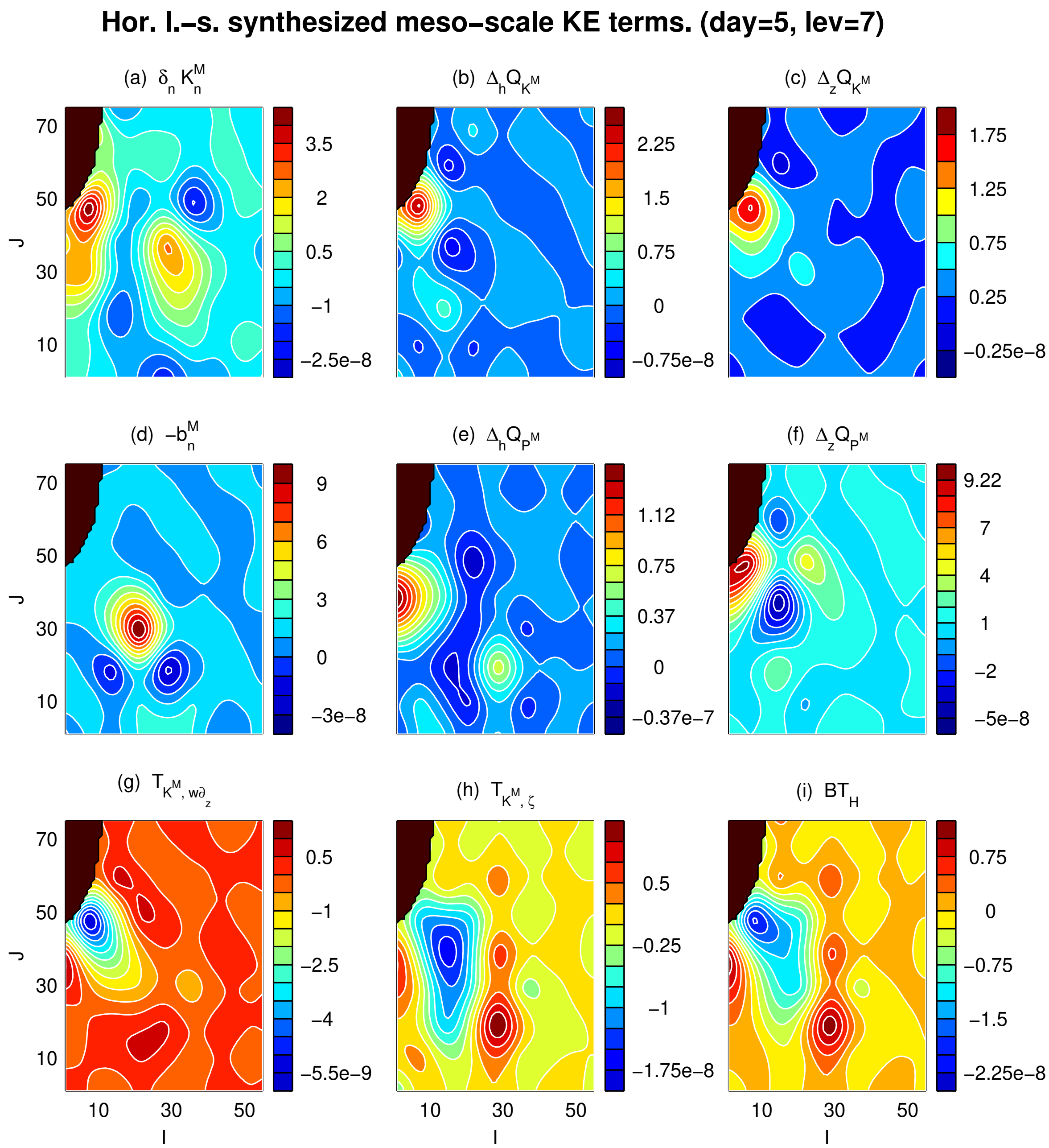

Figure 6.42: Same as Fig. 6.22, but for depth 300 m.

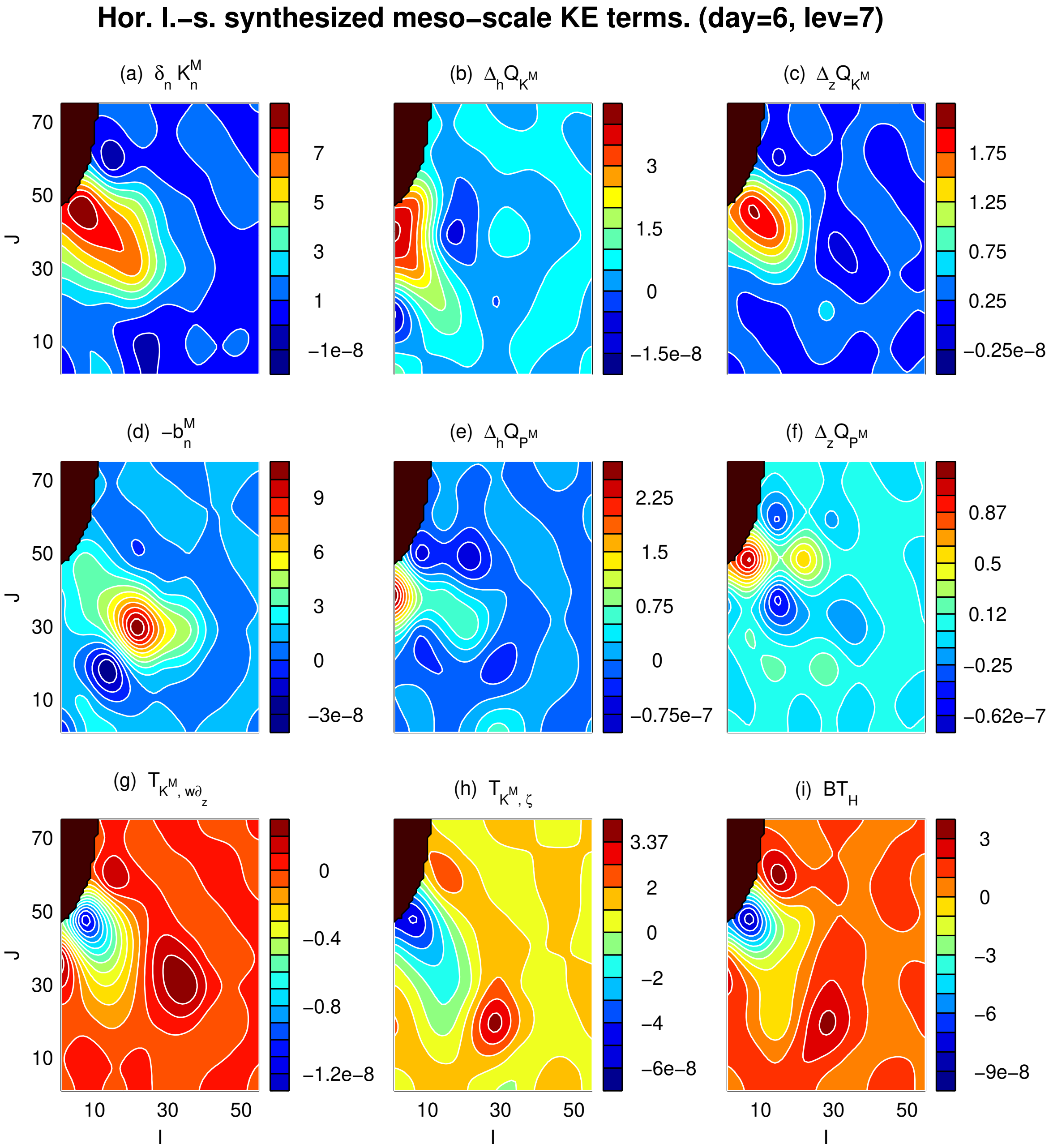

Figure 6.43: Same as Fig. 6.42, but for day 6.

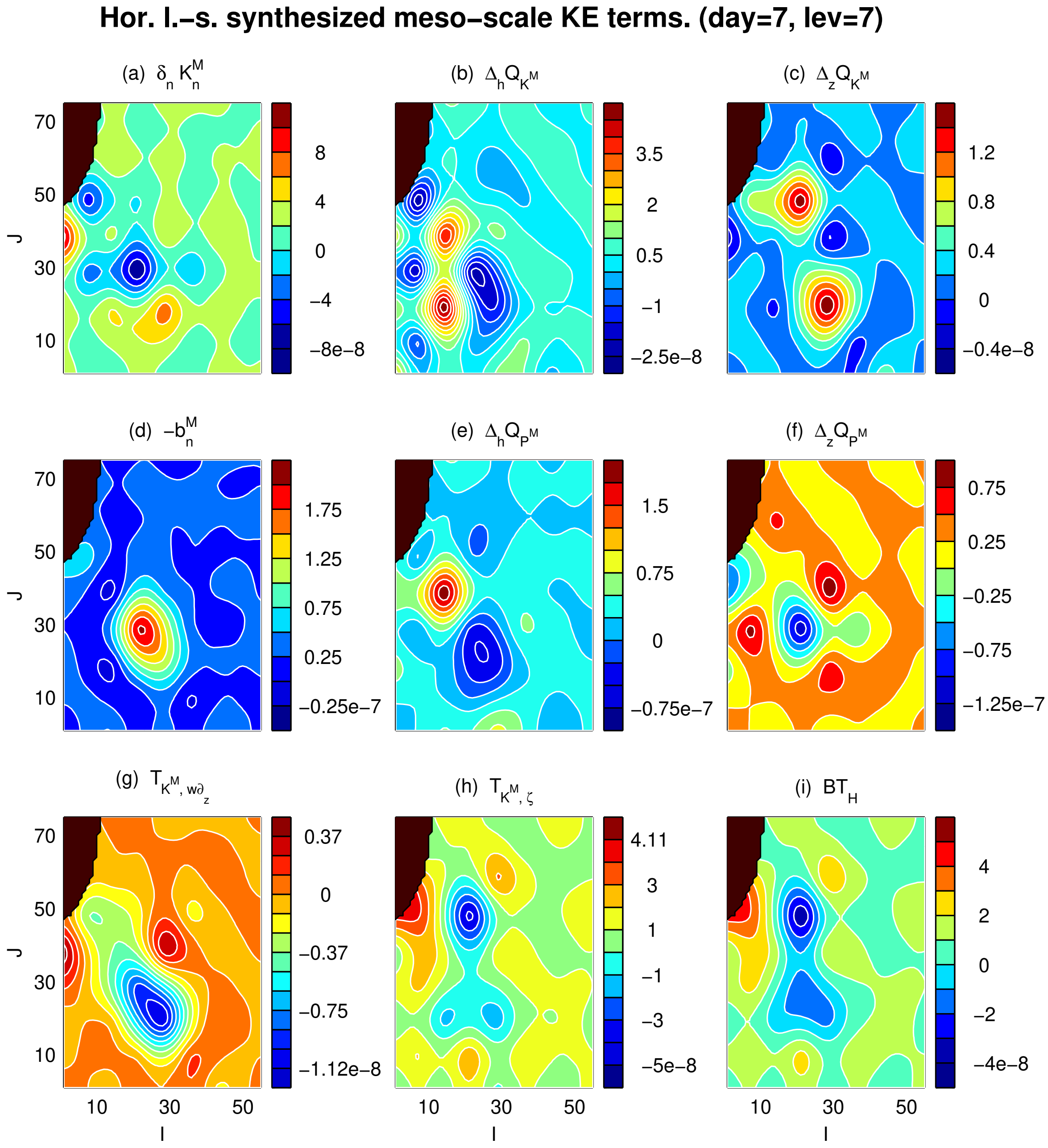

Figure 6.44: Same as Fig. 6.42, but for day 7.

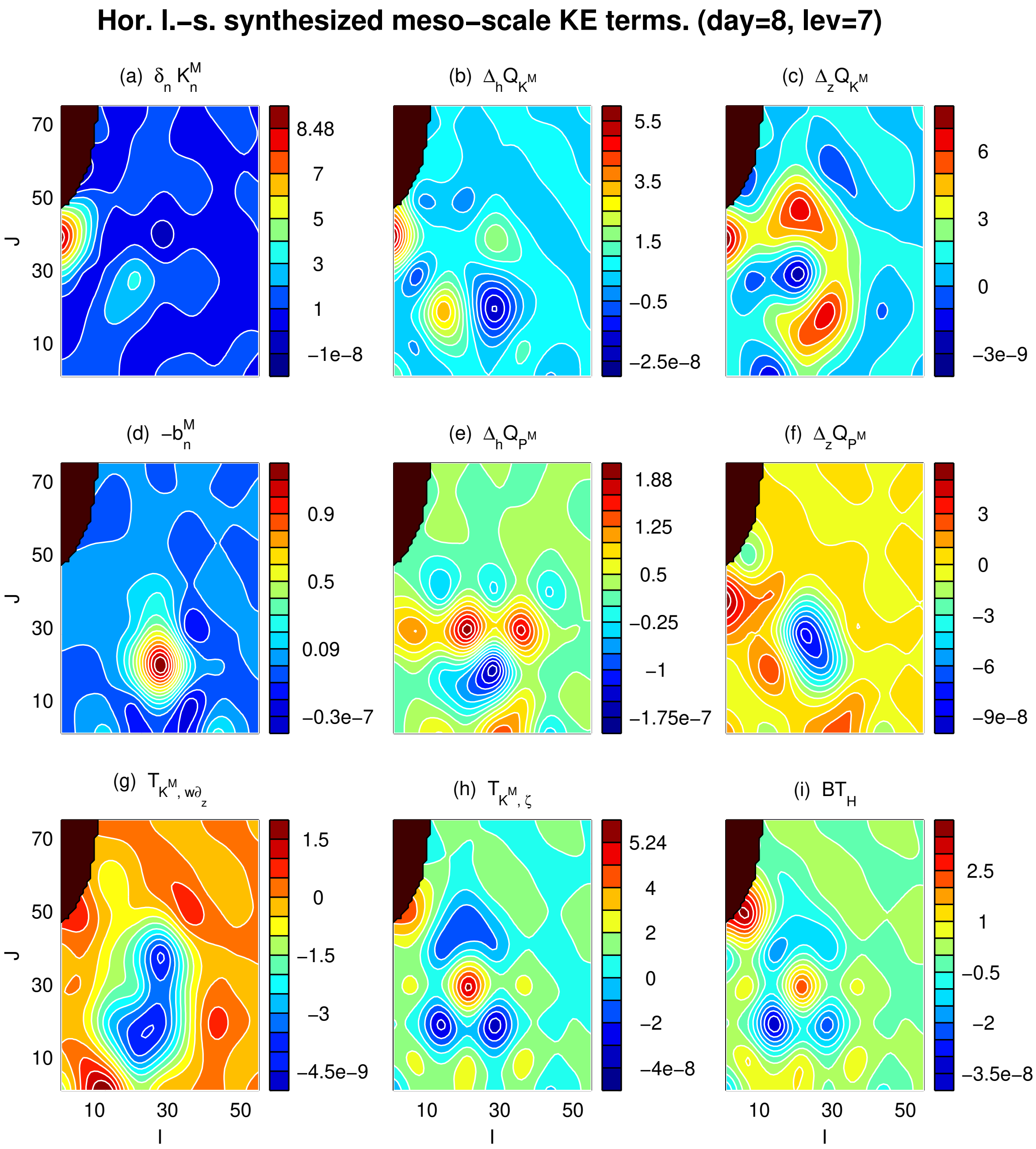

Figure 6.45: Same as Fig. 6.42, but for day 8.

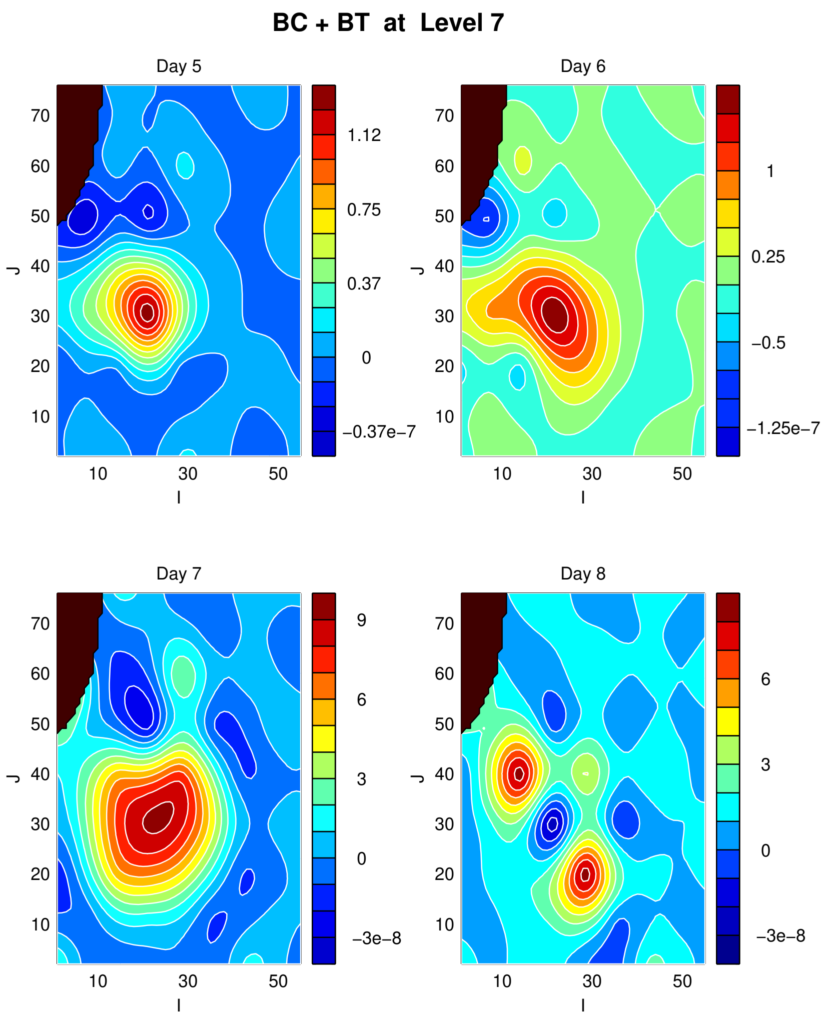

Figure 6.46: Total meso-scale energy transfer from the large-scale window at depth 300 m.

6.3.3 More about the baroclinic instability

A conspicuous feature identified from the above horizontal energetic maps is the solitary center of potential energy transfer from the large-scale window to the meso-scale window. For the deep layers, as we have shown, this means that the system is undergoing a baroclinic instability. But for the surface layers, the scenario is complicated by the kinetic energetics, and the MS-EVA expression as a whole does not support an Eady-like instability above some depth. Apparently, many things, particularly the vertical structure of the process, are still not clear. We need to find out the vertical extent of the baroclinic instability and collect more evidence to substantiate our previous claim.

We use a sectional distribution to illuminate this problem. For the sake of convenience, we want to make it comparable with the one we have done before for the Eady model. A section for this purpose is generally hard to choose, however, as the front is not a straight one here. We hence analyze a zonal section and a meridional section instead. These sections are chosen so as to cross the positive transfer center on the BC maps. For simplicity, we pick only day 7 for the study. A very strong signal of baroclinic instability has been identified on the horizontal maps at that time.

Figure 6.47 is the distribution of a selected collection of energetics on the zonal section J=30. The two instability indicators, BC and BT_H, are drawn in subplots (b) and (d), respectively. The other two are the buoyancy conversion b_n^M (a) and the vertical pressure working rate $\Delta_z Q_{PM}$ (c). Clearly, the positive BC is concentrated below 150 m (approximately), and maximized at 300 m from the available calculation. In the upper layers (above 150 m), there is a negative barotropic transfer center between I=15-25, and away from that the distribution of BT_H is rather plain. This vertical dichotomy is especially clear in Fig. 6.47c. A depth roughly at 150 m separates the $\Delta_z Q_{PM}$ profile near the intrusion area into two regimes, with the lower one negative in value, and the upper one in the opposite. Unfolded on horizontal maps, the negative regime corresponds to the b_n^M center, as well as the the BC hotspot.

On the meridional section I=20 (Fig. 6.48) the distribution is similar to that of Fig. 6.47: The surface regime is occupied by a center with negative BT_H and positive $\Delta_z Q_{PM}$, while the lower regime is characterized by a larger positive BC pool with negative b_n^M and $\Delta_z Q_{PM}$. By the Eady model results, this scenario implies strongly the existence of a baroclinic instability.

Recall that the Eady model also shows a vertically uniform distribution for the horizontally averaged BC (see Fig. 4.11d). The upward extent of the BC here then tells that the baroclinic instability is limited below a depth of approximately 150 m.

More evidence can be found for the baroclinic instability from the tilting pattern of b_n^M and $\Delta_z Q_{PM}$. By the Eady energetics, an unstable system has a buoyancy conversion and a vertical pressure work both distributed in a way tilting with height toward the left on a vertical section with the density gradient directing toward the paper (see Fig. 4.6). In the IFF region, the density field has a horizontal pattern which evolves generally as the temperature does (cf. Fig. 6.10), with the water from the north a little heavier than that from the south. This structure implies, if the system is baroclinically unstable, one should see a pattern of b_n^M and $\Delta_z Q_{PM}$ tilting with height toward the west on a zonal section. This scenario has been reproduced exactly as expected in Figs. 6.47(a) and (c), between the grid lines I=15 and I=35. On a meridional section, no general conclusion can be drawn for the pattern, as the background density distribution varies from location to location and from time to time. On day 7, the grid line I=20 happens to be near the western flank of the cold (and dense) tongue intrusion (see Fig. 6.10), the density gradient is thence outgoing from the paper. The corresponding tilting structure of buoyancy conversion and vertical pressure work then should be the opposite of that on section J=30, if a baroclinic instability really happens. Again, this is indeed the case, as shown in Figs. 6.48(a) and (c). Both of b_n^M and $\Delta_z Q_{PM}$ slope to the right (north) with height.

For a baroclinic instability dominated process, its signature will also be exhibited on the sectional distributions for perturbation fields. Here the term "perturbation" should be understood as the meso-scale feature, and until the end of this subsection, by the "meso-scale" feature of a given field p we mean a meso-scale time synthesis of p followed by a meso-scale space reconstruction. With this tacitly assumed, we perform a meso-scale decomposition for the vertical velocity w and density anomaly ρ. The scale window indices used are the same as the standard experiment: j_0=3, j_1=5, and j_0^{sp}=5, j_1^{sp}=9.

Figures. 6.49 and 6.50 are the distributions of the day-7 meso-scale w and ρ for the two sections we have examined before. Given the background density structure, the perturbation w is expected to have a pattern with phase lines sloping with height toward the west on the zonal section J=30, and toward the east on the meridional section I=20, for an Eady-like baroclinically unstable process (cf. Fig. 4.2). And, on the same conditions, the corresponding meso-scale density anomaly should also have a tilting pattern, but toward the opposite direction. This counter-tilting pattern is revealed on both the zonal section and the meridional section. In Fig. 6.49, it is roughly located between I=25 and I=35, seemingly a little shifted to the right compared to the center identified on the distribution of Fig. 6.47. But in Fig. 6.50, where the tilting phenomenon is very conspicuous, the inferred unstable region agrees well with its energetic counterpart, the one from Fig. 6.48.

The existence of a baroclinic instability is therefore substantiated, by a variety of signatures from the MS-EVA distributions to the prognostic field perturbation on two sections for a typical forecast day. Moreover, this instability happens mainly below the depth 150 m. Signatures from the section distributions of other days also support this assertion. In fact, results from other days before the meandering intrusion (day 8) show similar patterns for both section I=20 and section J=30. The only exception lies on day 6, when the barotropic transfer BT_H is significant in the west of the domain. In that case, the negative center on the BT_H distribution for the I=20 section is replaced by a pool positive in value.

6.3.4 Other transfers from the interaction analysis

In an MS-EVA with more than two scale windows, a mere energy transfer between the large-scale and meso-scale windows are still not sufficient to illustrate the growth of meso-scale disturbances. Contributions from other windows, though usually small, could sometimes be significant enough to alter the energy stream between scales. For example, we have seen a significant part of energy residing in the sub-mesoscale window on the time spectrum of v (e.g., Fig. 6.6). If this part of energy is from the meso-scale process, can we still expect some energy from that gained from the unstable large-scale system left to power the meandering intrusion? Apparently, the remaining transfers other than BC and BT from the interaction analysis also merit some examination.

We first write the total APE and KE transfers to the meso-scale window as T_{A^M} and T_{K^M}, respectively. In terms of the transfer terms defined before (see Table 3.2), they are

$$T_{A^M} \equiv T_{A^M,\partial_h\rho} + T_{A^M,\partial_z\rho} + TS_{A^M}, \tag{6.1}$$

$$T_{K^M} \equiv T_{K^M,\zeta} + T_{K^M,w\partial_z} + T_{K^M,\partial_z v}. \tag{6.2}$$

With the technique developed in §3.7, these two transfers can be analyzed as:

$$T_{A^M} = T_{A^M}^{0\to1} + T_{A^M}^{2\to1} + T_{A^M}^{2\oplus0\to1} + T_{A^M}^{1\to1}, \tag{6.3}$$

$$T_{K^M} = T_{K^M}^{0\to1} + T_{K^M}^{2\to1} + T_{K^M}^{2\oplus0\to1} + T_{K^M}^{1\to1}. \tag{6.4}$$

In the superscripts, the numbers 0, 1, and 2 represent respectively the large-, meso-, and submeso-scale windows, and the arrow indicates the direction of energy flow. Particularly, the superscript $0\oplus2\to1$ is used for the transfer to the meso-scale window due to the large-submeso-scale interaction. We next draw plots for these interaction terms.

By definition, the first terms on the right hand sides of Eqs. (6.3) and (6.4) are merely the BC and $BT_H + BT_V$. They have already been shown in the above subsections (except BT_V, which is insignificant compared to BT_H) but we re-draw them here for comparison. We pick day 6 and depth 125 m for the study, as we find the interaction analysis terms other than BC and BT are most significant on this day at this depth. These terms are graphed in Fig. 6.51 and Fig. 6.52. Particularly, BC and BT are plotted in Figs. 6.51a and 6.52a.

Look at $T_{A^M}^{1\to1}$ and $T_{K^M}^{1\to1}$ first. A term with superscript $1\to1$ arises from the local analysis representing the contribution from the same scale window. From Figs. 6.51d and 6.52d, it could be large, and in some region such as the positive BT center, it even exceeds BT (2.3×10^{-8} m^2/s^3 versus 1.0×10^{-8} m^2/s^3) and dominates the transfer. This fact elucidates from one aspect that transfers without interaction analysis might not be useful in identifying inter-scale processes.

Next look at the transfer from the large-sub-mesoscale interaction, $T_{A^M}^{0\oplus2\to1}$ and $T_{K^M}^{0\oplus2\to1}$. This part is usually negligible, as we have estimated before in §3.7. Here it is at least one order of magnitude smaller than the part from the large-scale window, both in Fig. 6.51 and in Fig. 6.52.

Of most interest among the interaction terms beside $T_{A^M}^{0\to1}$ ($T_{K^M}^{0\to1}$) is $T_{A^M}^{2\to1}$ ($T_{K^M}^{2\to1}$). In Fig. 6.51b, there is a negative core located close to where the BC hotspot lies (Fig. 6.51a), implying a significant part of energy transferred from the background field has been passed to the sub-mesoscale window. Our issue is, if this transfer cancels out the amount shown on the BC hotspot, then the energy gained from the baroclinic instability will not do anything for the meso-scale window, and hence can not account for the cause of the meso-scale meandering intrusion. Fortunately, the negative center of $T_{A_M}^{2\to 1}$ is far from significant in comparison to hotspot on the BC map. Consequently, the baroclinic instability does cause the growth of meso-scale disturbances, though some sub-mesoscale processes might share a part of its contribution.

In Fig. 6.52b, there is also a negative region lying where the positive BT sits. But this time it is significant in strength (-4×10^{-9} to -7×10^{-9} m^2/s^3 vs. $+0.1\times10^{-7}$ m^2/s^3 on the BT map), and the energy from the large-scale window through barotropic transfer is reduced by more than a half because of $T_{K_M}^{2\to 1}$. The reduced part of the transfer goes to the sub-mesoscale window, leading to the high-frequency oscillation which has been identified as a conspicuous feature on the time spectrum of v. In this sense, the barotropic instability occurred in the upper layers might not contribute much to the meso-scale deep-sock intrusion.

Described above are the meso-scale energetics of the standard experiment. For all that have been observed, the IFF system is undergoing a baroclinic instability around the meandering cold intrusion, and moreover, the instability is believed to occur below a certain depth of roughly 150 m. The energy gained from this process goes to the sub-mesoscale disturbances as well as the meso-scale processes, but most of it stays in the meso-scale window, serving to fuel the growth of the intrusion event.

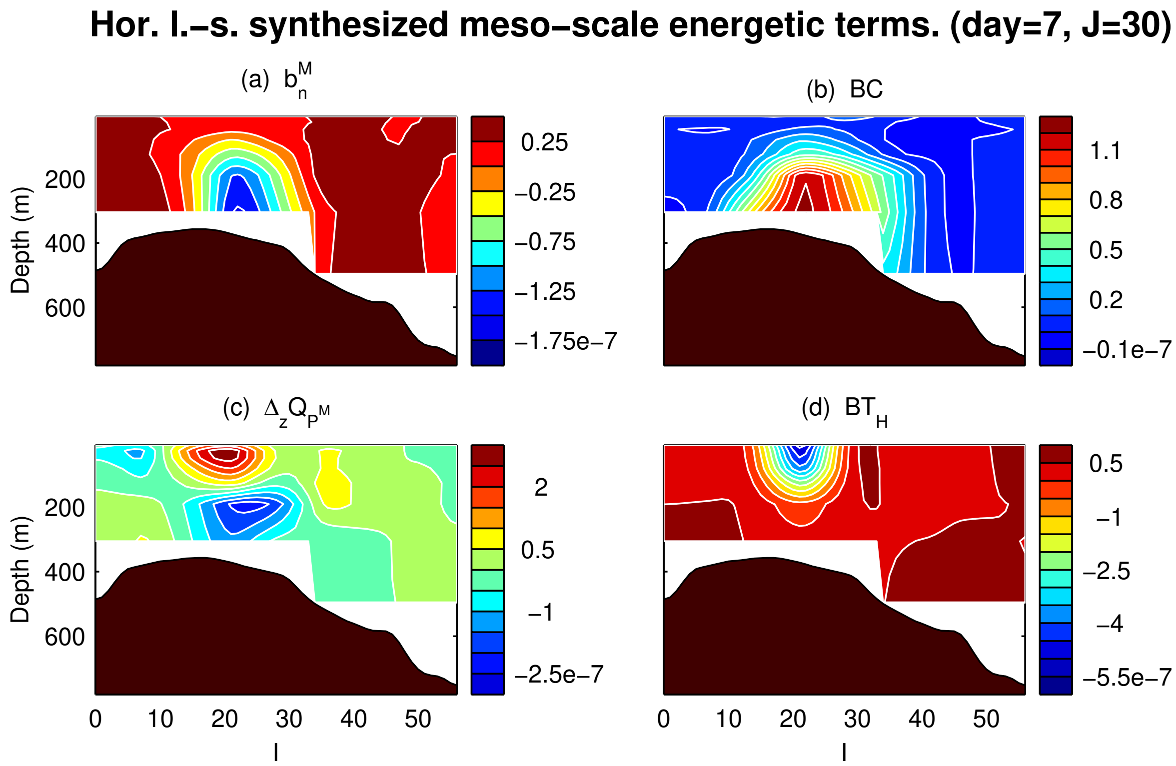

Figure 6.47: Horizontally (local) averaged meso-scale energetics for day 7 on the zonal section J=30.

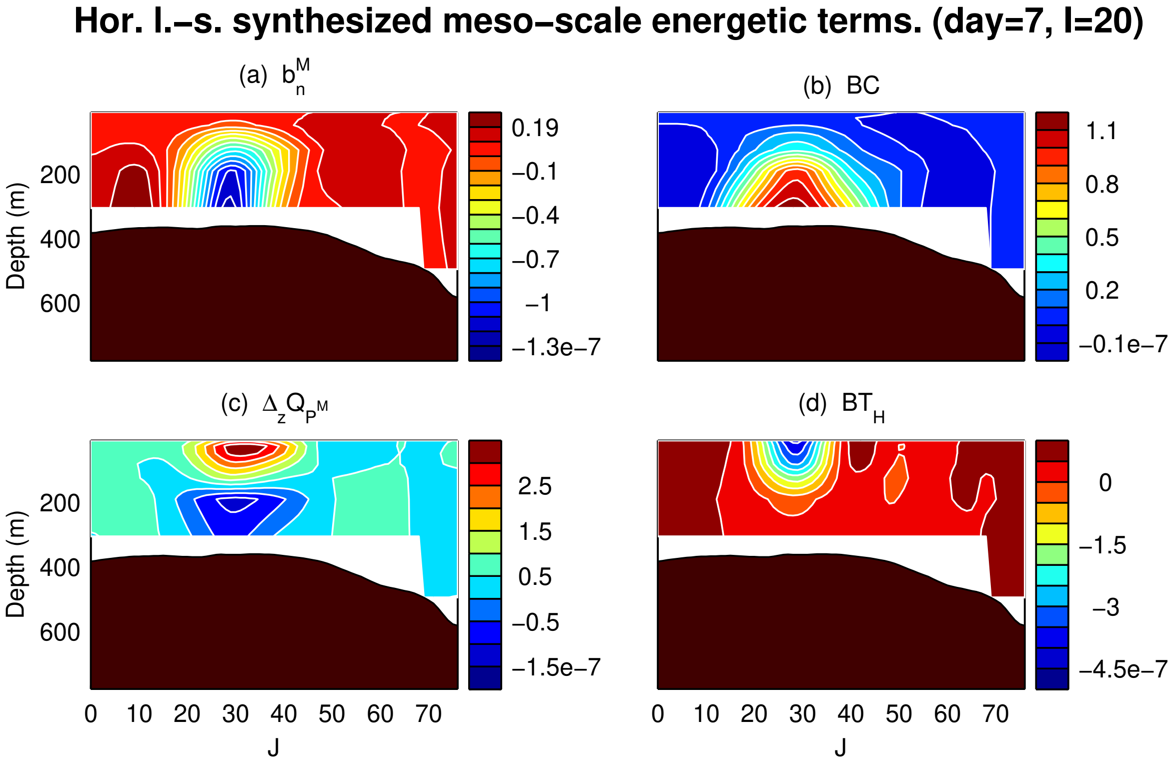

Figure 6.48: Horizontally (local) averaged meso-scale energetics for day 7 on the meridional section I=20.

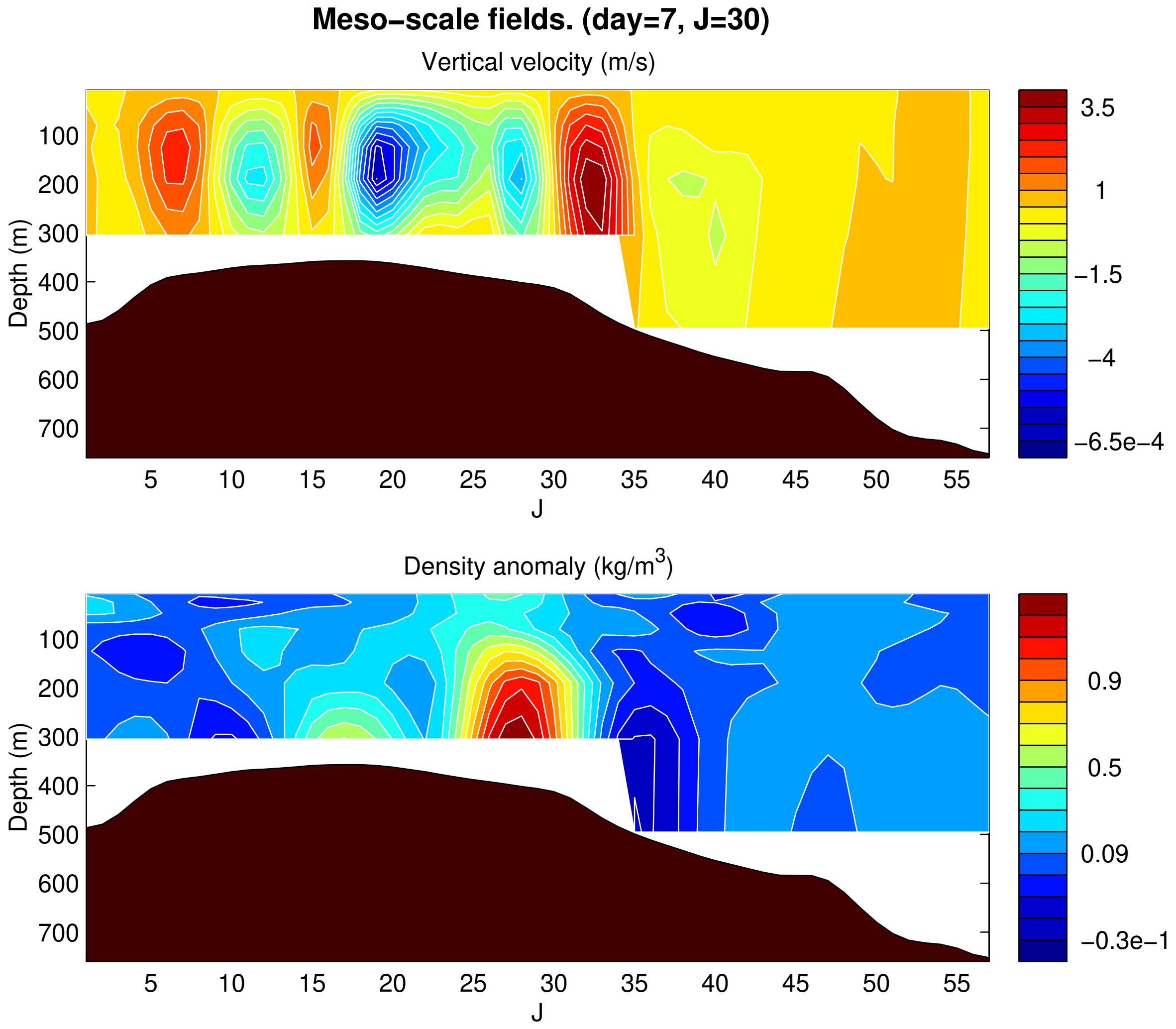

Figure 6.49: The meso-scale features of vertical velocity (top) and density anomaly (bottom) on the zonal section J=30 for day 7. In this figure by the "meso-scale" feature of a field we mean its time meso-scale synthesis with window bounds j_0=3 and j_1=5 followed by a horizontal meso-scale synthesis with bounds j_0^{sp}=5 and j_1^{sp}=9.

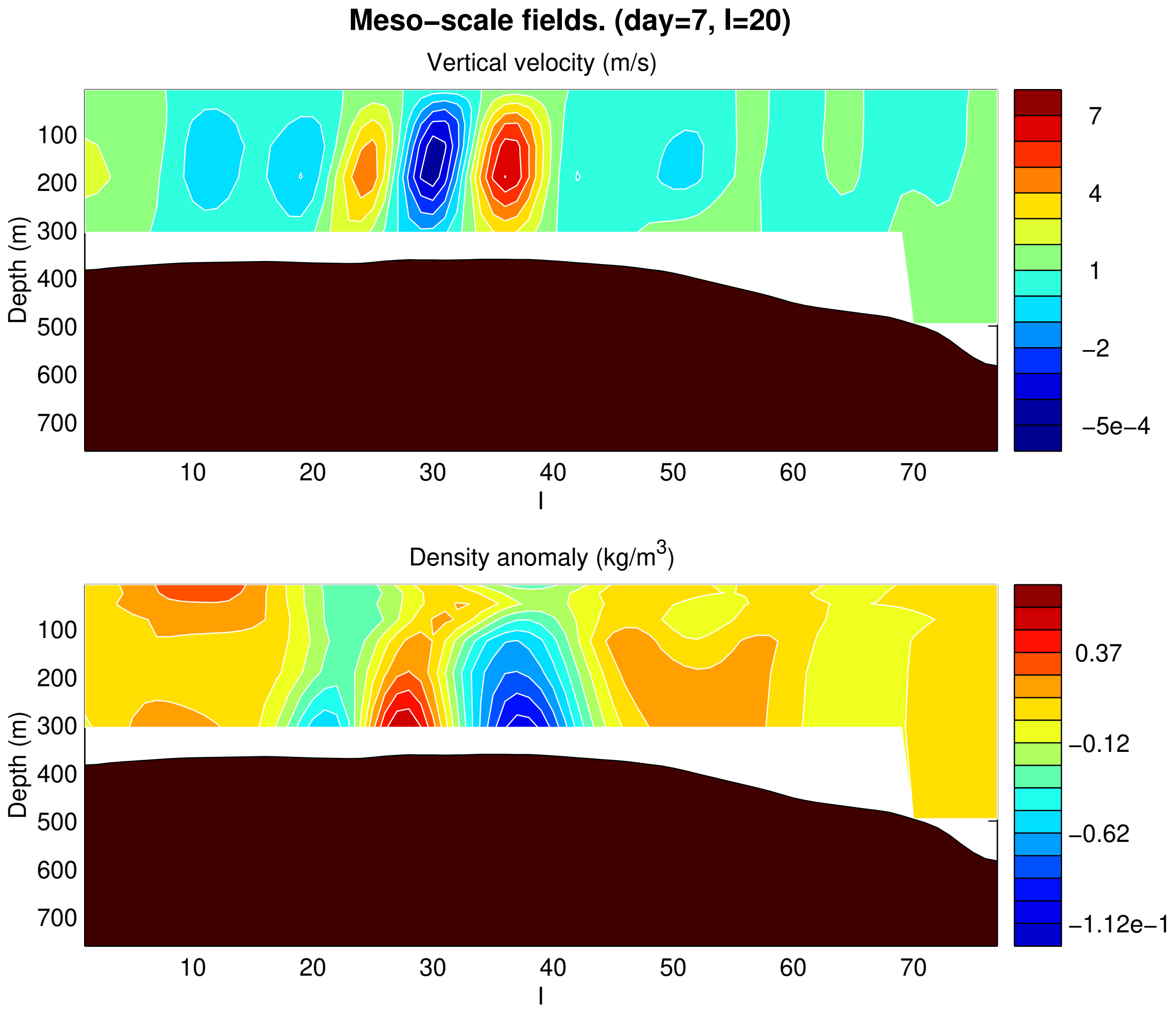

Figure 6.50: Same as Fig. 6.49, except for the meridional section I=20.

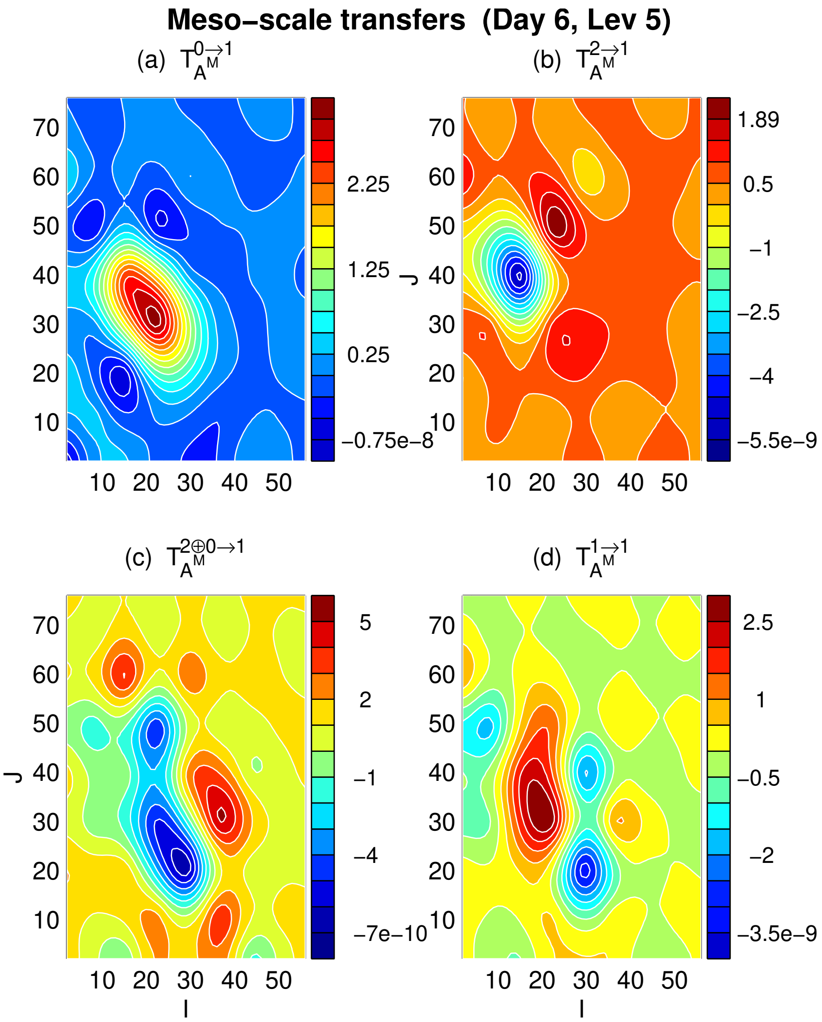

Figure 6.51: Interaction analysis of the total APE transfers, $T_{A^M} = T_{A^M,\partial_h\rho} + T_{A^M,\partial_z\rho} + TS_{A^M}$ (cf. Table 3.2), to the meso-scale window for day 6 at the 125-m level. The integers 0, 1, and 2 in the superscripts represent the large-, meso-, and submeso-scale windows, respectively. (Refer to Eq. (6.3) for the interaction analysis) Contoured in subplot (a) is simply the BC shown before.

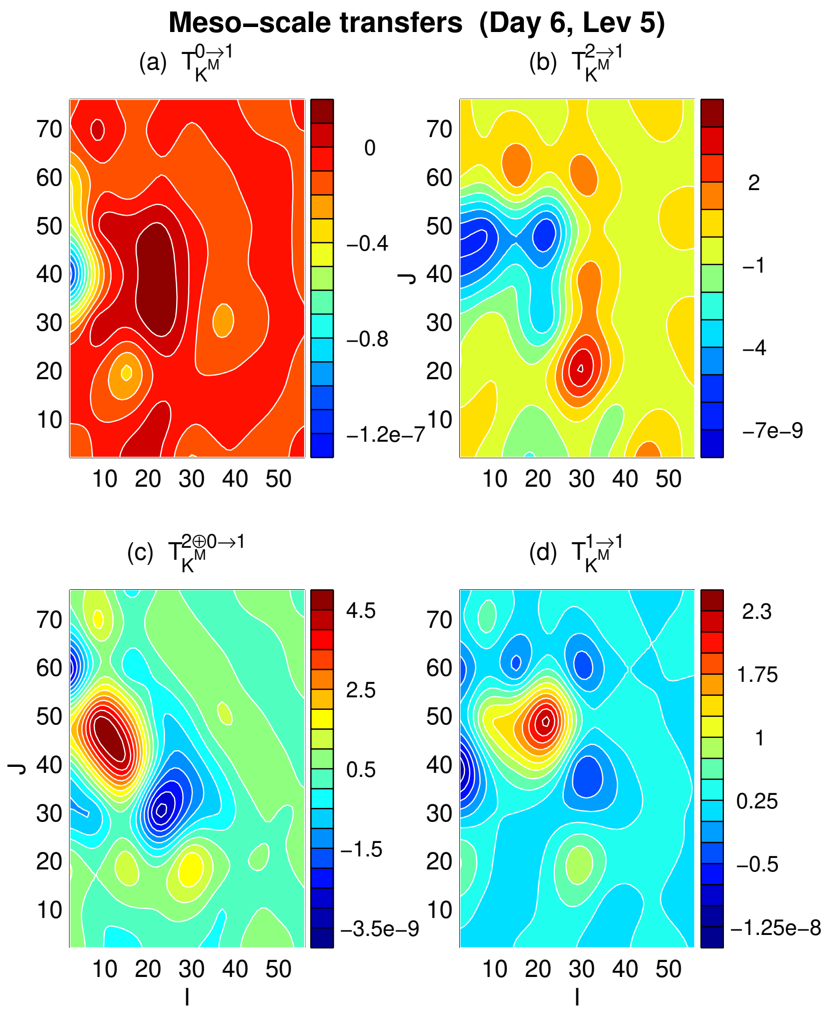

Figure 6.52: Interaction analysis of the total KE transfers, $T_{K^M} = T_{K^M,\zeta} + T_{K^M,w\partial_z} + T_{K^M,\partial_z v}$ (cf. Table 3.2), to the meso-scale window for day 6 at the 125-m level. The integers 0, 1, and 2 in the superscripts represent the large-, meso-, and submeso-scale windows, respectively. (Refer to Eq. (6.4) for the interaction analysis) Contoured in subplot (a) is simply the $BT = BT_H + BT_V$ shown before.

6.4 Sensitivity study

So far all we have done concerns the standard experiment with a certain combination of window bounds. As we mentioned in §6.2.3, there could be some ambiguity in setting these bounds. This section presents experiments with different parametric combinations, in the hope of getting rid of the ambiguity. Since in the standard experiment, it is mainly the APE mechanism that accounts for the formation of the meandering cold intrusion, we focus our attention on the sensitivity of the potential energetics, particularly the meso-scale buoyancy conversion b_n^M and the baroclinic instability indicator BC, to the parametric change. We choose day 7 and depth 125 m to address this issue. For reference, the related experiments are listed in Table 6.2.

6.4.1 Time window bounds

For the time window, we have adopted a j_0=3 and a j_1=5 for the standard experiment (Exp. 1 in Table 6.2) described in the preceding section. This choice makes a meso-scale window with scale levels j=3 and j=4. While these two levels are for sure what we want for the intrusion event based on the spectral analysis, the reason to rule out other possibilities is still not sufficient. For example, it is not impossible to have a j_1 exceeding 5, while j_0 could be made smaller than 3. In the following we wish to eliminate the ambiguity by experiments.

We first examine the effect of an extension of j_1 beyond j=5. Given j_0=3, j_0^{sp}=5, we have tried all the possibilities with j_1=6-10 (Exp. 2). The resulting energetics are basically the same as those obtained in the standard experiment. If no local average is performed, the result is still similar to its counterpart in Figs. 6.12-6.13, except for some small-scale features coming in. In a word, the choice of j_1 is not sensitive provided that the process with j=3-4 is included.

The real problem of the time parametric study comes from the choice of j_0. In addition to the standard experiment setting, we have three more options: j_0=0, j_0=1, j_0=2. Fig. 6.53 shows the b_n^M and BC of Exp. 3, with $j_0 = 0$ and other parameters the same as the standard experiment, for the seventh forecast day at the 125-m depth. In comparison to its counterpart in Fig 6.18, the buoyancy conversion b_n^M turns positive in the cold intrusion region, and a negative tongue appears between J=35 and J=50 right to the western boundary. On the BC map, the solitary hotspot is replaced by a relatively weak transfer pair with opposite signs, and a strong positive region at a location roughly corresponding to the negative tongue on the b_n^M map. By location the correlation between the processes identified here and the intrusion event observed in the forecast is far from significant. Apparently, The addition of processes with j=0, 1, and 2 have compounded the features of interest.

Similar result can be obtained with j_0=1 (Exp. 4), which forms a meso-scale window with the level $j = 1$ or scale 10.7 days included. The new features in the above observation with respect to Fig. 6.18 are therefore mainly from events with j=1 and 2.

Most of the j_0 ambiguity arises in the case j_0=2, as we have mentioned before in §6.2.3. While j=0 and 1 can be easily excluded from the meso-scale window because of the gap observed on the time spectrum, the peak at j=2 is in general mixed together with features at higher levels and hence it is a little subjective to decide whether to take it in or not. We have seen the meso-scale energetics in the standard experiment without j=2. In the following we show what will happen if it is taken into account.

Figure 6.54a shows the result of Exp. 5, which has a j_0=2, and other parameters the same as the standard experiment. Compared to its counterpart in Fig 6.18e, clearly the negative b_n^M center is still there. Its strength is also not changed. Both Fig. 6.54a and Fig. 6.18e have a maximal APE-to-KE conversion 9×10^{-8} m^2/s^3. The difference between the two maps is the appearance of a positive center immediately to the east. This comparison tells us that the APE-to-KE conversion on day 7 at depth 125 m in the deep-sock intrusion region happens at scale levels $j \geq 3$ only. The addition of j=2 only introduces on the right hand side of the intrusion a positive area, which implies a conversion from KE to APE. Similar observations can be made from the buoyancy conversion maps for other days at different depths. By all that accounts for, we may safely conclude that j=2 is not relevant to the intrusion event in which we are interested.

While the addition of scale level j=2 to the meso-scale window still keeps the negative b_n^M center, it totally changes the pattern of the BC distribution. The conspicuous hotspot in Fig. 6.18i finds no counterpart on the BC map in this experiment. Replaced at the same location in Fig. 6.54b is a negative core peripheral with some positive BC bands. In comparison

Table 6.2: MS-EVA experiments for the IFF variability study. Other parameters include $j_2 = 10$ and $j_2^{\rm sp} = 12$, and $j_1^{\rm sp} = 9$ is not essential here.

Experiment		Time W. bounds		Space W. B.	Remarks
Type	No.	j_0	j_1	j_0^{sp}	
Time window experiment	1	3	5	5	The standard experiment
	2	3	6-10	5	Almost the same as Exp. 1.
	3	0	5	5	b_n^M turns positive in the intrusion region. A new negative tongue appears.
	4	1	5	5	Similar to Exp. 3.
	5	2	5	5	A new positive b_n^M region appears, and the transfer information is disguised on the BC map.
Space window experiment	6	3	5	3	Similar to Exp. 1 except that the hotspot is smeared.
	7	3	5	4	Same as above.
	8	0	5	3	Some conspicuous features in Exp. 3 disappear.

to the standard experiment, a different process apparently has edged in, with an energy transfer pattern quite unexpected for our purpose. The scale level j=2 thus does not characterize the cold intrusion event, and hence a time window bound must be chosen such that j_0>2.

6.4.2 Space window bounds

Besides the time window bounds, we also have space window bounds as parameters in the application of MS-EVA. The purpose of a space decomposition is to perform a local average for the obtained energetics. We hence only need to consider $j_0^{\rm sp}$. Other bounds, such as $j_1^{\rm sp}$, are irrelevant in this sense.

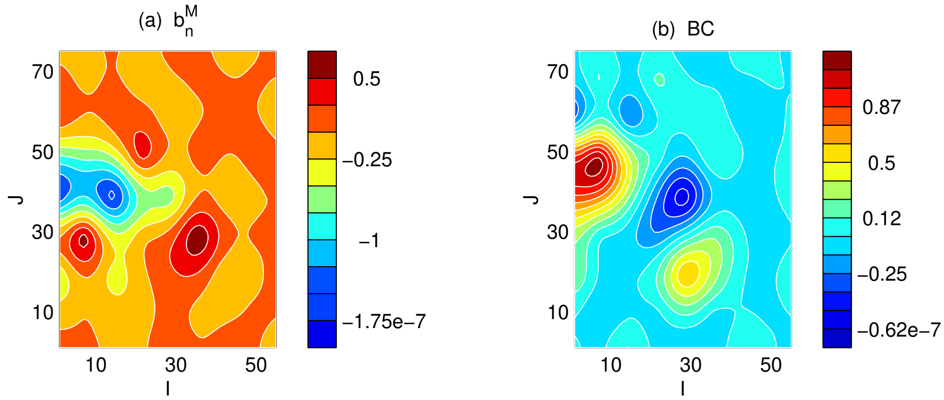

Figure 6.53: Meso-scale buoyancy conversion (a) and baroclinic instability indicator (b) for the seventh day at depth 125 m, with time window bounds j_0=0, j_1=5, and space window bound j_0^{sp}=5. All the energetics are in m^2/s^3.

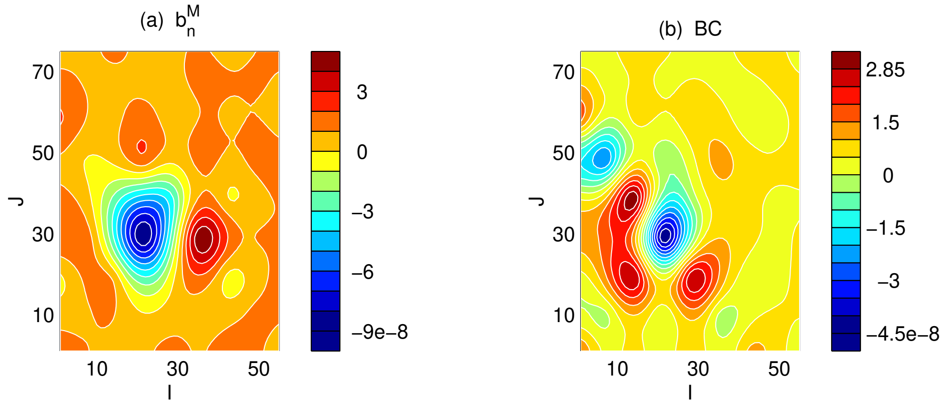

Figure 6.54: Same as Fig. 6.53, but for j_0=2.

In the present study, we care about only the meso-scale dynamics. An ideal $j_0^{\rm sp}$ should be such that it helps average out all the meso-scale oscillations so as to unravel the residual of the MS-EVA terms, while retaining as many as possible the features for the meso-scale energetics. We have chosen $j_0^{\rm sp}$=5 in the standard experiment. We here try options other than this scale level.

Shown in Fig. 6.55 are the meso-scale buoyancy conversion and large-to-meso-scale transfer (day 7, depth 125 m) for Exp. 6, where $j_0^{\rm sp}$=3, and all other parameters are the same as Exp. 1. Clearly, they are very similar to their counterparts in Fig. 6.18. Both the negative b_n^M center and the BC hotspot are there, though in the present experiment they have been smeared horizontally. This remarkable feature still exists when $j_0^{\rm sp}$ is switched to $j = 4$ (Exp. 7), a conspicuous peak as well as $j = 3$ identified on the space spectrum. Consequently, for the standard experiment, the meso-scale MS-EVA result is insensitive to the choice of $j_0^{\rm sp}$.

But when the time window bound j_0 is not correctly set, the effect of $j_0^{\rm sp}$ does show its significance. In Fig. 6.56 we re-graph the b_n^M and BC obtained in Exp. 3 (Fig. 6.53), but with a low $j_0^{\rm sp}$=3. Apparently, the negative b_n^M (positive BC) region near the western boundary is made much more clear, but the positive/negative BC pair in the middle area has been totally eliminated.

Consequently, the key to the success of the IFF MS-EVA analysis is to choose a j_0 which correctly reflects the deep-sock meandering. Addition of any processes with scale level below j=3 into the meso-scale window will disguise the dynamics of interest, making the obtained energetics not useful for our purpose. When a correct time window is set, the resulting meso-scale MS-EVA terms are not sensitive to the choice of space window bounds.

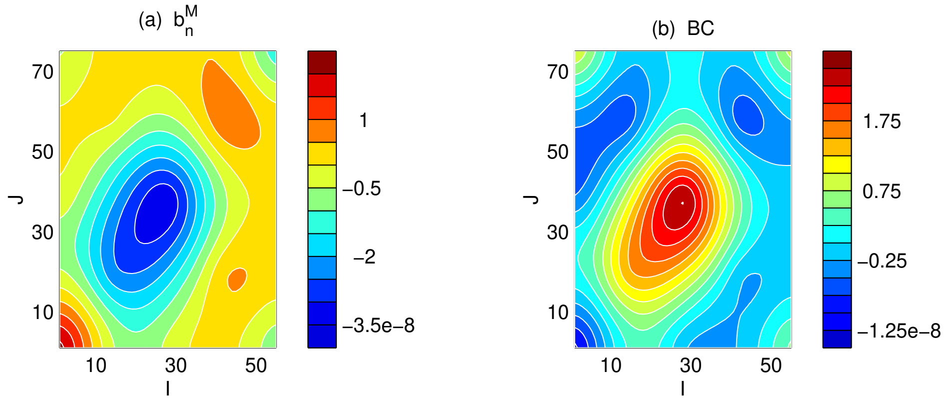

Figure 6.55: Meso-scale potential energetics (in m^2/s^3) for day 7 at the 125-m level with time window bounds set as the standard experiment (j_0=3, j_1=5), but the space window bound $j_0^{sp} = 3$.

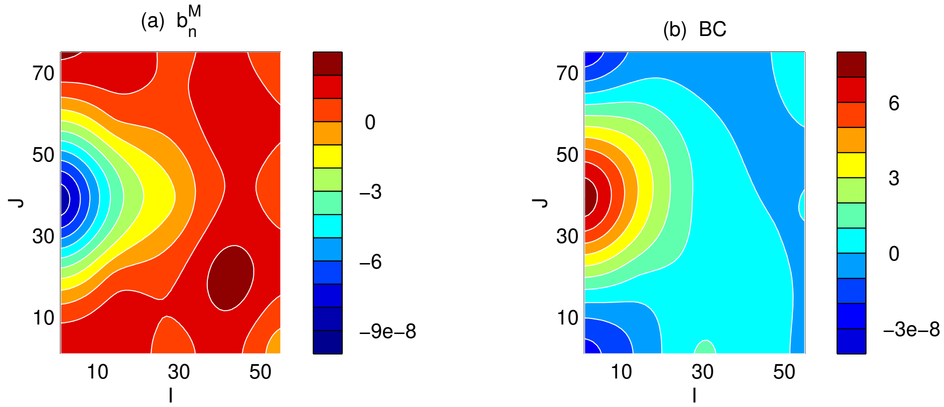

Figure 6.56: Same as Fig. 6.55, but for $j_0 = 0$.

6.5 Summary

We have performed an application of MS-EVA to study the IFF variability, with the results obtained in the preceding chapter as the input. The meandering cold intrusion observed in the forecast has been faithfully represented in a meso-scale window, and the dynamics behind it pinpointed and substantiated.

The application begins with a determination of the scale window bounds, which is achieved through a one-dimensional time wavelet spectral analysis and a two-dimensional space wavelet spectral analysis. A general observation from the spectra is that the IFF variability occurs mainly on two scales, either in time or in space. Energy is maximized at the time scales 10.3 days and 1.3-5.3 days, and the space scales 20-30 km and 40-60 km. This double-peak spectral structure allows us to pick for the meandering intrusion a reasonable window bound combination, which is, in the present configuration (cf. Table 6.1), $(j_0, j_1, j_2;\ j_0^{\mathrm{sp}}, j_1^{\mathrm{sp}}, j_2^{\mathrm{sp}}) = (3, 5, 10;\ 5, 9, 12)$.

With the window bounds determined, the MS-EVA is set up and the output visualized. We have given a detailed description of the meso-scale energetics for the intrusion event. Generally speaking, the meso-scale activities are concentrated in a small region where the cold tongue is observed. The locally averaged energetics as a whole is characterized by a solitary positive hotspot of BC, the potential energy transfer from the large-scale window, and a negative center of b_n^M, the meso-scale buoyancy conversion, for all the depths in the water column. The hotspot coincides approximately with the negative center in locality. In a general trend both of them get stronger and stronger in terms of absolute magnitude and coverage, from day 5 until the meander is fully developed on day 8. This pattern and evolutionary trend is also reflected on the deep-layer maps of the total large-to-mesoscale transfer, $BC + BT$, where BC dominates the distribution. By what we have shown in Chapter 4, this means that, in the area of concern, the system is undergoing an instability, and moreover, a baroclinic instability. But this process is operative below a certain depth. Above 150 m, the map of $BC + BT$ shows no solitary center of positive transfer as that of BC does, or even there is such a center (on day 6), the transfer is dominated by the barotropic instability indicator, BT.

The identified baroclinic instability and its vertical structure are also evidenced on the distributions of other energetics. The negative b_n^M mentioned above indicates clearly a conversion from APE to KE within the meso-scale window. In deep layers, a large part of the converted energy is brought upward/downward through the vertical pressure work, $\Delta_z Q_{PM}$. But in surface layers (above 150 m), $\Delta_z Q_{PM}$ functions quite differently. The whole scenario confirms the existence of the vertically limited baroclinic instability, which is further verified by the tilting patterns of the perturbation fields on related sections.

It should be pointed out that our deep-layer unstable structure is consistent with the "surface-intensified baroclinic instability" identified by Miller et al. (1995). Apparently, they have assigned to the words "surface" and "deep" a meaning different from what we have here. In their study, only depths below 200 m are considered, which obviously fall within our unstable regime.

We have also shown that the energy gained from the baroclinic instability is indeed used to power the meso-scale meandering. An interaction analysis indicates that the energy leaving the meso-scale window to sub-mesoscale processes is inconsequential in comparison to the part obtained through the instability.

To test the robustness of our application against the parametric change, we have tried a series of experiments with different combinations of window bounds. In the time direction, we find, the choice of j_0 is crucial. Failure to set a j_0 characteristic of the meandering event will result in energetics being totally misleading. When j_0 is correctly set (i.e., $j_0 = 3$), the result is insensitive to space window bound change.

The moral of the sensitivity study is that, in order to pinpoint the right dynamics for a process, we need to capture the right scale characteristic of that process. A fixed j_0 cannot be expected to handle all the situations which are different in general. Recall that the classical multiscale energetic analysis (i.e., Harrison and Robinson, 1978) is equivalent to our MS-EVA with $j_0 = 0$ and a periodic extension. If a classical analysis is performed for the IFF variability, our sensitivity study shows it will lead to nowhere for the meandering event. This is a good example that illustrates from one perspective the advantage of our MS-EVA.

Bibliography

[1] Allen, J. T., D. A. Smeed, and A. L. Chadwick, 1994: Eddies and mixing at the Iceland-Faeroe Front. *Deep-Sea Res.*, 41, No.1, 51-79.

[2] Harrison, D. E, and A. R. Robinson, 1978: Energy analysis of open regions of turbulent flows-mean eddy energetics of a numerical ocean circulation experiment. *Dyn. Atmos. Oceans*, 2: 185-211.

[3] Killworth, P., D. N, Paldor, and M. E. Stern, 1984: Wave propagation and growth on a surface front in a two-layer geostrophic current. *J. Mar. Res.*, 42, 761-785.

[4] Maskell, S. J., A. D. Weathershaw, and C. E. Stretch, 1992: Topographic and eddy effects in a primitive equation model of the Iceland-Faeroe Front. *J. Mar. Sys.*, 3, 343-380.

[5] Miller, A. J., et al., 1995a: Quasigeostrophic forecasting and physical processes of Iceland-Faeroe Frontal variability. *J. Phys. Oceanogr.*, 25, 1273-1295.

[6] Miller, A. J., P.-M. Poulain, A. R. Robinson, H. G. Arango, W. G. Leslie, and A. Warn-Varnas, 1995b: Quantitative skill of quasi-geostrophic forecasts of a baroclinically unstable Iceland-Faeroe front. *J. Geophys. Res.*, 100, 10 833 - 10 849.

[7] Miller, A. J., P. J. Lermusiaux, 1996: A topographic-Rossby mode resonance over the Iceland-Faeroe Ridge. *J. Phys. Oceanogr.*, 26, 2735-2747.

[8] Niiler, P. P., S. Piacsek, L. Neuberg, and A. Warn-Varnas, 1992: Sea surface temperature variability of the Iceland-Faeroe Front. *J. Geophys. Res.*, 97, 17777-17785.

[9] Pinardi, N., and A. R. Robinson, 1986: Quasigeostrophic energetics of open ocean regions. *Dyn. Atmos. Oceans* 10, 185-219.

[10] Read, J. F., and R. T. Pollard, 1992: Water masses in the region of the Iceland-Faeroe Front. *J. Phys. Oceanogr.*, 22, 1365-1378.

[11] Robinson, A. R., H. G. Arango, A. J. Miller, A. Warn-Varnas, P.-M. Poulain, and W. G. Leslie, 1996: Real-time operational forecasting on shipboard of the Iceland-Faeroe Frontal variability. *Bulletin of the American Meterological Society*, 243-259.

[12] Scott, J. C., and A. L. McDowall, 1990: Cross-frontal cold jets near Iceland: in water, sattelitte imfrared, and Geosat altimeter data. *J. Geophys. Res.*, 95, 18005-18014.

[13] Smart, J. H., 1984: Spatial variability of major frontal systems in the North Atlantic-Norweigian Sea area: 1980-81. *J. Phys. Oceanogr.*, 14, 185-192.

[14] Spall, M., 1989: Regional primitive equation modeling and analysis of the POLYMODE data set. *Dyn. Atmos. Oceans*, 14, 125-174.

[15] Swift, J. H., 1984: The circulation of the Denmark Strait and Iceland-Scotland overflow waters in the North Atlantic. *Deep-Sea Res.*, 31, 1339-1355.

[16] Fofonoff, 1985: UNESCO International Equation of State (IES 80). *J. Geophys. Res.*, 90, C2, pp. 3332-3342.

[17] Willebrand, J., and J. Meincke, 1980: Statistical analysis of fluctuations in the Iceland-Scotland frontal zone. *Deep-Sea Res.*, 27A, 1047-1066.

Chapter 7

Summary and conclusions

A new methodology, Multiscale Energy and Vorticity Analysis (MS-EVA), has been developed, validated, and applied to a real ocean problem. In Chapter 1, the research issues were addressed and the solution strategy presented. The development began in Chapter 2, with the construction of a functional analysis tool, the multiscale window transform, in the framework of multi-resolution analysis. Specifically, we have defined and built three scale windows, which are bounded by three scale levels j_0, j_1, and j_2, and constructed a transform and a synthesis with respect to each window. Cases with a one-dimensional field and a two-dimensional field have been investigated. Either of them gives an orthonormal, local, and self-similar transform, which is windowed on scales. We first worked on the real line, then moved to a finite domain, extended with either a periodic or a symmetric scheme. We have proved that both schemes will yield a transform without energy leakage, so long as a symmetric scaling function is adopted. Such an energy-conserving transform has many properties. A most important one is the property of marginalization, which relates a phase space representation of the product of two signals to its physical space expression. More precisely, marginalization of the product of two multiscale window transforms results in a duration/domain average of the product of the corresponding multiscale window syntheses. A realization of this kind of transform was made with the aid of cubic splines. A fast algorithm was presented and the computational procedure outlined.

The multiscale window transform thus obtained was then exploited to formulate the MS-EVA. In Chapter 3, the large-, meso-, and submeso-scale potential and kinetic energy and enstrophy equations were derived and interpreted. Energetic processes represented by the terms of these equations are classified into four categories: transport, transfer, conversion, and dissipation/diffusion, and we have studied in detail the first three processes. Buoyancy conversions occur on each scale window without inter-scale energy transfer involved. The major issue in this process decomposition was the separation of transports and transfers from the nonlinear terms in the multiscale energy equations. We introduced the concept of perfect transfer and showed that all these terms can be split as expected. The transfer terms obtained include, by governing mechanism, the transfer due to vorticity evolution, the transfer due to vertical advection, the transfer due to the vertical shear of horizontal velocities, and the transfer due to density gradient evolution. These terms were further decomposed with interaction analysis to distinguish the inter-scale energy transfer from transfers from other sources. The whole formulation has been connected to the classical energetics analysis. We proved, with the aid of the property of marginalization, that the classical formulation is a very particular case of our MS-EVA. They are identical in a two-scale decomposition when $j_0 = 0$ and a periodic extension scheme is adopted.

The developed MS-EVA was validated with the stability problem for a zonal jet stream. In the first part of Chapter 4, we proved that the stability of a system in the classical sense is entirely dependent on the transfer terms from the MS-EVA analysis. For a two-scale window decomposition, let BC denote the sum of the large-to-meso-scale transfers (obtained through an interaction analysis) from the APE equation, and BT signify the same sum but from the KE equation, then their local averages, $\langle BC \rangle^{\Omega}$ and $\langle BT \rangle^{\Omega}$, proved to be two indicators of system instability. When $\langle BC + BT \rangle^{\Omega} > 0$, the system is unstable and vice versa. Particularly, if it is $\langle BC \rangle^{\Omega}$ that makes $\langle BC + BT \rangle^{\Omega}$ positive, then there exists a baroclinic instability; otherwise if it is $\langle BT \rangle^{\Omega}$ that dominates a positive $\langle BC + BT \rangle^{\Omega}$, the instability must be barotropic. With these points the MS-EVA was put to test with the well-known Eady and Kuo models, each admitting only one of the two types of instability. In each case, a positive residual results from the instability indicator for any chosen unstable mode. Specifically, the Eady model gives a positive horizontally averaged BC while its BT sums to zero over the space domain; the Kuo model results in the opposite, just as expected for a barotropically unstable system. The MS-EVA is thereby validated.

Chapter 4 also presented spatial structures that characterize the Eady- and Kuo-like unstable systems. With an Eady-like instability, the meso-scale buoyancy conversion is always negative—that is to say, corresponding to a baroclinic instability there is a buoyancy conversion on the meso-scale window, and the conversion is from available potential energy (APE) to kinetic energy (KE). Vertically, the conversion is maximized in the middle of the water column and minimized at the surface and the bottom. The energy converted from APE is drained upward and downward by the vertical pressure work, leading to an accumulation of velocity disturbances close to the two vertical boundaries. At the same time, the perturbation APE is also intensified toward the bottom and the surface, as the transfer is uniform through the water column while the largest conversion happens in the middle. The surface and bottom intensification is evident on the meso-scale feature maps for both density and horizontal velocity components (u or v).

While the Eady unstable system is characterized by a conspicuous pattern in the vertical direction, the transfer in the Kuo model is structured horizontally. The zonally averaged MS-EVA shows that the energy transfer is limited within two latitudes, and becomes negligible beyond that limit. In general the transfer is greater than zero (i.e., from the large-scale window to the meso-scale window), but near the two latitudes there exist two narrow strips where BT_H (a dominant part of BT due to rotation and vertical advection) is negative. These strips provide regions for the transferred energy to feedback to the large-scale window, functioning like two walls that bar the energy transfer from happening further in the north and south. These MS-EVA scenarios, both with the Eady model and the Kuo model, appear consistent with what we have already known. In this sense, the MS-EVA is validated again.

With the validated MS-EVA, we set up an application to a real ocean problem, the Iceland-Faeroe Frontal (IFF) variability study. We first conducted a forecasting in Chapter 5 with a dataset gathered during the 1993 NRV *Alliance* cruise, which is composed of an initialization survey, a zigzag survey, and a validation survey. The forecasting was set in a real-time operational mode, with fields sequentially updated as data come in. For regions not covered by the survey, a feature model was used to facilitate the initialization. The whole procedure was then completely data-driven. With this we have launched two major forecasts, labeled as F2 and F5 respectively. The former made the prediction with the initialization data only, while the latter started from the former and further took in the zigzag survey information. The zigzag data assimilation was basically through optimal interpolation, but with an ingredient of multiscale treatment. The forecasts were verified with the validation survey data. The feature reproduced proved to be highly correlated with the observations, both in terms of the defined forecast skill and by visual inspection. The F5 forecast was found to be especially successful.

From the observations and the forecast, a conspicuous event identified for the 1993 IFF variability is the intrusion of a cold tongue, or the formation of a deep-sock meander, on August 22. An application of the MS-EVA was set up in Chapter 6 in the hope of nailing down the dynamics behind this phenomenon. The application began with a wavelet spectral analysis, in order to determine the time and space window bounds needed in the MS-EVA set-up. The variability was observed mainly to be happening on two scales, both in time and in space. The cold meandering intrusion was shown to be characterized by a 2.7-day peak on the time spectrum, and by a 20-30-km maximum on the space spectrum. The spectral structure sets a rule for the scale decomposition and the window bounds were chosen accordingly. An MS-EVA diagnosis with these bounds was launched and the output visualized. From the energetics, meso-scale activities are accumulated in a small area around the deep-sock meander. The locally averaged energetics (achieved through a horizontal large-scale synthesis) reveals a clear baroclinic instability during August 19 through August 21 around the cold intrusion, and shows that the instability is limited to the region beneath depth 150 m (approximately). This observation is consistent with the fingerprints on the related vertical energetics, and was further verified by the counter-tilting patterns of perturbation density and vertical velocity. We showed, through an interaction analysis, that the energy gained from this baroclinic instability is indeed used to power the meandering intrusion. The leakage to the sub-mesoscale processes is negligible in comparison to portion retained for meso-scale perturbation growth. All these results have been tested against change of MS-EVA parameters. In this example, we found, a correct large-meso-scale decomposition is the key to a successful MS-EVA application.

Summarized above are the major results from this thesis. The MS-EVA study is still in progress, however. At this stage, many research issues are still open. Firstly, the missing vertical and horizontal diffusion/dissipation need to be considered. They could be significant, especially in cases where mixing is important. Secondly, a more appropriate APE equation might be necessary to be commensurate with the Harvard Ocean Prediction System (HOPS). The HOPS computes temperature and salinity, and uses the UNESCO equation of state to obtain density. The advection-diffusion equations of temperature and salinity together with a nonlinear equation of state does not in general result in an APE equation exactly as we obtained in Chapter 3. Thirdly, the interpolation of a σ-level system to a z-level coordinate could invoke inaccuracies. These three issues are the main problems that may cause imbalances on the energetics computed, particularly in dealing with real problems where all kinds of processes could be present. For this reason, they may also referred to as computational issues. Another computational issue is that the differencing scheme for MS-EVA equations cannot be made exactly the same as the one used for forecast, as energy/vorticity evolution is different from momentum integration, anyhow. This accounts for a certain amount of error in the diagnostic fields, though theoretically it can be made as small as one wants by increasing the resolution of the mesh grid.

The last and most important issue regards the dynamic fingerprinting. More research on fundamental processes is desired to build up the library of fingerprints. For MS-EVA, a larger library of dynamical fingerprints implies more power. It will doubtlessly broaden the horizon of application. As a short-term goal, one may feel interested in problems such as how energy is accumulated from seemingly irrelevant processes to form a coherent structure of weather importance (e.g., a two-dimensional turbulence), how a wave loses its stability and how its energetic structures evolve with the change of propagation properties, how waves and currents coexist whereas strong interactions are present, to name but a few. Besides, we have not yet touched the cascade and inverse cascade (i.e., transfer in the phase space) of "vorticity energy" with the enstrophy equations developed. Because of the importance of vorticity in geophysical fluid dynamics (as we have introduced in Chapter 1), a large pool of related problems are to be investigated.

Ultimately, the capabilities of the MS-EVA can only be demonstrated through applications. This is not just because of the importance of application per se. The complexity of real problems in turn gives the impetus for fundamental process research. We have examined the frontal variability for the Iceland-Faeroe region. Studies with other regions with benchmark results are needed for our purposes. In a word, while continuing to improve what we already have with the methodology, we need to look for more and more applications with real problems. This forms the major research direction of MS-EVA for the near future.

Index