Ocean Modelling 40 (2011) 164–189

Contents lists available at SciVerse ScienceDirect

Ocean Modelling

journal homepage: www.elsevier.com/locate/ocemod

Statistical field estimation for complex coastal regions and archipelagos ☆

Arpit Agarwal, Pierre F.J. Lermusiaux *

Massachusetts Institute of Technology, Cambridge, MA 02139, USA

ARTICLE INFO

Article history:
Received 14 April 2011
Received in revised form 26 July 2011
Accepted 3 August 2011
Available online 17 August 2011

Keywords:
Field mapping
Gauss-Markov estimation
Coastal objective analysis
Data assimilation
Fast Marching Method
Level Set Method
Geostrophy
Levitus Climatology
World Ocean Atlas (WOA)
Archipelago

ABSTRACT

A fundamental requirement in realistic ocean simulations and dynamical studies is the optimal estimation of gridded fields from the spatially irregular and multivariate data sets that are collected by varied platforms. In this work, we derive and utilize new schemes for the mapping and dynamical inference of ocean fields in complex multiply-connected domains and study the computational properties of these schemes. Specifically, we extend a Bayesian-based multiscale Objective Analysis (OA) approach to complex coastal regions and archipelagos. Such OAs commonly require an estimate of the distances between data and model points, without going across complex landforms. New OA schemes that estimate the length of shortest sea paths using the Level Set Method (LSM) and Fast Marching Method (FMM) are thus derived, implemented and utilized in idealized and realistic ocean cases. An FMM-based methodology for the estimation of total velocity under geostrophic balance in complex domains is also presented. Comparisons with other OA approaches are provided, including those using stochastically forced partial differential equations (SPDEs). We find that the FMM-based OA scheme is the most efficient and accurate. The FMM-based field maps do not require postprocessing (smoothing). Mathematical and computational properties of our new OA schemes are studied in detail, using fundamental theorems and illustrations. We find that higher-order FMM's schemes improve accuracy and that a multi-order scheme is efficient. We also provide solutions that ensure the use of positive-definite covariances, even in complex multiply-connected domains.

1. Introduction and motivation

Statistical field estimation theory was introduced by Gandin (1965) to meteorology and was extended to oceanography by Bretherton et al. (1976) where it is commonly referred to as Objective Analysis (OA). The theory, based on the Gauss-Markov theorem (Plackett, 1950), provides a sound basis for interpolating irregularly spaced data onto a computational grid. Up to specifics of multiscale oceanic and atmospheric fields, classic OA schemes are equivalent to utilize the update step of the Kalman Filter to grid the irregularly-spaced data. Specifically, the data is gridded based on specified prior field estimates and error covariances. The methodology has been well formulated for open oceans without any landforms (convex simply-connected domains), but OA in complex coastal regions (multiply-connected domains) is one of the 'last' mapping problems which remains to be studied in detail. This is one of the main research questions of the present work.

Our research uses the Multidisciplinary Simulation, Estimation and Assimilation System (Haley and Lermusiaux, 2010; MSEAS, 2010). MSEAS consists of a set of mathematical models and computational methods for ocean predictions and dynamical diagnostics, for data assimilation and data-model comparisons, and for optimization and control of autonomous ocean observation systems. It is used for fundamental research and for realistic simulations and predictions, recently including monitoring (Lermusiaux, 2007), real-time acoustic-ocean predictions (Xu et al., 2008; Lermusiaux et al., 2010) and environmental management (Cossarini et al., 2009). Several dynamical models are part of MSEAS, including a free-surface primitive-equation dynamical model which uses implicit two-way nesting (Haley and Lermusiaux, 2010). This new multiscale free-surface code builds on the primitive-equation model of the Harvard Ocean Prediction System (HOPS, Haley et al. (2009)). Additionally, barotropic tides are calculated from an inverse tidal model (Logoutov and Lermusiaux, 2008; Logoutov, 2008).

In the multiscale OA schemes of MSEAS, the Kalman updates are carried out successively, from the largest scale to the smallest scale, using sequential processing of observations and scale

☆ We are grateful to the Office of Naval Research for research support under Grants N00014-07-1-1061 and N00014-08-1-1097 (ONR6.1), N00014-07-1-0473 (PhilEx), S05-06 and N00014-08-1-0680 (PLUS) to the Massachusetts Institute of Technology.

* Corresponding author.
E-mail addresses: arpit@mit.edu (A. Agarwal), pierrel@mit.edu (P.F.J. Lermusiaux).

1463-5003/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ocemod.2011.08.001

separation. In a two-scale version, a two-staged OA approach (Lermusiaux, 1997, 1999) maps the data onto oceanic fields in two steps: the larger and the smaller scale steps. The main inputs to one of these steps are the statistical description of the field being estimated and the observational noise covariance. While the latter is dependent on the measurement sensor, knowledge of the field statistics does not come easily in oceanography due to the scarcity of observations. The field statistics is often provided by analytical correlation functions which depend on the spatial separation distance and the spatial–temporal scales (Carter and Robinson, 1987). Other MSEAS schemes also utilize 4D dynamical models to construct covariances (Lermusiaux et al., 2000; Lermusiaux, 2002). These dynamical models have been successfully simplified to diffusion models (Lynch and McGillicuddy, 2001) and this approach is also used here for benchmarking.

Our work is motivated by the Philippines Straits Dynamics Experiment (PhilEx, Gordon et al. (2011)). The goal of PhilEx is to enhance understanding of the oceanographic processes and features arising in and around straits, and to improve the capability to predict the inherent spatial and temporal variability of complex Archipelago regions using models and advanced data assimilation techniques. In addition to the Philippines, we have used our new schemes in several coastal regions with and without islands, including the Taiwan region, New England shelf, Dabob Bay and Monterey Bay (Xu et al., 2008; Lermusiaux et al., 2010; Haley and Lermusiaux, 2010). Other OA schemes have been used in coastal regions (Hessler, 1984; Stacey et al., 1988; Paris et al., 2002), but without satisfying coastline constraints, in particular, there should be no direct relationship across landforms. In ocean regions with complex 3D geometries, we found that such schemes generate field estimates that lead to major issues when used to initialize simulations. Efficient and accurate methodologies for field (e.g. temperature, salinity, biology, and velocity) mapping in complex multiply-connected coastal domains and archipelagos were thus necessary.

Our schemes estimate the sea paths between data and model points using the Level Set Method (LSM) (Osher and Sethian, 1988; Sethian, 1999b) and the Fast Marching Method (FMM) (Sethian, 1996; Sethian, 1999b), which are techniques to evolve boundaries using appropriate partial differential equations (PDEs). The FMM-based OA methods are shown to be cheaper and more robust than others, in particular than those based on solving diffusion-based PDEs. We find that higher-order discretizations of the level-set PDEs increase the accuracy of distance estimates; second-order schemes being sufficient for most applications. We show that the covariance matrices are not necessarily positive definite because the Weiner Khinchin and Bochner theorems for positive definiteness (e.g. Papoulis, 1991), are only valid for convex simply-connected domains. Several approaches to overcome this issue are presented and evaluated. The solutions we propose include introducing a small process noise or, better, reducing the covariance matrix based on the dominant singular value decomposition.

Our new methods are expected to have many applications, in particular to improve the World Ocean Atlas (WOA) climatologies in complex multiply-connected domains. The WOA provides global ocean climatology containing monthly, seasonal and annual means of temperature (T) and salinity (S) fields at standard ocean depths. The temperature and salinity climatologies of the WOA (Levitus, 1982), which is also termed as 'Levitus Climatology'; and its updates in 1994 (Levitus and Boyer, 1994; Levitus et al., 1994), 1998 (Antonov et al., 1998a,b,c; Boyer et al., 1998a,b,c), 2001 (Stephens et al., 2002; Boyer et al., 2002) and 2005 (Locarnini et al., 2006; Antonov et al., 2006; Garcia et al., 2006a,b), have proven to be valuable tools for studying the hydrographic structures of the World's oceans. The WOA climatologies have been particularly useful for providing initial and boundary conditions to ocean circulation models. The OA procedure for the 'Levitus Climatology' requires the use of an analytical correlation function to determine the covariance (or weight function, as described by Levitus (1982)). If the "straight Euclidean distance" (the straight line distance between two points) is used in such analytical correlation functions, the distance estimate is inappropriate for complex multiply-connected domains, as it goes across land and so violates all coastline/bottom constraints. In particular, unconnected water masses are then erroneously blended across landforms, leading to artificial water masses, spurious currents and other fictitious features. The aim of our new methodologies is to satisfy all geometric constraints arising in complex multiply-connected domains and so rectify these issues.

The paper is organized as follows. The problems addressed are described in Section 2. In Section 3, we review the two staged multi-scale statistical field mapping approach from MSEAS. In Section 4, we introduce the new OA methodologies based on the Level Set and Fast Marching Methods. An approach for computing the transport streamfunction and total velocity under geostrophic balance by optimizing the unknown inter-island transports is also discussed. The OA approach based on the stochastically forced partial differential equations (SPDE) is introduced in Section 5. In Section 6, applications of our new methodologies, for the complex regions of Dabob Bay and Philippines Archipelago are presented. In Section 7, we study the computational properties of our new mapping schemes. Section 8 consists of a summary and conclusions. The scheme to compute the 'Levitus Climatology' maps is summarized in Appendix A, the FMM algorithm in Appendix B and the algorithm for optimizing unknown inter-island transports in Appendix C.

2. Problem statement

A domain is said to be convex if, for every pair of points within the domain, every point on the straight line segment that joins them is also within the domain. A domain is simply-connected if any closed curve within it can be continuously shrunk to a point without leaving the domain. A domain which is not simply-connected is multiply-connected.

A main research question is field mapping via OAs in complex multiply-connected coastal domains. OA schemes require a description of field statistics which is often provided by analytical correlation functions (Carter and Robinson, 1987; Lam et al., 2009). Such analytical functions are dependent on the spatial separation distance. Using "straight Euclidean" distances in complex multiply-connected domains is not appropriate since there is no direct relationship across landforms. An appropriate measure of distance should be longer. The most straightforward is the length of the shortest sea path, i.e. the shortest path without going across complex landforms. Examples of such paths that we computed for the Monterey Bay, Massachusetts Bay, Dabob Bay and Philippines Archipelago are illustrated in Fig. 1. Our new methodology efficiently measures these distances. It also allows for altering distances to account for dynamical or other effects. For example, we can estimate 3D shortest sea paths and weight vertical distances more than horizontal ones, hence accounting for effects of reduced correlations across depths. In general, any coordinate system can be used, e.g. Cartesian, terrain-following or density-based. If, instead of depth, density surfaces are used, diapycnal distances can be weighted more than isopycnal ones. All of these generalizations of the shortest sea path, as well as correlation functions that are constrained by dynamical or feature-based considerations, can be easily accommodated in our OA methodology.

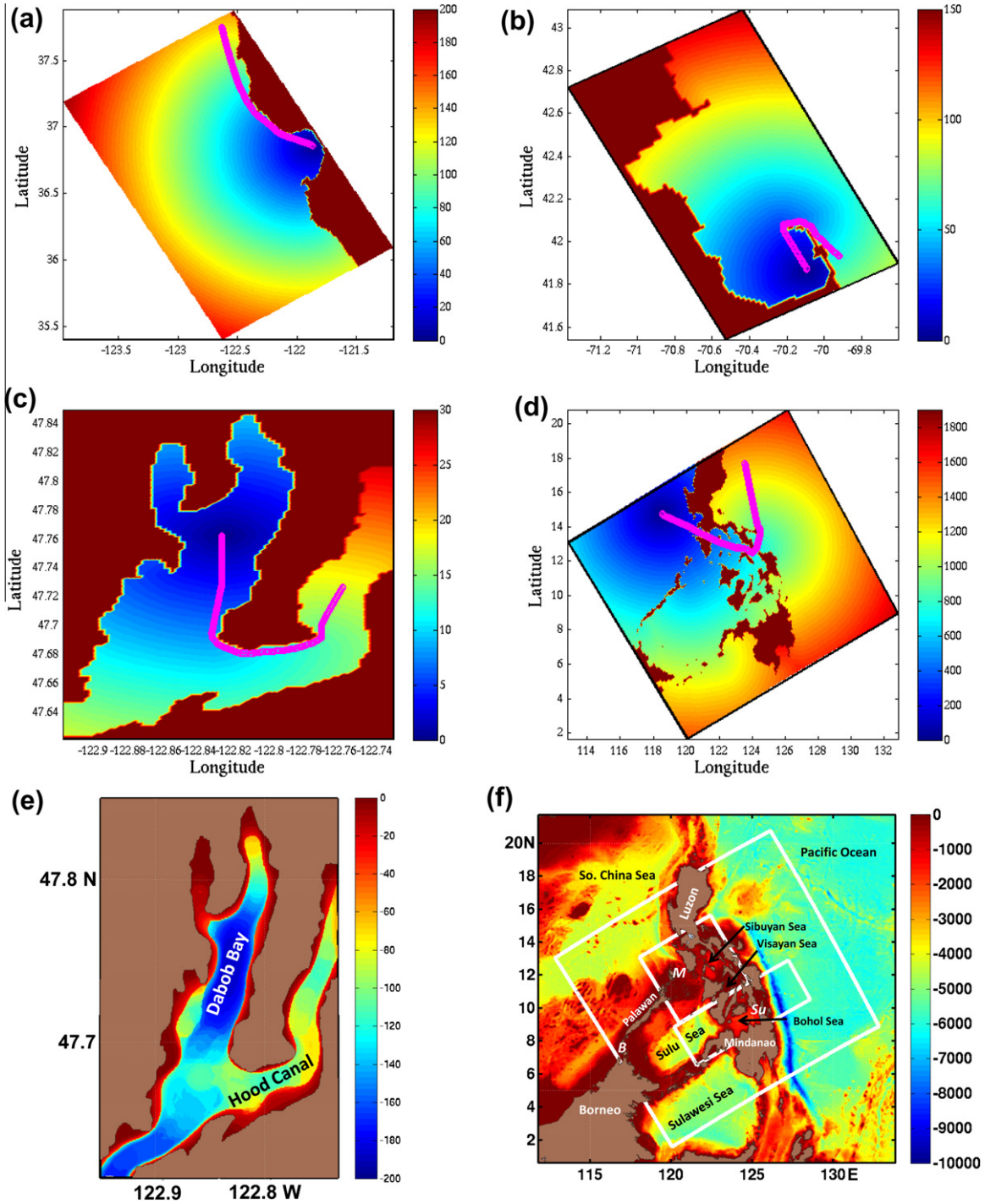

Fig. 1. Examples of optimal shortest sea paths overlaid on shortest distance (km) fields, both computed using the Level Set Method in: (a) Monterey Bay, (b) Massachusetts Bay, (c) Dabob Bay, (d) Philippines Archipelago. Also shown are domains overlaid on bathymetry (m) with locations of islands, straits and seas in (e) Dabob Bay and (f) Philippines Archipelago (abbreviations: B, Balabac Strait; M, Mindoro Strait; Su, Surigao Strait).

The shortest sea paths, or the above generalizations of such paths, in complex multiply-connected regions, can be efficiently obtained using the following techniques: the Level Set Method and the Fast Marching Method. These methods model the propagation of evolving boundaries using appropriate PDE's. Here, we illustrate their applications for realistic OAs in the Philippines Archipelago and Dabob Bay (WA, USA) regions. Other optimization methods for path planning, e.g. Dijkstra's algorithm (Bertsimas and Tsitsiklis, 1997) and Bresenham-based line algorithm (Bresenham, 1965), can also be used for mapping in complex domains, but we find and show that the FMM and LSM schemes are computationally more efficient and accurate. We also compare our results to the OA approach based on solving stochastically forced PDEs (Balgovind et al., 1983; Lynch and McGillicuddy, 2001).

The FMM and LSM can also be utilized for estimating the minimum vertical area along any path between two islands. This estimation is very efficient, for all island pairs in complex domains with many islands. Such areas are needed to estimate total velocities and transports under a geostrophic constraint (Wunsch, 1996). Specifically, these vertical areas are used here in the

Table 1
Comparison between the Kalman filter and the MSEAS OA update equations (for a univariate variable and one scale).

KF update equations	MSEAS OA update equations
Kalman gain: $\mathbf{K}_t = \mathbf{P}_{t\|t-1}\mathbf{H}_t^T \times [\mathbf{H}_t\mathbf{P}_{t\|t-1}\mathbf{H}_t^T + \mathbf{R}_t]^{-1}$	OA gain: $\mathbf{K} = Cor(\mathbf{x},\mathbf{X}) \times [Cor(\mathbf{X},\mathbf{X}) + \mathbf{R}]^{-1}$
State estimate update: $\hat{\mathbf{x}}_\mathbf{t} = \hat{\mathbf{x}}_{\mathbf{t}\|\mathbf{t-1}} + \mathbf{K}_\mathbf{t}(\mathbf{y}_\mathbf{t} - \mathbf{H}_\mathbf{t}\hat{\mathbf{x}}_{\mathbf{t}\|\mathbf{t-1}})$	State estimate update: $\psi^{OA} = \bar{\psi} + \mathbf{K}[\mathbf{d} - \bar{\mathbf{d}}]$
Error covariance update: $\mathbf{P}_\mathbf{t} = (\mathbf{I} - \mathbf{K}_\mathbf{t}\mathbf{H}_\mathbf{t})\,\mathbf{P}_{\mathbf{t}\|\mathbf{t-1}}$	Error covariance update: $\mathbf{P}^{OA} = Cor(\mathbf{x},\mathbf{x}) - \mathbf{K}^{OA} \times Cor(\mathbf{X},\mathbf{x})$

inversion for the transport streamfunction along the island coastlines. The resulting temperature, salinity and velocity field estimates can then be used as first-guess in 3D mapping of primitive-equation fields and error covariances (Lermusiaux, 2002).

Mathematical and computational properties of the new mapping schemes are also investigated in detail. To reduce the computational cost and to understand the impact of individual data, sequential processing of observations (Parrish and Cohn, 1985; Cho et al., 1996) is utilized. By definition, the prior covariance matrix should be positive definite. According to the Wiener-Khinchin and Bochner theorem (Papoulis, 1991; Yaglom, 2004; Dolloff et al., 2006), a covariance matrix based on an analytical correlation function will be positive definite if the Fourier transform (or the spectral density of the correlation function) is non-negative for all frequencies. These theorems are valid only for convex simply-connected domains. In our complex multiply-connected domains, the covariance matrix may become negative due to: (a) numerical errors in the computation of the shortest sea path's length using FMM/LSM OA schemes, or, (b) the presence of landforms. These issues may lead to divergence problems (Brown and Hwang, 1997) in the field mapping. Therefore, the following two questions were investigated and resolved: (a) What are the computational errors in the sea path lengths computed using the FMM/LSM and how can they be reduced?, and (b) What are the computational issues, including non-positive definite covariances, that arise in mapping data in a multiply-connected coastal domain and how can they be remedied? Answering these questions was indispensable for the development of FMM/LSM OA schemes for complex multiply-connected domains.

3. MSEAS objective analysis approach

Bayesian-based OA schemes are well established for mapping heterogeneous, multivariate, irregular data (Gandin, 1965; Bretherton et al., 1976; Carter and Robinson, 1987; Daley, 1993) in open oceans, without islands or archipelagos. Most OA schemes utilize the Gauss-Markov or minimum error variance criterion (Plackett, 1950) to map observations to the numerical grid and they require the computation of Euclidean distances between all data and model points. Within MSEAS, our multi-scale OA scheme consists of the successive utilization of Kalman update steps, one for each scale and for each correlation across scales (Lermusiaux et al., 2000). In particular, our two-scale OA version is summarized in Lermusiaux (1997, 1999).

Considering one scale or one interaction between two scales, let us denote the vector of numerical grid point locations as $\mathbf{x}$ and the vector of measurement locations as $\mathbf{X}$, then the OA estimate of the field for that scale or interaction (ψ^{OA}) based on the latest background field ($\bar{\psi}, \bar{\mathbf{d}}$) is given by:

$$\begin{aligned}\psi^{OA} &= \bar{\psi} + Cor(\mathbf{x},\mathbf{X})[Cor(\mathbf{X},\mathbf{X}) + \mathbf{R}]^{-1}\left[\mathbf{d} - \bar{\mathbf{d}}\right] \\ &= \bar{\psi} + \mathbf{K}^{OA}\left[\mathbf{d} - \bar{\mathbf{d}}\right],\end{aligned} \tag{1}$$

where $Cor(\mathbf{x},\mathbf{X})$ is the correlation matrix between grid and data points (for multivariate OAs or 3D OAs, it is a normalized covariance matrix, see Lermusiaux (2002)), $\bar{\mathbf{d}} = \mathbf{H}\bar{\psi}$, $\mathbf{H}$ is the observation matrix, $\mathbf{d}$ is the sensor data vector, $\mathbf{R}$ is the error covariance matrix for the sensor data $\mathbf{d}$ (for the scale considered) at data points, and the gain $\mathbf{K}^{OA}$ is given by:

$$\mathbf{K}^{OA} = Cor(\mathbf{x},\mathbf{X})[Cor(\mathbf{X},\mathbf{X}) + \mathbf{R}]^{-1}. \tag{2}$$

The error covariance of the estimated field (for one scale) is then given by (where $E[]$ denotes the expectation operator):

$$\begin{aligned}\mathbf{P}^{OA} &= E\left[(\mathbf{x} - E[\mathbf{x}])(\mathbf{x} - E[\mathbf{x}])^T\right] \\ &= Cor(\mathbf{x},\mathbf{x}) - \mathbf{K}^{OA}Cor(\mathbf{X},\mathbf{x}).\end{aligned} \tag{3}$$

A comparison between our above update equations for the OA for one scale and the Kalman filter (KF) update equations (using underscore t to indicate time t is made in Table 1).

Thus, if covariances in time are not considered, the update equations of the OA of one scale are equivalent to the update equations of the discrete Kalman filter algorithm. The background error correlation matrix for the field-to-data points, $Cor(\mathbf{x},\mathbf{X})$, and the background correlation matrix at the data points, $Cor(\mathbf{X},\mathbf{X})$, are directly related to the KF a priori error covariance matrix $\mathbf{P}_{t|t-1}$ i.e. $Cor(\mathbf{x},\mathbf{X}) = \mathbf{P}_{\mathbf{t}|\mathbf{t-1}}\mathbf{H}_\mathbf{t}^\mathbf{T}$ and $Cor(\mathbf{X},\mathbf{X}) = \mathbf{H}_\mathbf{t}\mathbf{P}_{\mathbf{t}|\mathbf{t-1}}\mathbf{H}_\mathbf{t}^\mathbf{T}$ ($\mathbf{H}_t$ is the observation matrix). In 2D horizontal OAs for a single variable, the matrix $\mathbf{R}$ is often chosen diagonal with a uniform non-dimensional observational error variance σ_d^2, i.e. $\mathbf{R} = \sigma_d^2\mathbf{I}$. In MSEAS, the correlation matrices for a given scale are usually generated from the isotropic function:

$$Cor(r) = \left(1 - \frac{r^2}{L_0^2}\right)\exp\left[-0.5 \times \left(\frac{r^2}{L_e^2} + \frac{\Delta t^2}{\tau^2}\right)\right]. \tag{4}$$

Here, Δt is the difference between the observation and estimation times, and τ is the decorrelation time scale. This time effect extends the Kalman update step at a single time to a smoothing OA step using data from different but synoptic times. The parameters L_0 and L_e are the zero-crossing and the e-folding length scales. The scalar r is the spatial separation distance.

The MSEAS OAs are often carried out in two stages (Lermusiaux, 1999). In the first stage, the largest dynamical scales (denoted LS) are mapped onto the computational grid using the parameters $(\tau, L_0, L_e)_{LS}$. The background field for this stage is often chosen to be equal to the horizontal mean of all the observations. In the second stage, the smaller scales are mapped using the coefficients $(\tau, L_0, L_e)_{ME}$ often corresponding to the most energetic (meso) scales. The background field for this stage is the first stage OA. A major assumption in this scheme is that the errors in the largest and the most energetic stages are statistically independent. A 3D and dynamics-based extension of this scheme, including multiscale interactions, is presented in Lermusiaux (2002); this 3D multiscale approach also benefits from our new efficient estimation of shortest sea paths. Of course, the accuracy of the field estimates also depends on the spatial and time scale parameters used in the analytical correlation function, as well as on the correlation function itself. The 2D horizontal version of the MSEAS OA has many similarities with the approach used for 'Levitus Climatology' maps which is described in Appendix A.

The 'Levitus Climatology' and MSEAS OA mapping schemes compute the covariance or weight factors by providing Euclidean

distances as inputs to correlation functions. If they are not employed in open ocean conditions, actual sea distances between data and model points without going across complex landforms or through bathymetry are needed. The LSM or FMM presented next in Section 4 are used to obtain such shortest sea distances in complex (e.g. multi-island) multiply-connected coastal regions.

4. Methodologies for estimating the length of shortest sea paths in complex coastal regions and archipelagos

The shortest sea paths between data and model-grid points in complex multiply-connected coastal regions are efficiently computed using the LSM and FMM. These paths are then input to our MSEAS software for multiscale OAs. The LSM and FMM methods are both more accurate and computationally cheaper than the conventional Bresenham-based line algorithm (Bresenham, 1965) and Dijkstra's algorithm (Bertsimas and Tsitsiklis, 1997). Comparisons to these other algorithms are discussed in (Agarwal, 2009); key results are summarized in Section 7.1.

4.1. Objective analysis using the level set Method (LSM)

A level set of a real-valued function ϕ of n variables is a set of the form:

$$\{(x_1,\ldots,x_n)|\phi(x_1,\ldots,x_n)=c\}, \tag{5}$$

where c is a constant and x_i are the n variables. That is, a level set is the set of points where the function ϕ takes on a given constant value c.

Osher and Sethian (1988) proposed a numerical technique, called the Level Set Method, to implicitly represent and model the propagation of evolving level set interfaces under the influence of a given velocity field using appropriate PDEs. An initial value formulation describing the interface motion is now discussed. The initial position of interfaces are given by level sets of the function ϕ. The evolution of this function ϕ is linked to the propagation of the interface through a time-dependent level set equation. Interfaces can be represented explicitly (parameterized interfaces, i.e. interfaces given by $\mathbf{x}=\mathbf{x}(s)$, where s is the parameter) or implicitly (e.g. interfaces given by the zero level set i.e. $\phi(\mathbf{x})=0$). Using the implicit representation $\phi(\mathbf{x})$, where $\mathbf{x}$ is the position vector, a convection equation can be solved to propagate level sets advected by a velocity field $\mathbf{v}$:

$$\phi_t+\mathbf{v}\cdot\nabla\phi=0. \tag{6}$$

In many cases, one is interested only in the motion normal to the boundary. Therefore, the velocity $\mathbf{v}$ can be represented using the scalar speed function F and the normal direction $\mathbf{n}$. Thus:

$$\mathbf{v}=F\mathbf{n}=F\frac{\nabla\phi}{|\nabla\phi|}. \tag{7}$$

The hyperbolic, non-linear (Hamilton–Jacobi equation) level set equation, obtained from Eqs. (6) and (7), is given by

$$\phi_t+F|\nabla\phi|=0. \tag{8}$$

Integrating the level set equation is an initial value problem which tracks the evolution of the level sets ϕ = constant assuming F is given by the specifics of the evolution of the ϕ for a particular problem. The following first order upwinded finite difference approximation can be used to integrate this Eq. (8) (two-dimensional in space) (Osher and Sethian, 1988; Sethian, 1999b):

$$\phi_{ij}^{n+1}=\phi_{ij}^{n}-\Delta t\left[\max(F,0)\nabla_{ij}^{+}+\min(F,0)\nabla_{ij}^{-}\right],$$

where,

$$\begin{aligned}\nabla_{ij}^{+}&=\Big[\max\left(D^{-x}\phi_{ij}^{n},0\right)^2+\min\left(D^{+x}\phi_{ij}^{n},0\right)^2\\&\quad+\max\left(D^{-y}\phi_{ij}^{n},0\right)^2+\min\left(D^{+y}\phi_{ij}^{n},0\right)^2\Big]^{1/2},\\\nabla_{ij}^{-}&=\Big[\min\left(D^{-x}\phi_{ij}^{n},0\right)^2+\max\left(D^{+x}\phi_{ij}^{n},0\right)^2\\&\quad+\min\left(D^{-y}\phi_{ij}^{n},0\right)^2+\max\left(D^{+y}\phi_{ij}^{n},0\right)^2\Big]^{1/2}.\end{aligned} \tag{9}$$

Here, D^{-x} is the first order backward difference operator in the x-direction; D^{+x} is the first order forward difference operator in the x-direction, etc. Mathematically, these operators are given by:

$$D^{-x}\phi_{i,j}=\frac{\phi_{i,j}-\phi_{i-1,j}}{\Delta x},\quad D^{+x}\phi_{i,j}=\frac{\phi_{i+1,j}-\phi_{i,j}}{\Delta x}. \tag{10}$$

The above numerical technique of the Level Set Method can be used to solve the Eikonal equation. If the scalar speed function of the front F is non-negative, then the steady state boundary value problem, known as the Eikonal equation, can be formulated to evaluate the arrival time function T ($\mathbf{x}$). The Eikonal equation representing the time T ($\mathbf{x}$) for the "frontal interface" to reach the position $\mathbf{x}$ from its initial position is given by

$$F|\nabla T|=1. \tag{11}$$

The Eikonal equation simply states that the gradient of the arrival time function is inversely proportional to the local speed of the front. To solve the Eikonal equation, a time dependent problem is proposed. The time evolved steady state solution of the resultant Hamilton–Jacobi equation is the Eikonal equation. Mathematically, this is written as:

$$T_t+F|\nabla T|=1\overset{\text{steady}}{\rightarrow}F|\nabla T|=1. \tag{12}$$

This Hamilton–Jacobi equation (Eq. (12) (left)) can be discretized using the numerical scheme for the Level Set equation. The steady state solution of this Hamilton–Jacobi equation will be the solution of the Eikonal equation (Eq. (12) (right)).

The Level Set Method has been used in a wide variety of applications; including arrival time problems in control theory, generation of minimal surfaces, flame propagation, fluid interfaces, shape reconstruction, etc. In the oceanic context, the method can be used to determine shortest sea path lengths as follows. The scalar speed function F is set to 0 for the grid points on land and 1 for the grid points on water. The level set $T(\mathbf{x})$, which is the arrival time function, then also represents the shortest sea distance from the starting position to the position vector $\mathbf{x}$. This is because the level set T, which is the arrival time, when multiplied by the local speed of the front (equal to 1 in this case) gives the level set T itself for the shortest sea path length estimate. Once these sea distances between all data points and model points are available, the prior correlation functions can be evaluated and the correlation matrices filled in Eq. (1). An OA can then be computed.

Operation count for the LSM: The computation of shortest sea paths via the LSM requires evolving all the level sets in Eq. (8) and not simply the zero level set corresponding to the front itself. The LSM thus has an operation count of $O(N^3)$ in two dimensions for N^2 grid points (Sethian, 1999b). It is computationally expensive since an extra dimension is added.

A modified method named 'Fast Marching' significantly reduces the operation count. Roughly speaking, the two possible ways to obtain steady-state are either iteration towards the solution, or direct construction of the stationary solution T. While the LSM constructs the solution to the Eikonal equation (Eq. (11)) by iterating towards it, the FMM directly constructs it.

4.2. Objective analysis using the Fast Marching Method (FMM)

The Fast Marching Method (FMM) for monotonically advancing fronts has been proposed by Sethian (1996, 1999b). This method leads to a very fast scheme for solving the Eikonal equation (Eq. (11)). The LSM relies on computing the evolution of all the level sets by solving an initial value PDE using numerical techniques from hyperbolic conservation laws. This is because the LSM iteratively solves the level set equation to compute the steady state solution of Eq. (12). As an alternative, an efficient modification is to perform the work only in the neighborhood of the zero level set, as this is known as the 'narrow band approach'. The basic idea is to tag the grid points as either "alive", "land mines" or "far away" depending on whether they are inside the band, near its boundary, or outside the band, respectively. The work is performed only on "alive" points, and the band is reconstructed once the "land mine" points are reached.

The FMM solves boundary value problems without iterations. The method is applicable to monotonically advancing fronts, i.e. the front speed $F \geqslant 0$ or $F \leqslant 0$ in the level set equation (Eq. (12)). The steady state form of the level set equation is the Eikonal equation (11). For the two dimensional case, the stationary boundary value problem is given by:

$$|\nabla T| F(x,y) = 1 \quad \text{s.t.} \quad \Gamma = \{(x,y) | T(x,y) = 0\}, \tag{13}$$

where Γ is the starting position of the interface. The first order finite difference discretization form of the Eikonal equation (Sethian, 1999b) at the grid point (i,j) is given by:

$$\left[\max\left(D_{ij}^{-x}T,0\right)^2 + \min\left(D_{ij}^{+x}T,0\right)^2 + \max\left(D_{ij}^{-y}T,0\right)^2 + \min\left(D_{ij}^{+y}T,0\right)^2\right]^{1/2} = \frac{1}{F_{ij}},$$

or,

$$\left[\max\left(\max\left(D_{ij}^{-x}T,0\right), -\min\left(D_{ij}^{+x}T,0\right)\right)^2 + \max\left(\max\left(D_{ij}^{-y}T,0\right), -\min\left(D_{ij}^{+y}T,0\right)\right)^2\right] = \frac{1}{F_{ij}^2}. \tag{14}$$

Eq. (14) is essentially a quadratic equation for the value at each grid point (assuming that values at the neighboring nodes are known). An iterative algorithm for computing the solution to Eq. (14) was introduced by Ruoy and Tourin (1992). FMM is based on the observation that the upwind difference structure of Eq. (14) means that the information propagates "one way", i.e. from the smaller values of T to the larger values. Therefore, FMM rests on solving Eq. (14) by building the solution outward from the smallest time value T. The front is swept ahead in an upwind manner by considering a set of points in a *narrow band* around the existing front and bringing new points into the *narrow band* structure. The fast marching algorithm is discussed in detail in Appendix B (see also Agarwal, 2009).

The use of higher-order FMMs (or LSMs) to reduce errors in the estimation of shortest sea path lengths is discussed in Section 7.2. They are computationally more expensive but can be necessary for robust and accurate OAs because in complex multiply-connected domains, we found that covariance matrices were sensitive to the accuracy of these lengths. These findings are discussed later in Sections 7.2 and 7.3.

Operation count for the FMM: Once again, for estimating the optimal distance, the scalar speed function F is set to 0 for the grid points on land and 1 for the grid points on water. However, the FMM has a significantly lower operation count of $O(N^2 \text{ Log } N)$ for N^2 grid points (Sethian, 1999b). It is computationally much cheaper than the LSM explained above.

The Fast Marching Method, as discussed above, is an efficient way to compute the sea distance between any two locations. These sea distances can then be used for setting up the covariance matrix using any distance-dependent analytical correlation function (e.g. Eq. (4)). Note that the cost of the OAs proper are the same for both the LSM and FMM.

4.3. Total velocity under geostrophic balance: estimating the minimum vertical area in complex coastal regions and archipelagos

Classically, the synoptic ocean data that are most abundant are hydrographic (temperature and salinity) measurements. If these data are first gridded by OAs, they can be used to estimate a velocity field under the constraint of geostrophic shear (Wunsch, 1996) or other momentum balance assumptions, including full momentum conservation of the primitive-equations (Lermusiaux et al., 2000; Lermusiaux, 2002). If geostrophic shear is used as the constraint, to compute transport estimates from the hydrographic OAs, a reference velocity is required. In complex domains, an estimate of the area of the sea cross-sections between any two landforms (e.g. islands) is also often necessary to set the inter-islands transports. The FMM can be directly used to compute the minimum of these cross-sectional areas.

In our case, we utilize an optimization scheme to estimate these inter-island transports; see (MSEAS, 2010) and the summary in Appendix C. Its objective is to find a set of values for the transport streamfunction (Ψ) along the island coastlines that produce a suitably smooth (initial) velocity field, e.g. without unrealistic velocities. If prior estimates of specific transports between islands are known, they are utilized with their uncertainties as inputs to the optimization scheme. If such prior estimates are not available, they are set using a minimum energy principle: a norm of the total velocity between the corresponding islands is minimized under the constraint of geostrophic velocity shear balancing the hydrographic OA maps. To do so, the weight functions require an estimate of the cross-sectional area between islands. This is not easy to compute exactly without a FMM/LSM approach.

With the FMM/LSM schemes, the minimum vertical area can be obtained if we solve the Eikonal equation (Eq. (11)) setting the scalar speed function to be $F(x,y) = 1/H(x,y)$. The Eikonal equation thus simplifies to $|\nabla T| = H$, which shows that the solution $T(x,y)$ of this Eikonal equation will be the minimum vertical area. This new approach is used in Section 6 to obtain velocity estimates from our hydrographic FMM-based OA maps.

5. Objective analysis using stochastically forced partial differential equations (SPDE's)

Another OA approach that accounts for landforms uses SPDE's. The central idea is to represent the underlying field variability as an outcome of a stochastic process using a SPDE where the stochasticity represents the uncertainty in this differential equation. The SPDE is defined only over the sea domain so as to account for geometric constraints. The covariance matrix for the field is then constructed numerically, by solving a set of SPDEs over the sea domain. For example, the stochastically forced Helmholtz equations in 1-D and 2-D in space for the field ψ in an unbounded domain (Balgovind et al., 1983) are associated with the following covariance functions, respectively:

$$\begin{aligned} &\frac{\partial^2\psi}{\partial x^2} - k^2\psi = \epsilon(x) \Longleftrightarrow C_{\psi\psi}(r) = (1+kr)e^{(-kr)}, \\ &\nabla^2\psi - k^2\psi = \epsilon(x,y) \Longleftrightarrow C_{\psi\psi}(r) = krK_1(kr) \\ &\qquad\qquad \simeq \left(\frac{\pi}{2}kr\right)^{1/2}\left(1+\frac{3}{8kr}\right)e^{-kr}, kr \to \infty \end{aligned} \tag{15}$$

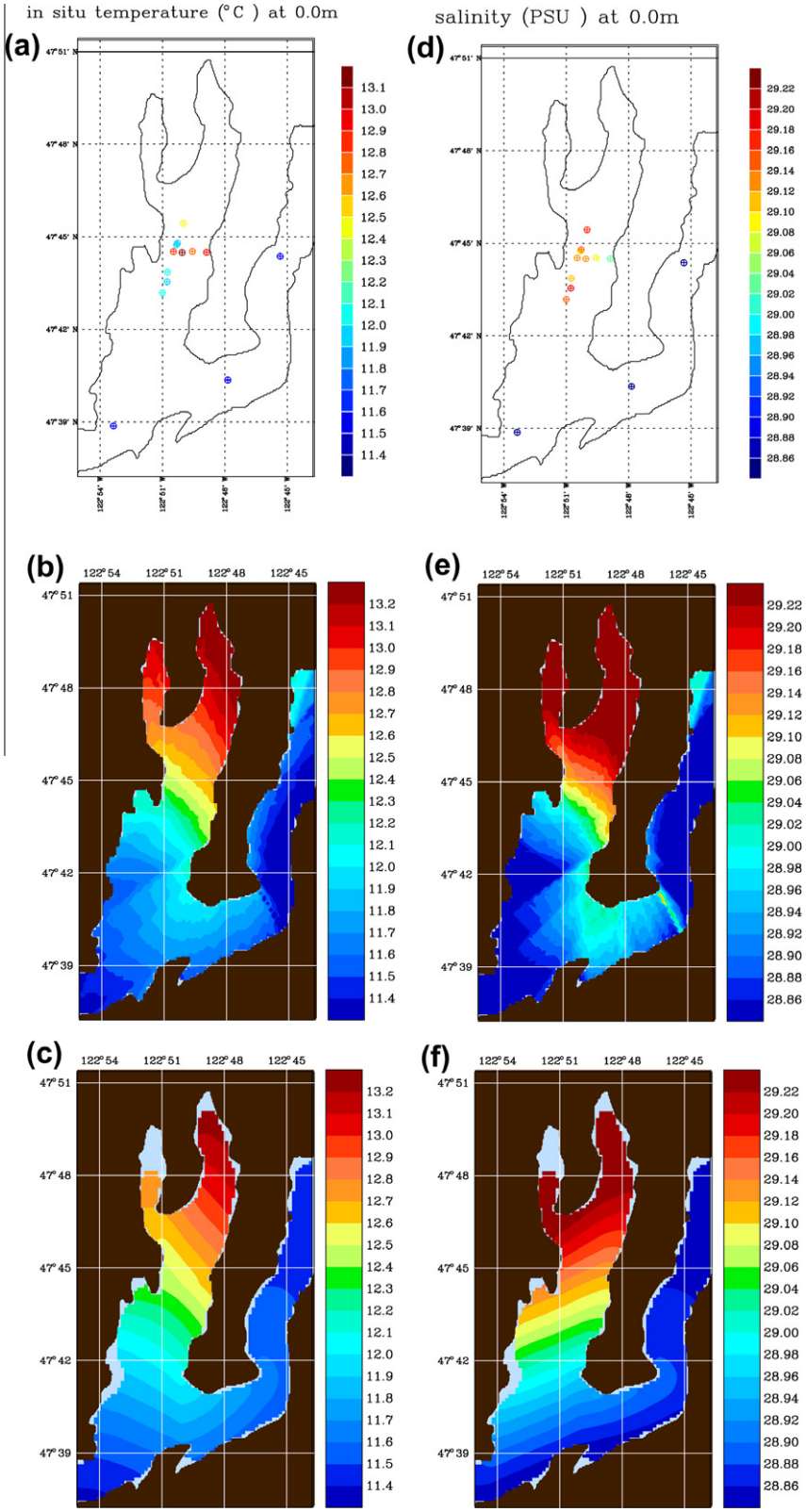

Fig. 2. Temperature (T,°C) (a) and Salinity (S, PSU) (b) data in Dabob Bay. OA fields for T (left) and S (right) from the optimal path length computed using: (c and d) Bresenham-based line algorithm; (e and f) Fast Marching Method, clearly showing the issues of the Bresenham-based line algorithm.

where K_1 is the Bessel function of the second kind. The process noise ϵ is a random disturbance with mean 0, standard deviation 1 and no spatial correlation. Also, the length scale corresponds to the inverse of the SPDE parameter (k). Denman and Freeland (1985) and Weaver and Courtier (2001) have proposed other functions which can also be linked to appropriate SPDE's.

A major advantage is that the field-to-field covariance $Cor(\mathbf{x},\mathbf{x})$ can be computed numerically from the discretized SPDE along with

appropriate boundary conditions (i.e. no flux across islands) to directly account for the coastline constraints (Lynch and McGillicuddy, 2001). The discretization of SPDEs (such as Eq. (15)) or any other differential operator defined on the sea domain usually amounts to solving a matrix equation of the form:

$$[A]\{\psi\} = \{e\}, \tag{16}$$

where $\{e\}$ is the spatial discretization of the process noise ϵ. All the coastline constraints are then incorporated automatically in this matrix form (16). Since $[C_{ee}] = [I]$, the covariance matrices for field-to-field points and field-to-data points are directly obtained from Eq. (16):

$$\begin{aligned} Cor(\mathbf{x},\mathbf{x}) &= [A]^{-1}[C_{ee}][A]^{-T} = \left([A]^T[A]\right)^{-1}, \\ Cor(\mathbf{x},\mathbf{X}) &= [A]^{-1}[C_{ee}][A]^{-T}[H]^T = \left([A]^T[A]\right)^{-1}[H]^T. \end{aligned} \tag{17}$$

The covariance matrix (17) obtained using the SPDE approach can be used along with Gauss-Markov Estimation theory (see Table 1) to perform OAs in coastal regions. A limitation of this approach is that the resulting fields can be affected by the discretization error associated with the discretized form of the SPDE. In fact, we found that we often need to postprocess (smooth out) the SPDE-gridded fields to remove spurious field gradients. Such gradients, even when small, can lead to spurious velocities by aggregate integration in the vertical for the estimation of total velocity under geostrophic balance. It has also been verified that that the SPDE approach is computationally expensive when compared to our new FMM-based methodology.

A variant of the above methodology represents the covariance function ($C_{\psi\psi}$), instead of the field (ψ), by a SPDE, e.g. a stochastic Helmholtz equation (Logutov, personal communication). The advantage is that the covariances required are then computed directly, without the need of Eq. (17), which is much cheaper. However, the noise in the resulting OA fields are then found to be even larger (Agarwal, 2009). A heuristic reason is that this simpler representation corresponds to carrying out a "smoothing" step using the Helmholtz equation only once as compared to twice in the original representation (Eq. (17)). Both of these methods, the SPDE specified for the field (ψ) and the SPDE specified for the covariance ($C_{\psi\psi}$) were implemented. They are utilized for comparisons with our LSM-based and FMM-based schemes.

Even though many different SPDE's could be utilized for mapping a field, in our examples, we selected the stochastically forced Helmholtz equation for three reasons. First, the dynamics of the atmosphere can be approximately governed on the time scale of a few days by a Helmholtz-like equation, which is the equation for the conservation of potential vorticity under the assumptions of a quasi-geostrophic, frictionless, shallow water model without topography (Balgovind et al., 1983; Pedlosky, 1987). Second, a

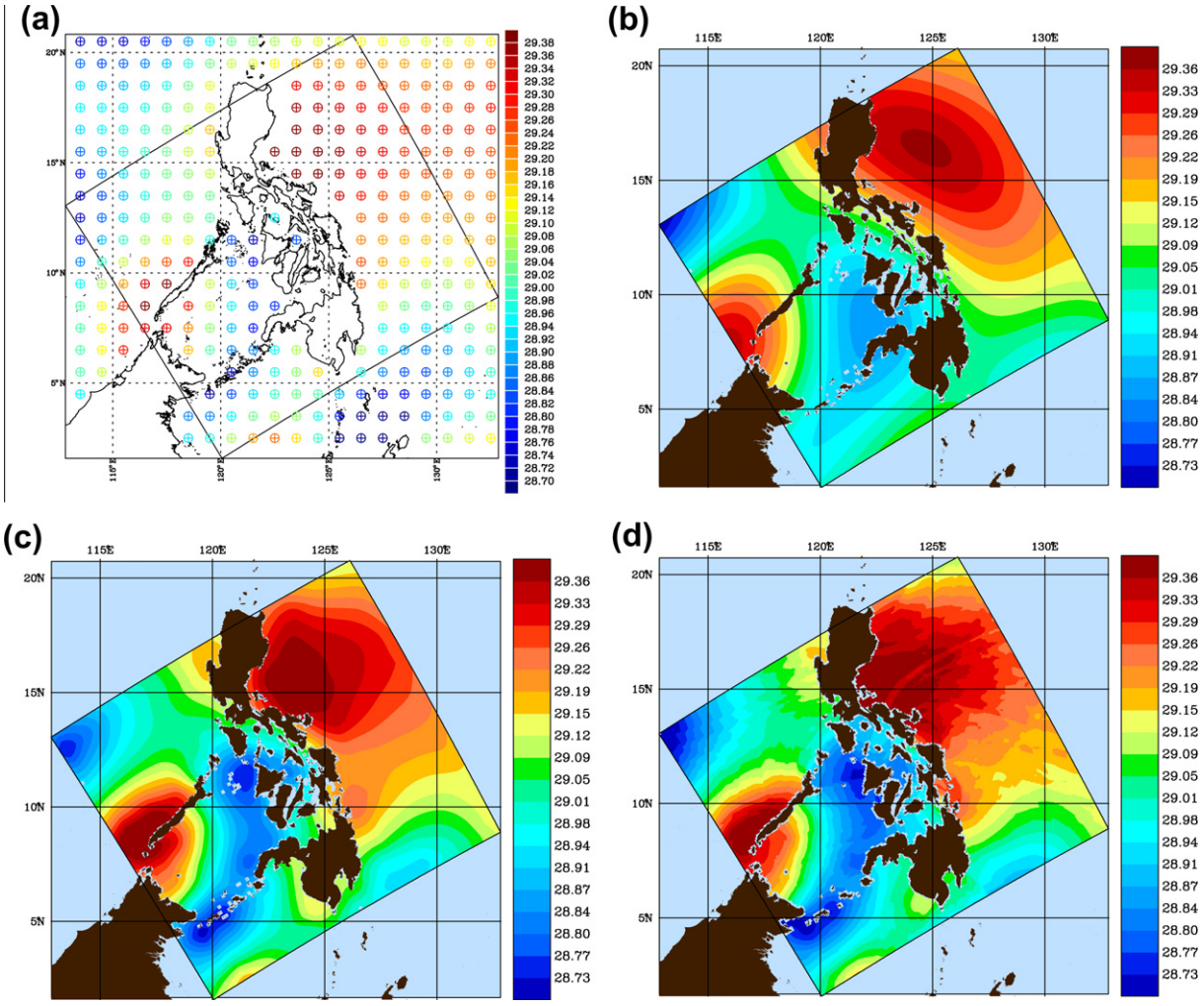

Fig. 3. (a) World Ocean Atlas 2005 Climatology in situ temperature (°C) at 0 m. Temperature (°C) OA Fields obtained using: (b) standard OA without taking islands into account, (c) Fast Marching Method, (d) SPDE approach (representing the field by a stochastically forced Helmholtz equation).

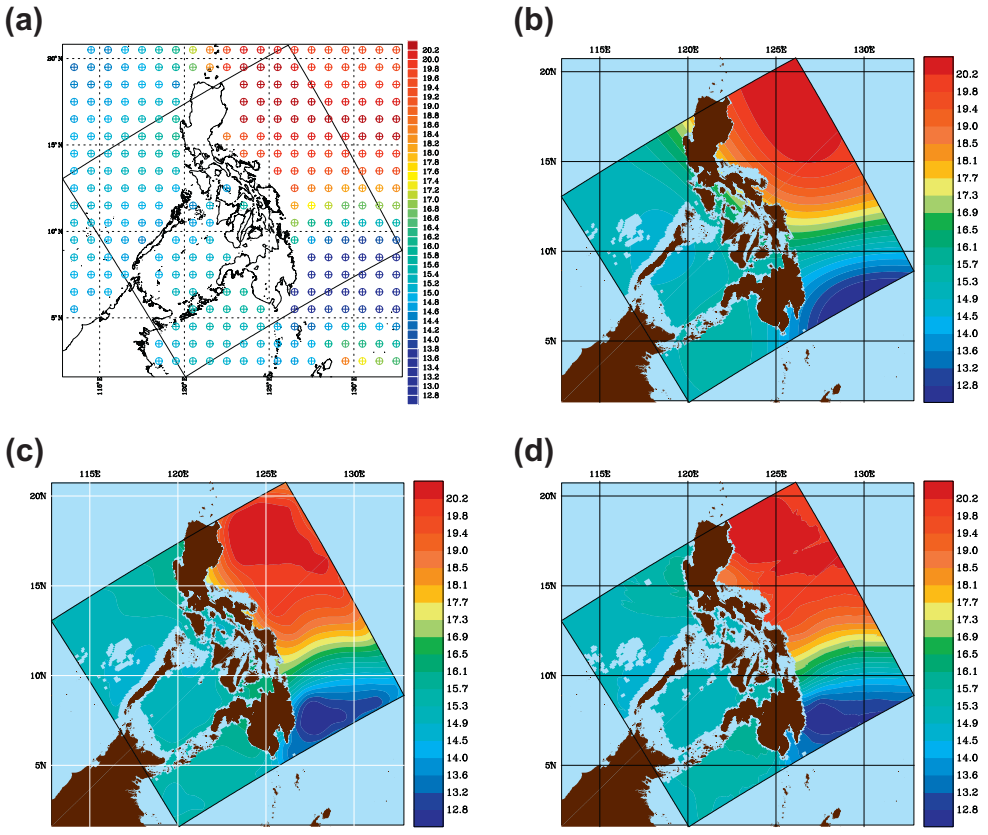

Fig. 4. As Fig. 3, but at 200.0 m.

Helmholtz equation can be obtained from the diffusion or wave equations and background correlations are seldom modeled as Gaussian, by solving a pseudo-diffusion equation (Derber and Bouttier, 1999). In these linear PDE's, if the solution is assumed separable in time and space, one obtains for the time variation an ordinary differential equation of the first order. For the spatial variations, one always obtains a Helmholtz equation (Selvadurai, 2000), which is the equation that would be used for spatial mapping. Thirdly, the Helmholtz equation is equivalent to a steady diffusion–reaction equation.

Operation count for the SPDE-based OAs: The cost of the SPDE computations of covariances with one data point is at least in $N^2 n$ but most likely in N^3 where n is the number of time-steps to reach steady state.

Meaningful comparisons among the different methods require comparable covariance parameters. Specifically, for our SPDE-based OA examples using Eq. (15), the SPDE parameter (k) is chosen such that the correlation function corresponding to the stochastically forced Helmholtz equation best fits the analytical correlation function used by our standard OA scheme and by our new LSM or FMM-based schemes, see Section 4 and Agarwal (2009). The results can then be compared to each other. This is done next in Section 6.2.

6. Applications illustrating the novel OA methodologies

Methodologies derived in Section 4 are now utilized to map temperature, salinity and biological (chlorophyll) fields using a 2-staged mapping scheme in both Dabob Bay and Philippines Archipelago (locations of islands, straits and seas are on Fig. 1). Specifically, Section 6.1 evaluates our schemes in Dabob Bay and shows that they are more effective than other classic distance optimizing algorithms, such as the Bresenham-based line algorithm (Bresenham, 1965). Section 6.2 compares methods of Sections 4 and 5 in the Philippines Archipelago region. The estimation of total velocity under geostrophic balance by minimizing unknown inter-island transports is also illustrated.

Of course, before evaluations with full data sets, we completed computational tests, both in idealized and realistic domains. These tests used either analytical (artificial) data or sub-sampled ocean data. Their aim was to ensure that our codes and algorithms were correct but also effective, giving the expected answer. One type of test was to compare innovation vectors due to one data point (for unit data, this is a row of the covariance matrix if the data point is located at a grid point). Some of these scalar data tests were also completed to study the positive definiteness of covariance matrices (see Section 7). We also compared innovation vectors based on a few data points. Finally, we completed cross-validations (e.g. Brankart and Brasseur, 1996): to do so, we sub-sample the available data to compute a field map with each method and then employ the unused part of the data to compare these field maps. These studies were done with the FMM/LSM, classic and SPDE OAs. We found that our new schemes were more accurate than regular OAs and less noisy than SPDE OAs. Most of these results are not shown, in part because the further away from data locations, the more estimates depend on the accuracy of the covariance functions or SPDEs employed. To mostly evaluate the OA algorithms and the inputs themselves to these functions or SPDEs, we focus next on the regions where there are data and we limit our examples to real data since analytical data led to the same conclusions.

6.1. Objective analysis in Dabob Bay

Dabob Bay data are used to illustrate the effectiveness of the FMM-based scheme over other distance optimizing algorithms such as the Bresenham-based line algorithm (Bresenham, 1965).

Fig. 2 shows maps of temperature and salinity fields for the spatially irregular data in the region, using the Bresenham-based line algorithm and the Fast Marching Method. The limitation of the former is that the resulting optimal distance is discontinuous. This leads to discontinuities in the covariance and also in the resultant field maps (Agarwal, 2009).

The temperature and salinity field maps (Fig. 2) were obtained using two-scale OAs: one with larger length scales $(L_0 = 60, L_e = 30)_{LS}$ and one with smaller length scales $(L_0 = 30, L_e = 15)_{ME}$; in both cases using a non-dimensional observational error variance $(\sigma_d^2 = 0.25)$. These parameter values were estimated based on data, see Agarwal (2009). Temperature and salinity data have higher values in the western arm (Fig. 2(a) and (b)). The eastern arm (Fig. 2(a) and (b)) has relatively low temperature and salinity. Effects due to the discontinuity in distance obtained from Bresenham-based line algorithm are clearly evident in Fig. 2(c) and (d). Numerical fronts with high temperature and salinity gradients exist at the intersection of the two arms. Such fronts lead to numerical problems in dynamical simulations. The geostrophic velocity obtained using these field maps is unrealistic and has high magnitudes along these fronts. A possible remedy, which reduces the discontinuity effects, is to smooth the distance by averaging distances of neighboring points (Haley, personal communication). We found that this averaging technique becomes numerically very expensive. In addition, the intensity of erroneous fronts are reduced when this averaged Bresenham-based line algorithm is used, but they still exist. Finally, when our new FMM-based scheme is used to compute distances and to compute the OAs, results are clearly devoid of any numerical fronts (Fig. 2(e) and (f)). The FMM-based scheme accurately satisfies the coastline constraints and is computationally inexpensive when compared to the Bresenham-based line algorithms.

6.2. Objective analysis in the Philippines Archipelago

A motivation of this study was the Philippines Straits Dynamics Experiment (PhilEx, Lermusiaux et al. (2011)). In such a complex coastal region, our new schemes were needed to map the very irregular datasets available and initialize simulations. Without them, major problems occurred: neither dynamical studies nor ocean forecasts could be initiated from standard OA schemes. To illustrate this, different OA schemes are compared next, specifically: our new OA methods based on the FMM, a standard OA scheme which ignores islands and uses the direct Euclidean distance, and a stochastically forced PDE scheme (SPDE specified for the field).

The World Ocean Atlas–2005 data for temperature and salinity are used. WOA05 data are data mapped using the 'Levitus Climatology' scheme (see Appendix A) and are regularly spaced. These data are used here to illustrate and discuss the comparison of dif-

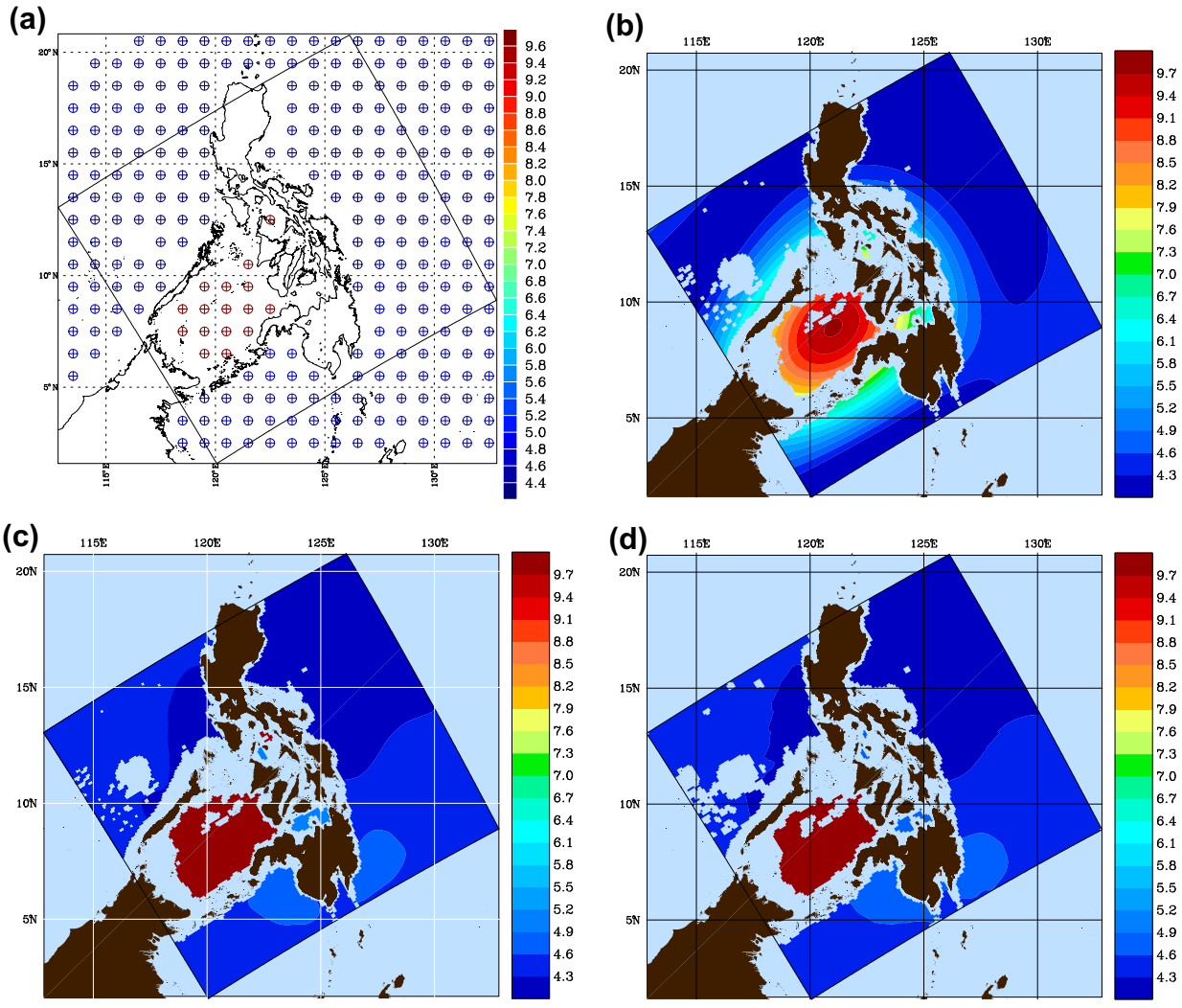

Fig. 5. As Fig. 3, but at 1000.0 m.

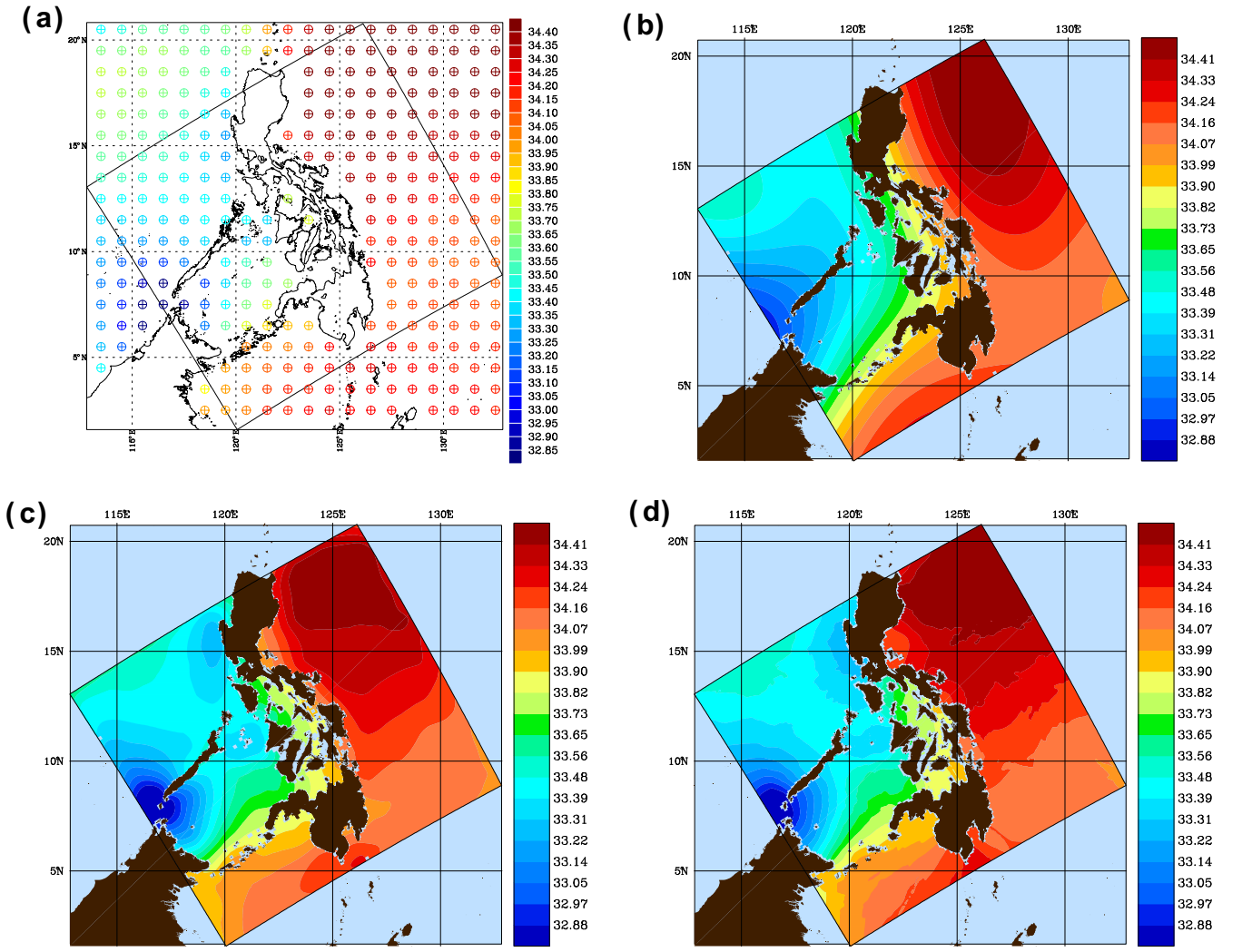

Fig. 6. (a) World Ocean Atlas 2005 Climatology in situ Salinity (PSU) at 0 m. Salinity (PSU) OA Fields obtained using: (b) standard OA without taking islands into account, (c) Fast Marching Method, (d) SPDE approach (representing the field by a stochastically forced Helmholtz equation).

ferent methodologies with a classical data set. Subsequently, synoptic in situ data sets are used for temperature, salinity and biological (chlorophyll) data.

6.2.1. Objective analysis using WOA05 data: methods comparison

Hydrographic field maps: We compare two-dimensional horizontal OA field maps of the WOA05 data (Figs. 3–8(a)) computed using schemes presented in Sections 4 and 5. Figs. 3–5 show the temperature field maps at the depth of 0 m, 200 and 1000 m, respectively. Figs. 6–8 show the salinity field maps at the depth of 0, 200 and 1000 m, respectively. For the two-scale OA schemes, the correlation function used for each scale is given in Eq. (4). The parameters are: large length scales $(L_0 = 540, L_e = 180)_{\text{LS}}$, most energetic length scales $(L_0 = 180, L_e = 60)_{\text{ME}}$ and observational error variance $\sigma_d^2 = 0.25$. These parameter values were estimated based on data, see Agarwal (2009). For the SPDE approach, the SPDE parameter k is set to 1/200 (this is a best fit to the correlation function used by the other schemes) and the observational error to $\sigma_d^2 = 0.25$.

The OA field maps from all methods (Figs. 3 and 4) indicate that the Philippines Sea and the region near Palawan island is warmer than the rest of the region near the surface (0, 200 m). The region south of the Sulu Sea around the Sulu Archipelago has relatively lower temperature. At levels below 500 m (see Fig. 5), there is a significant difference in the temperature of the Sulu Sea (warm) when compared to the rest of the region (cold) (Gamo et al., 2007; Gordon, 2009). These temperature fields show that direct correlation across landforms are likely weak. Similar observations can be made for salinity. Salinity in the Sulu Sea and South China Sea (Figs. 6 and 7) is lower than the salinity in the rest of the region near the surface (0, 200 m). At levels below 500 m, the salinity in the Sulu Sea (Fig. 8) is significantly lower than in the rest of the region. These salinity fields further support the hypothesis that direct correlation across landforms are weak.

The field maps obtained using the LSM and FMM are identical, but the FMM has a significantly lower computational cost. While the LSM constructs the distance estimate by iterating towards it, the FMM is based on the direct construction of the stationary solution (see Section 4). The OA fields obtained using LSM and FMM are very close because the FMM exactly constructs the solution of the discretized Eikonal equation whereas the LSM computes the solution within a desired tolerance limit. Thus, an OA based on FMM should clearly be preferred, as it is more accurate and less expensive. On the other hand, the SPDE approach leads to OAs that are much more noisy than those obtained using the FMM. Since the SPDE scheme is also more expensive, the FMM scheme is superior.

The comparison of the different methods for the temperature and salinity maps at 1000 m is shown in Figs. 5 and 8, respectively. The methods based on FMM (Figs. 3–8(c)) and SPDE (Figs. 3–8(d))

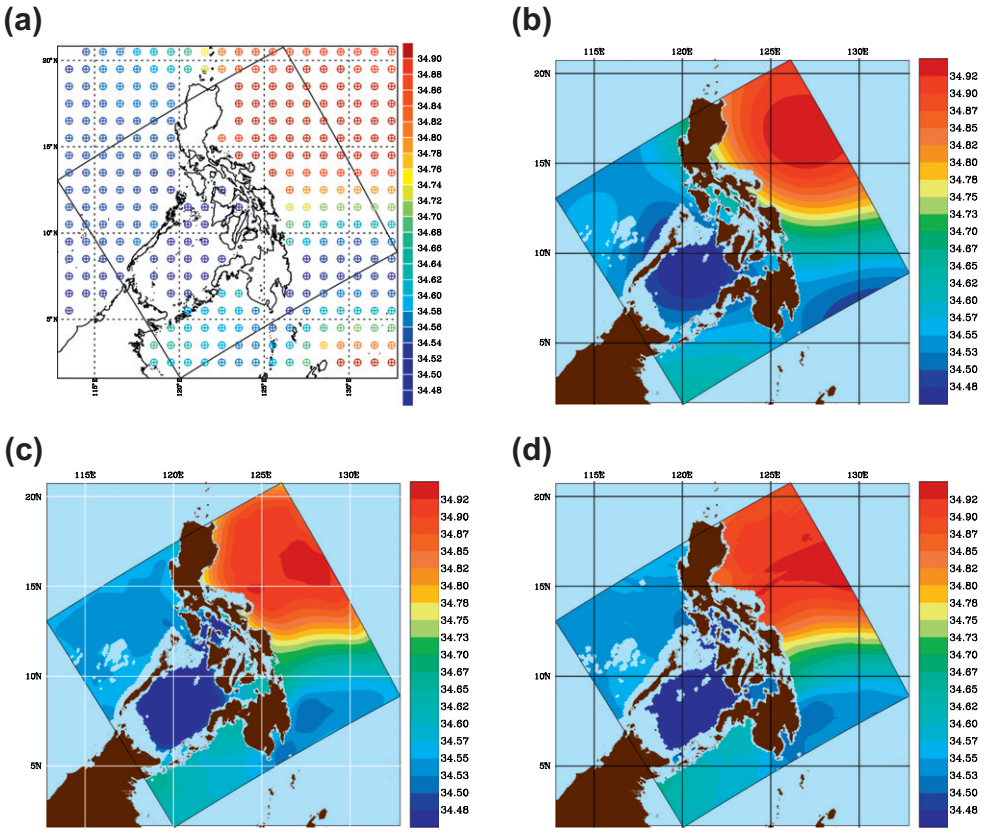

Fig. 7. As Fig. 6, but at 200.0 m.

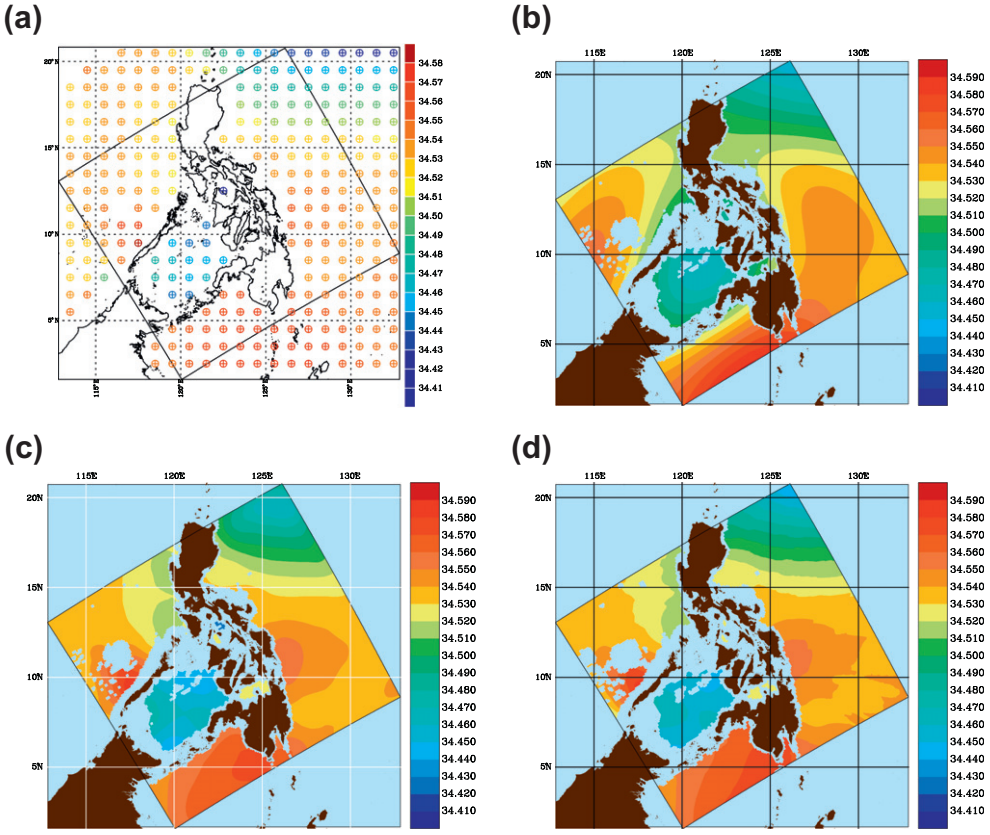

Fig. 8. As Fig. 6, but at 1000.0 m.

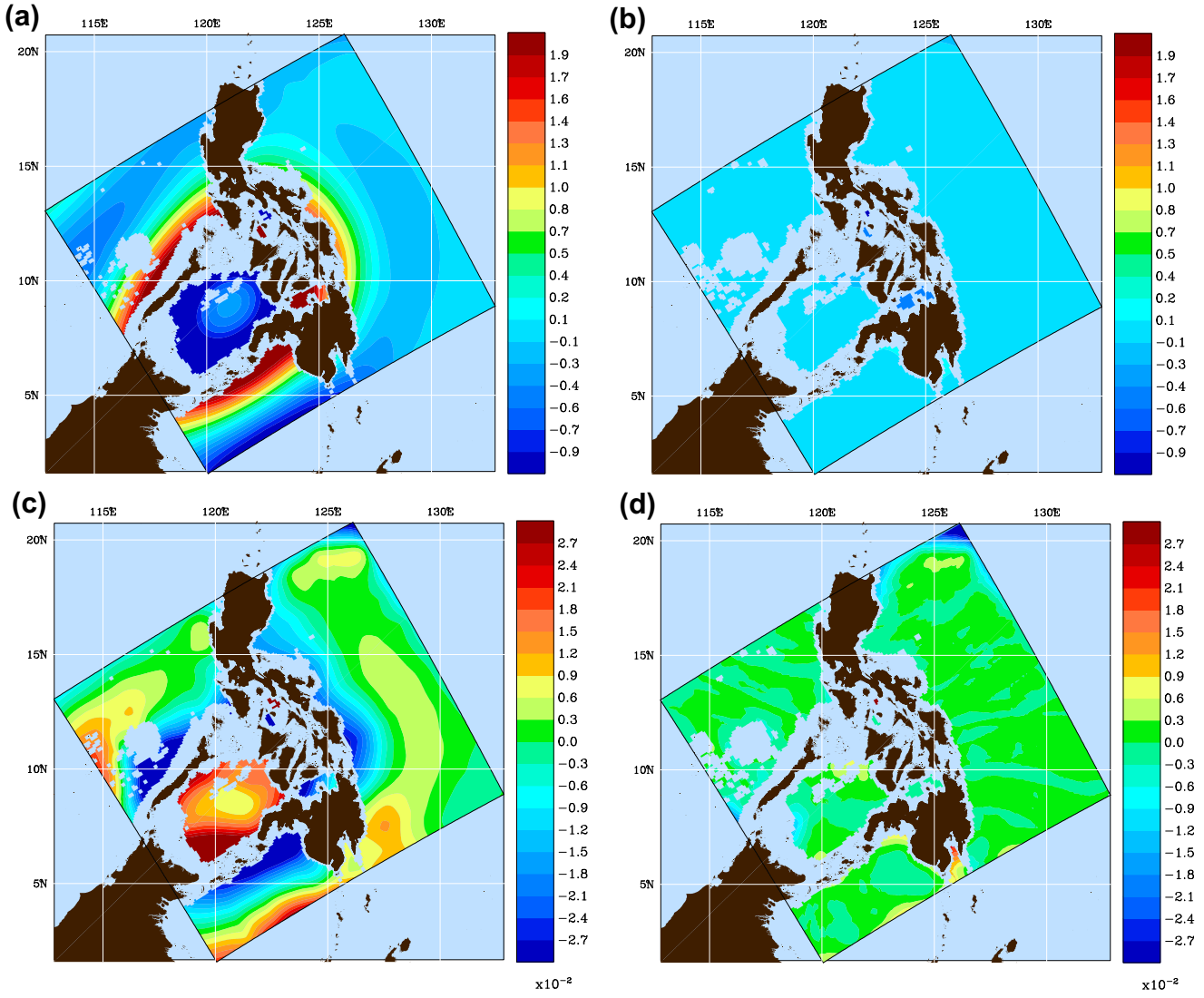

Fig. 9. Difference between Temperature (°C) fields at 1000 m obtained using Fast Marching Method and using: (a) Standard OA; (b) SPDE. Difference between Salinity (PSU) field 1000 m obtained using Fast Marching Method and using: (c) Standard OA; (d) SPDE. Both SPDEs represents the field by an Helmholtz equation.

clearly satisfy the coastline constraints. The data in the Sulu Sea, which has high temperature and low salinity compared to the remaining region, does not influence the field outside the Sulu Sea since the two regions are not connected by water at 1000 m (assuming only 2D horizontal correlations). On the other hand, the standard OA (Figs. 3–8(b)) does not satisfy the coastline constraints. Thus the data outside the Sulu Sea is correlated to the field inside the Sulu Sea. This is undesirable since the direct relationship across landforms is at best very weak. This leads to spurious high temperature and salinity gradients in the Sulu Sea, which creates large spurious geostrophic flow shear. Differences between temperature field maps and salinity field maps obtained using the FMM and using other OA methods at 1000 m are shown in Fig. 9. The differences between the field maps obtained using the FMM and standard OA are large. There are small differences between field maps obtained using the FMM and SPDE approaches because: (i) the SPDE scheme is more sensitive to truncation errors, and (ii) the analytical correlation function corresponding to the Helmholtz equation (used in the SPDE approach) is slightly different from the analytical correlation function in the FMM.

The SPDE approach satisfies the coastline constraints, but the discretization errors in the SPDE can be significant and this results in noisy spatial variations in the OA maps, even though this noise is not present in the monthly hydrographic data. This noise then also negatively affects the geostrophic flow shear, and additional smoothing (post-processing) is often needed to filter SPDE-based OA fields. Such post-processing is not required for our FMM-based scheme. As mentioned in Section 5, an SPDE approach can be implemented by specifying the SPDE for the field (as shown in Figs. 3–8(d)) or by specifying it directly for the covariance. The latter scheme is a bit cheaper than the former but it is a rough approximation and it further increases the undesired noise of the field maps. Finally, the computational time required by the SPDE approach was confirmed to be higher than that of the FMM, in accord with the operation counts of Section 5. Thus, the FMM appears to be the best among all the methods of Sections 4 and 5. This was confirmed in many other regions and the FMM scheme is thus used to map the spatially irregular synoptic data in the sections that follow.

Velocity field maps. We now illustrate the estimation of total velocity under geostrophic balance in the region using the above OA field maps of hydrographic WOA05 data. The algorithm for optimizing inter-island transports (Appendix C) is utilized to compute a smooth total flow field estimate under the constraint of geo-

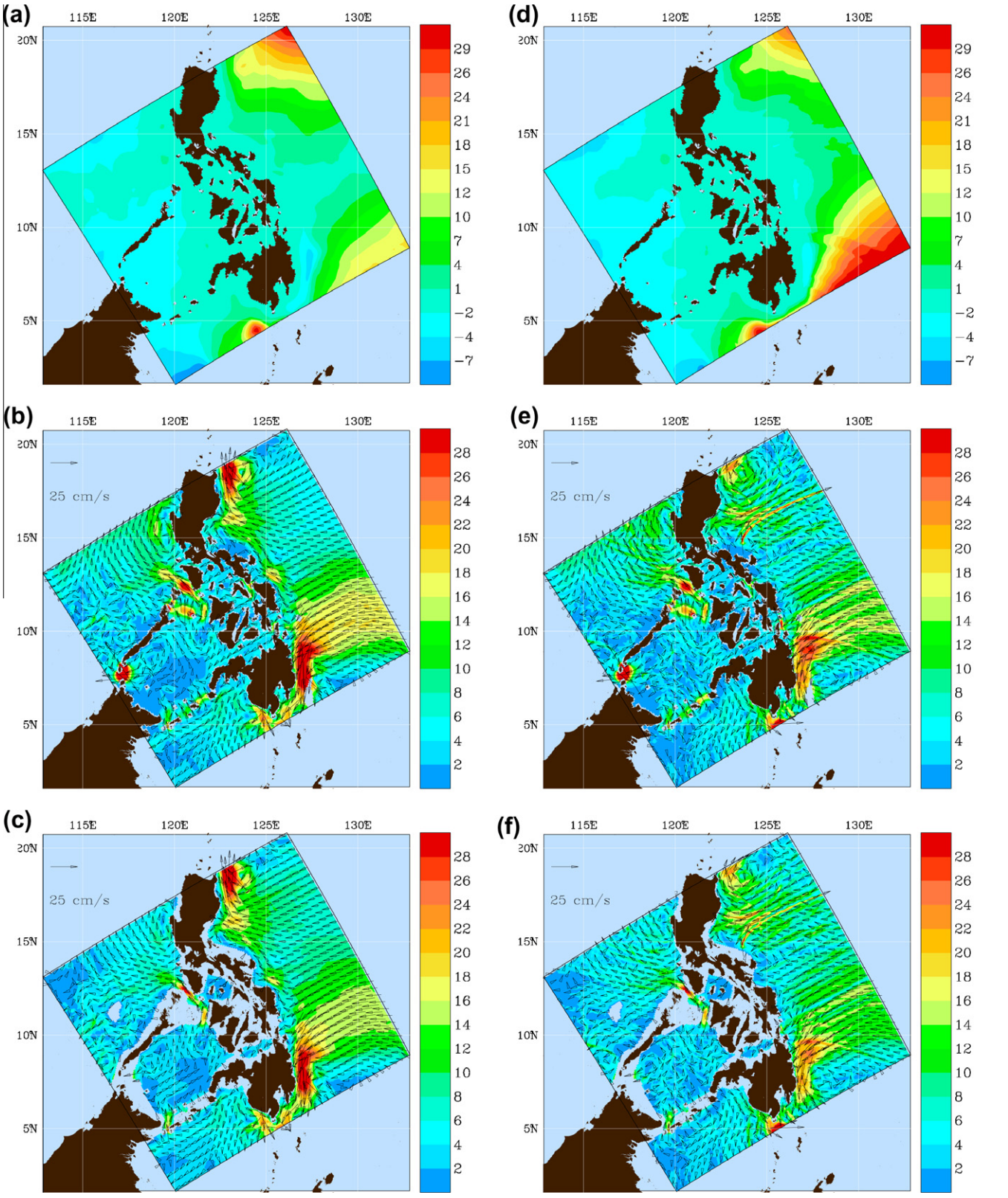

Fig. 10. Velocity estimation under geostrophic balance and optimized inter-island transports (weight functions based on minimum vertical areas among islands) from hydrographic field maps (WOA05) obtained using the FMM (left) and using the SPDE Approach (right): (a and b) Streamfunction, velocity at depths: (c and d) 0 m; (e and f) 100 m.

strophic shear balance. Weight functions based on the minimum vertical area across each pair of islands are computed and used in the algorithm. As described in Section 303a, the minimum vertical areas were estimated using the FMM by specifying the scalar speed function in the Eikonal equation (Eq. (11)) as $F(x,y) = 1/H(x,y)$, where H is the ocean depth. The temperature and salinity

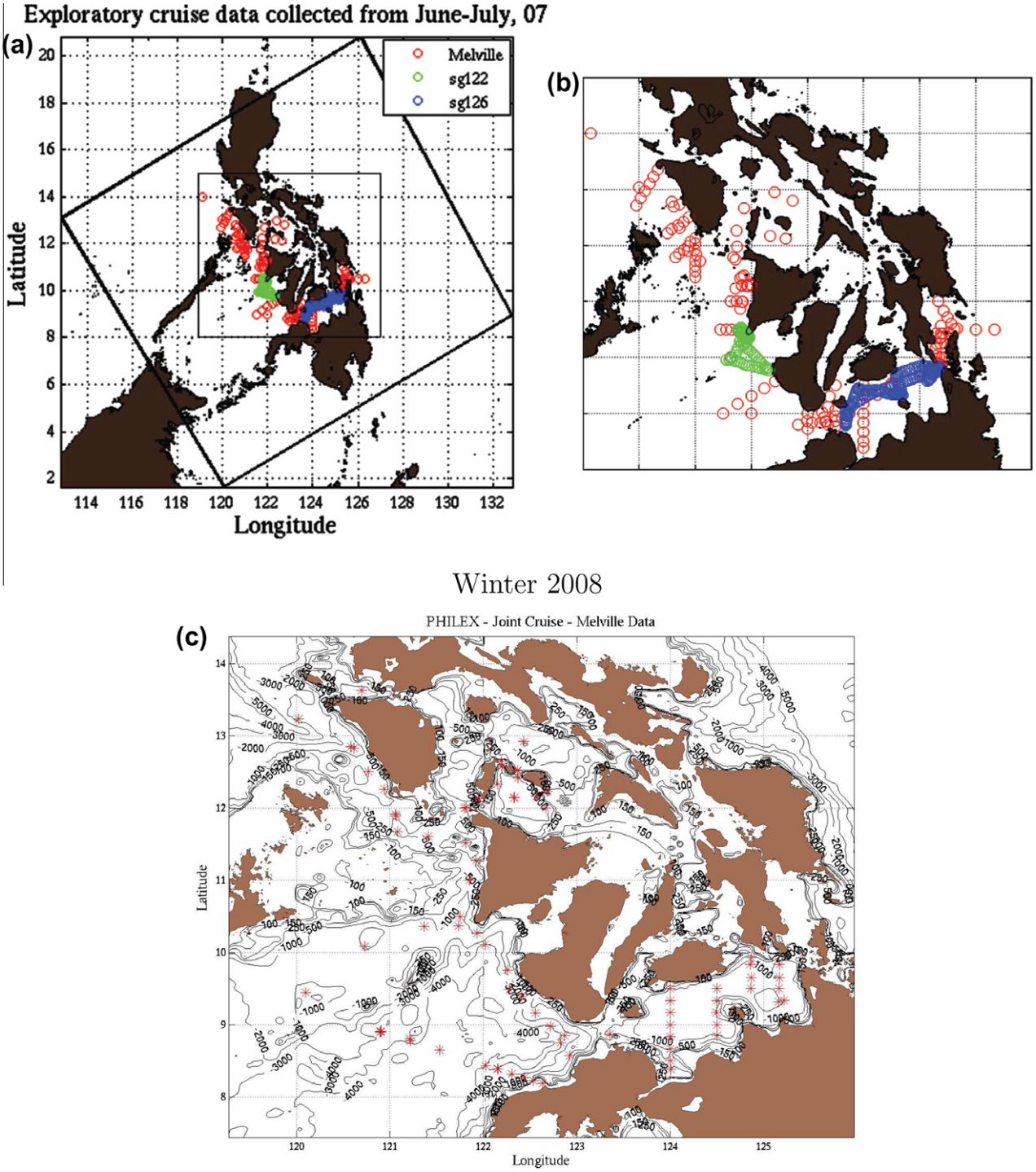

Fig. 11. Locations of: (a,b) Melville exploratory cruise and glider data (Summer 2007) and (c) Melville joint cruise Data (Winter 2008), both in the Philippines Archipelago. For the Summer 2007 OAs, we employ: R/V Melville, 142 CTD and 133 Chlorophyll casts; SG122, 110 CTD casts; and, SG126, 191 CTD casts. For the Winter 2008 OAs, we employ 86 CTD casts from the R/V Melville.

maps are those of our FMM-based OA scheme (Figs. 3–8(c)) and of the SPDE approach (Figs. 3–8(d)), with the Helmholtz equation employed for the field. The streamfunction and velocity fields (at depths 0, 100 m) are shown in Fig. 10. The estimates based on our FMM-based hydrographic OAs (Fig. 10 (left)) are in overall good agreement with those obtained using maps based on the stochastically forced Helmholtz equation (Fig. 10 (right)). However, the SPDE-based velocity fields are noisier, reflecting the spurious noise in the hydrographic OAs. On average, these monthly mean flow estimates suggest larger density-driven velocities in the Mindoro Strait, in the Mindanao current south of Mindanao Island and in the Balabac Strait. The maximum absolute velocity reaches 80 cm/s in the Balabac strait at the surface. At lower depths, velocities remain high in the Mindoro Strait and near Mindanao.

Weight functions based on the minimum inter-island distance, which can be obtained using the FMM by specifying the scalar speed function in the Eikonal equation (Eq. (11)) as 1 for sea points and 0 for land points, were also used. The velocity fields obtained

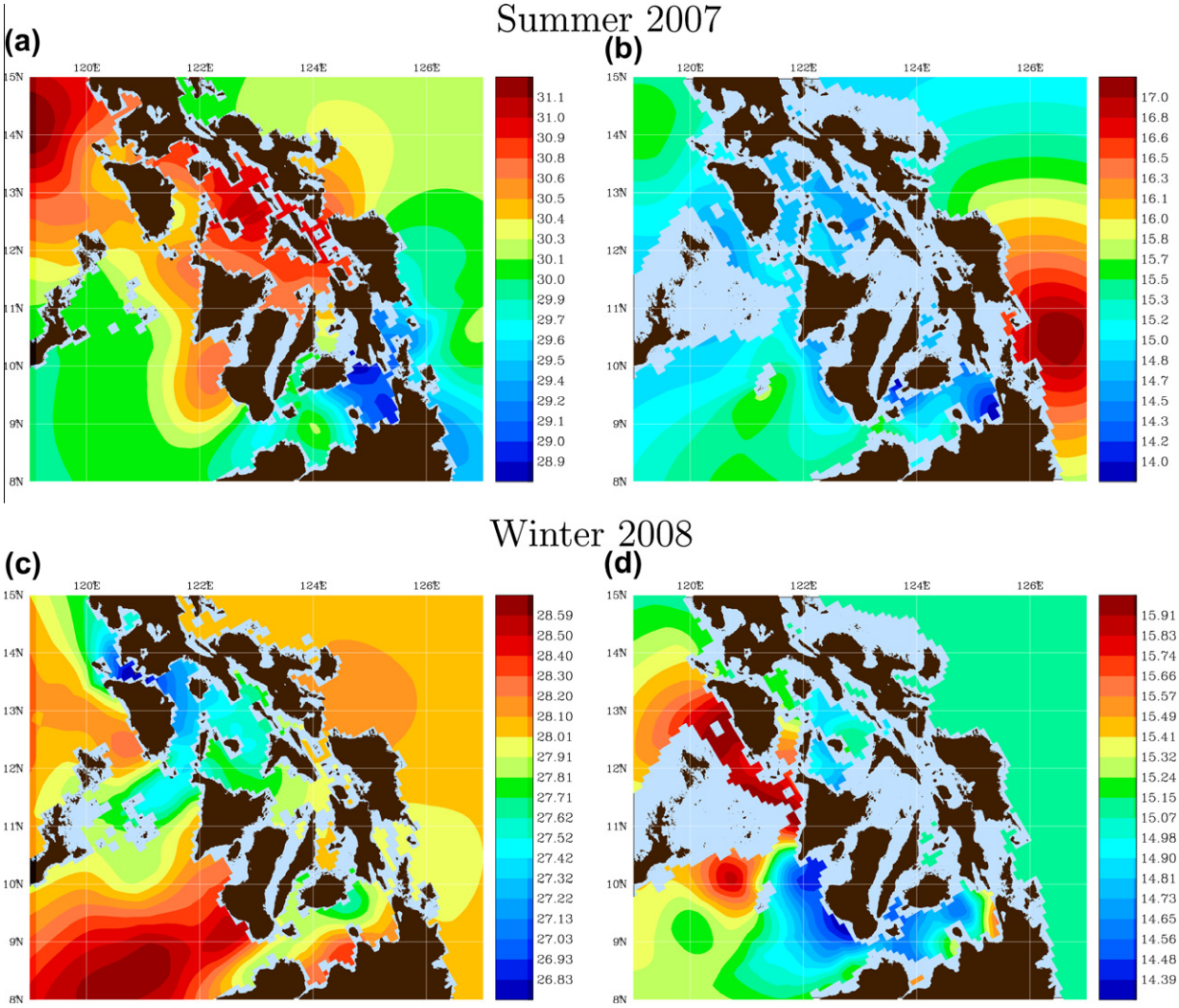

Fig. 12. Temperature (°C) OA Fields at 0 m (Left) and 200 m (Right) using the: (a,b) Melville exploratory cruise and glider data (Summer 2007); (c,d) Melville joint cruise data (Winter 2008). Colorbars are not the same for the two periods due to the winter and summer variability.

using these weight functions had much larger magnitudes, particularly in the Balabac Strait (Agarwal, 2009), where the maximum absolute velocity was 141 cm/s. Such high velocity magnitudes are very unlikely. These results show that weight functions based on the minimum vertical area (logical for transport estimates) are adequate.

6.2.2. Objective analysis of synoptic data for the Summer 2007

The data used in this example are from the Melville Exploratory cruise and sg122 and sg126 gliders for June-July'07 (A. Gordon and C. Lee, personal communications). The data coverage is shown in Fig. 11(a) and (b). A portion of the Archipelago is sampled and OA maps are computed in that region. The scales used were fit to: large length scales $(L_0 = 1080, L_e = 360)_{\mathrm{LS}}$ and most energetic length scales $(L_0 = 270, L_e = 90)_{\mathrm{ME}}$. The observational error is set to $\sigma_d^2 = 0.20$. These scale and error values were estimated based on data, see Agarwal (2009). The hydrographic field maps obtained using our FMM-based OA scheme are shown in Fig. 12(a) and (b) and Fig. 13(a) and (b), respectively at depths of 0 and 200 m. Once again, these maps clearly indicate that the coastline constraints are appropriately satisfied. At 0 m, the warmer regions to the west of Luzon island remain uncorrelated with the Pacific waters east of Luzon. The warm Sibuyan and Visayan Seas can be distinguished from the relatively cold Bohol Sea. At 450 and 1000 m (not shown), the data in the warm Sulu Sea and Bohol Sea do not impact the other regions; there is no direct relationship across landforms. Similar observations are made for the salinity (e.g. at 0 m, the low salinities west of Luzon island do not affect Pacific waters east of Luzon). We note that very few data are collected in the Pacific proper, so patterns there also reflect the sampling paths.

6.2.3. Objective analysis for early Winter 2008

The data used in this example is obtained from the joint Melville cruise for the November 2007 – January 2008 period. Data locations are shown in Fig. 11 (bottom). The OA parameters are as those of Summer 2007 (Section 6.2.2). The hydrographic field maps obtained using the FMM-based scheme are shown in Fig. 12(c) and (d) and Fig. 13(c) and (d), respectively, at 0 and 200 m depth. At the surface, the warm/fresher region west of Luzon is uncorrelated with the region east of Luzon. At 450 and 1000 m (not shown), the warm Bohol Sea is enclosed and at these depths, it does not affect other regions either.

Comparing Winter 2008 with Summer 2007, the largest differences in temperature and salinity are near the ocean surface (deeper than 200 m, fields are much closer). For example, the 0 m temperatures in the South China Sea and Pacific are significantly

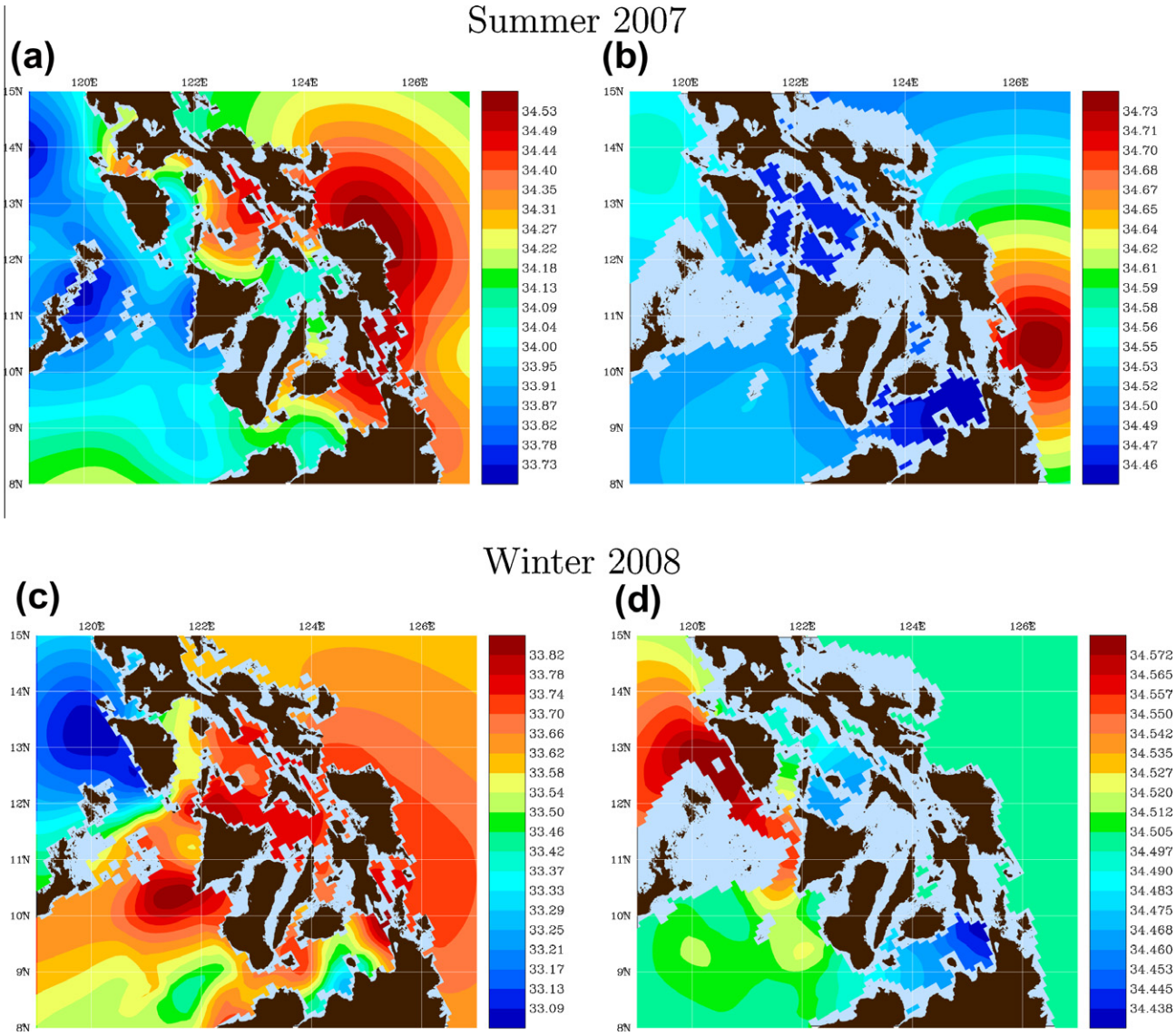

Fig. 13. Salinity (PSU) OA Fields at 0 m (left) and 200 m (right) using the: (a and b) Melville exploratory cruise and glider data (Summer 2007); (c and d) Melville joint cruise data (Winter 2008). Colorbars are not the same for the two periods due to the winter and summer variability.

lower in Winter 2008 than in Summer 2007. However, in the Sulu Sea, temperatures of the two seasons are much closer.

6.2.4. Objective analysis for biological fields (chlorophyll)

Of course, our new FMM-based scheme is not limited to physical fields. Its application to biology is illustrated here using the Exploratory cruise Summer 2007 data (Gordon, 2009). Our biological OAs (for chlorophyll, nitrate and ammonium) were utilized to initialize physics-biology modeling studies (Burton, 2009; Lermusiaux et al., 2011), combining in situ data with satellite images and a region-by-region biological feature model. Here, only the mapping of R/V Melville chlorophyll profiles is shown. The biological OA parameters are, once again, as in Section 6.2.2. The resulting chlorophyll maps are illustrated in Fig. 14 at depths of 0, 10, 50 and 150 m.

The concentration of biological fields such as chlorophyll, phytoplankton and zooplankton is substantial near the surface due to sunlight. The chlorophyll concentration is maximum near islands, often driven by winds or bathymetric upwelling. Away from islands, it tends to be more uniform, around a mean value. At 0 and 10 m, the maximum chlorophyll concentration is observed south of the Visayan Sea and in the Bohol Sea. At 50 m, chlorophyll concentrations remain significant there, but the largest chlorophyll concentrations are observed north of Palawan island. Concentrations at 150 m are largest in the Pacific (the only Pacific data is at the mouth of Surigao Strait, hence higher values are estimated there only). These results agree with the expected depths of subsurface Chl-maxima in the different seas: Sulu Sea (50–90 m), Visayan Sea (10–40 m) and Pacific (100–180 m), see (Burton, 2009). Biological concentrations at lower depths decrease rapidly.

7. Computational analysis

The computational properties of our methods for mapping irregular data in complex geometries are now described and studied. First, the computational costs are compared in Section 7.1. Then, schemes to resolve issues specific to complex multiply-connected coastal regions such as the need for accurate distance estimates (Section 7.2) and the need for positive-definite covariance matrices (Section 7.3) are discussed. These schemes are important because if the covariance matrix becomes negative, divergence problems occur in the Kalman updates (Brown and Hwang, 1997).

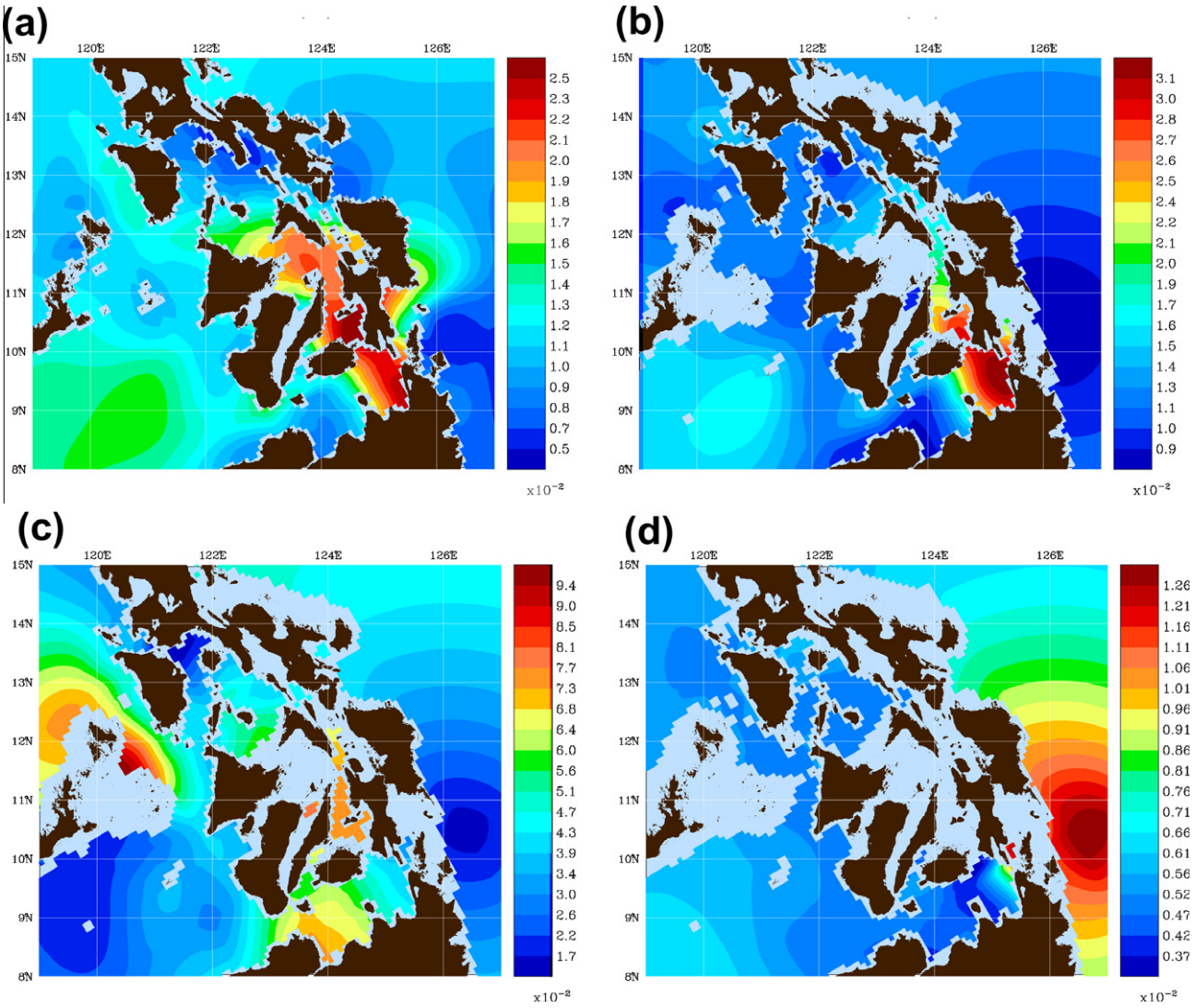

Fig. 14. Chlorophyll (μmol/kg) OA Fields for the Melville exploratory cruise data (Summer 2007), computed using the FMM at: (a) 0 m, (b) 10 m, (c) 50 m, (d) 150 m.

To motivate the computational studies, recall that we generate the covariance matrices using analytical correlation functions defined based on the Euclidean distance. These correlation functions are termed "positive definite" if they generate positive definite covariance matrix in a simply-connected convex domain. It has been well established using the Wiener-Khinchin and Bochner's theorems that if the Fourier transform (or the spectral density) of a correlation function is non-negative for all frequencies then the correlation function is positive definite (Yaglom, 1987; Papoulis, 1991; Yaglom, 2004; Dolloff et al., 2006). However, we found that for coastal regions, covariance matrices generated from "positive definite correlation functions" may not be positive definite due to: (a) numerical errors in the computation of the shortest path lengths, or (b) the presence of landforms which lead to multiply-connected or non-convex domains, invalidating assumptions in the Wiener-Khinchin and Bochner theorems (see Agarwal, 2009 for proof). This can lead to divergence problems in the mapping.

Such problems are illustrated using the WOA05 data (Spliced February and Winter Climatology) shown in Fig. 15 (a). To simplify, we consider single-scale OAs (all previous examples were two-scale OAs). The field maps obtained using our FMM-based scheme with length scales ($L_0 = 540, L_e = 180$) and length scales ($L_0 = 1080$, $L_e = 360$) are shown in Fig. 15. Fields obtained using the larger scales (Fig. 15(c)) clearly show divergence problems near Palawan. These problems are not encountered when the smaller length scales are used and are much smaller when a higher-order FMM scheme is used (Fig. 15(b)). Questions which motivate our studies next are thus (as introduced at the end of Section 2): (i) What are the computational errors in the shortest sea-path distances computed using the FMM/LSM and how can they be reduced?, and (ii) What are the computational issues, including non-positive definite covariances, that arise in a multiply-connected coastal domain and how can they be remedied? A higher-order FMM than the first-order one (Section 4.2) is discussed in Section 7.2. Higher-order FMMs significantly reduce errors in the distance estimates, i.e. the difference between the numerically computed and true distance, which limits divergence problems in the mapping. However, even if exact distances are used, when curved boundaries or islands are present in the domain, negative covariances can still occur. Methods to solve these issues are derived in Section 7.3.

7.1. Comparison of computational costs

For all OA schemes, we sequentially process observations (see Parrish and Cohn (1985), Cho et al. (1996), Lermusiaux (1997),

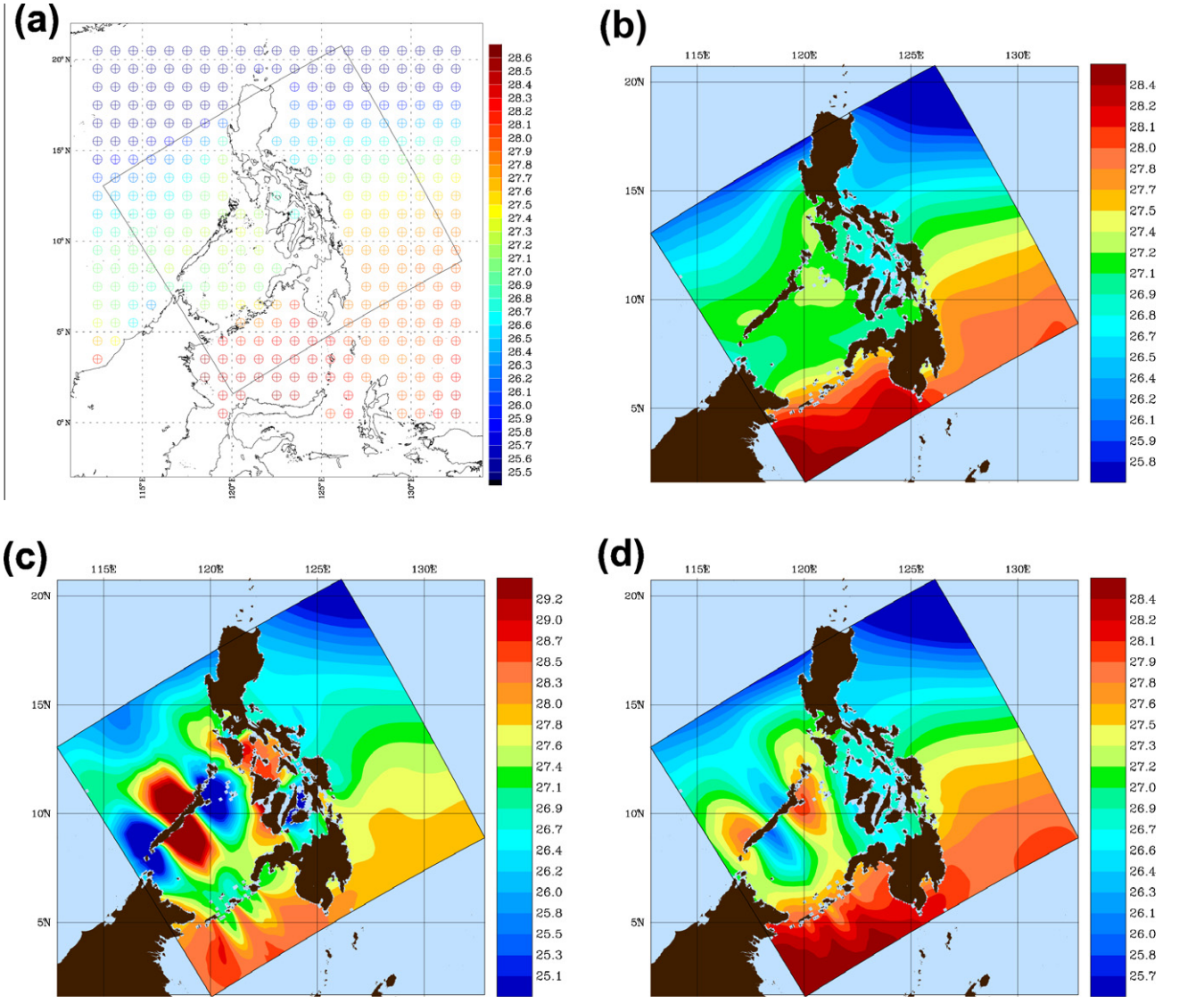

Fig. 15. (a) World Ocean Atlas 2005 (Spliced February and Winter Climatology) in situ temperature (°C) at 0.0 m; Temperature (°C) OA Fields using the FMM at the surface (0 m) with the following scheme and scales, (b) first order FMM and L_0 = 540 km, L_e = 180 km, (c) first order FMM and L_0 = 1080 km, L_e = 360 km, (d) higher order FMM and L_0 = 1080 km, L_e = 360 km.

Agarwal (2009)). Such sequential processing drastically reduces computational costs and also allows estimating the impact of individual data. Since data are processed sequentially, the costs for the OA schemes are compared by considering a single scalar data point. The main cost is then the computation of the covariance of that data point with all other grid points in the domain. For the FMM, LSM and Dijkstra's schemes, the operation count to do this is driven by the computation of the shortest distances from that data point with all other points. For the SPDE scheme, it depends on the diffusion equation used and on the iterations to reach state-state. For a 2-D domain with N points in each direction, these operation counts are given in Table 2.

There are a total of N^2 grid points at each level and the operation count for LSM is obtained from an optimistic guess that LSM will take roughly N steps to converge. In reality, the iterations can take much longer to converge, and the LSM is thus not efficient to compute these distances. On the other hand, FMM is an efficient technique which requires a fast method to locate the smallest value grid point in the *narrow band*. The MinHeap data structure with backpointers (Sedgewick and Wayne, 2011) is employed here to efficiently locate the grid point with the minimum value. The total work done in the DownHeap and UpHeap operations, which ensure that the updated quantities do not violate the heap properties, is $O(\log N)$. Thus, for a 2D domain with N grid points in each direction, the FMM has an operation count of $N^2 \log N$, which is a significant improvement over the LSM. An efficient SPDE scheme requires at least an order of $N^2 n$ where n is the number of iterations to reach steady state. We have observed that the SPDE approach is at least 15% more expensive computationally than the FMM scheme. Thus, the FMM-based scheme is computationally the most efficient.

Table 2
Operation counts for computing the covariances among one data point and each of the N^2 model grid points, as obtained using the LSM, FMM, SPDE (n iterations) and Dijkstra's schemes.

Method	Operation count
Level Set Method	$O(N^3)$
Fast Marching Method	$O(N^2 \log N)$
SPDE Method	$O(N^2 n)$
Dijkstra's Method	$O(N^3)$

7.2. Higher order Fast Marching Method

In a domain with no islands or landforms, the shortest path length obtained using the FMM/LSM should be equal to the Euclidean distance. But the FMM/LSM have discretization errors which lead to inaccurate length estimates. The Weiner Khinchin and Bochner theorems are valid for covariances computed using the Euclidean distance in a simply-connected convex domain. So, if the domain is simply-connected convex, the covariance matrix can only become negative definite due to the inaccurate length estimates. This may lead to divergence problems in the resultant field maps. Here, the goal is to estimate and reduce the computational errors in the shortest path lengths. To do so, we introduce the higher order FMM.

The FMM scheme presented in Section 4.2 is first order, since the first order discretization form (Eq. (14)) of the Eikonal equation (Eq. (11)) is used. A different implementation of FMM with higher accuracy (Sethian, 1999a; Sethian, 1999b) is discussed here. It employs the second order backward approximation to the first derivative T_x is given by:

$$T_x \approx \frac{3T_i - 4T_{i-1} + T_{i-2}}{2\Delta x} \Longleftrightarrow T_x \approx D^{-x}T + \frac{\Delta x}{2} D^{-x-x}T, \tag{18}$$

and the second order forward approximation to the first derivative T_x given by:

$$T_x \approx \frac{3T_i - 4T_{i+1} + T_{i+2}}{2\Delta x} \Longleftrightarrow T_x \approx D^{+x}T - \frac{\Delta x}{2} D^{+x+x}T. \tag{19}$$

Here D^{-x} and D^{+x} are the first order forward and backward approximations for the first derivative, respectively (Eq. (10)), $D^{-x-x} \equiv D^{-x}D^{-x}$ and $D^{+x+x} \equiv D^{+x}D^{+x}$.

Consider the switch functions defined by:

$$\begin{aligned} switch_{ij}^{-x} &= \begin{pmatrix} 1 & \text{if } T_{i-2,j} \text{ and } T_{i-1,j} \text{ are known('Alive')} \\ & \text{and } T_{i-2,j} \leqslant T_{i-1,j} \\ 0 & \text{otherwise} \end{pmatrix}, \\ switch_{ij}^{+x} &= \begin{pmatrix} 1 & \text{if } T_{i+2,j} \text{ and } T_{i+1,j} \text{ are known('Alive')} \\ & \text{and } T_{i+2,j} \leqslant T_{i+1,j} \\ 0 & \text{otherwise} \end{pmatrix}. \end{aligned} \tag{20}$$

Similar functions are defined in the y-direction. The higher accuracy scheme attempts to use a second order approximation for the derivative whenever the points are tagged as 'alive' (the points inside the band where the value of the arrival time function is frozen: see Section 4.2) but reverts to the first order scheme otherwise.

The modified discretization equation for the higher accuracy FMM is thus given by:

$$\begin{pmatrix} \max\left(\left[D_{ij}^{-x}T + switch_{ij}^{-x}\frac{\Delta x}{2}D_{ij}^{-x-x}T\right], \\ -\left[D_{ij}^{+x}T - switch_{ij}^{+x}\frac{\Delta x}{2}D_{ij}^{+x+x}T\right], 0\right)^2 \\ + \\ \max\left(\left[D_{ij}^{-y}T + switch_{ij}^{-y}\frac{\Delta y}{2}D_{ij}^{-y-y}T\right], \\ -\left[D_{ij}^{+y}T - switch_{ij}^{+y}\frac{\Delta y}{2}D_{ij}^{+y+y}T\right], 0\right)^2 \end{pmatrix} = \frac{1}{F_{ij}^2}. \tag{21}$$

We note that the above scheme is not necessarily of second order. Its accuracy depends on how often the switches evaluate to zero and how the number of points where the first order method is applied changes as the mesh is refined. When the number of points where the first order method is applied is relatively small (occurs only near the coastlines), the error is reduced considerably by using the second order FMM (Agarwal, 2009). Of course, third or higher-order approximations for the derivative T_x can be used to construct even more accurate FMM schemes, but this increases the computational cost. We found that the relative error in the distances computed by the FMM is higher near the data point and it decays as the distance increases. To keep the computational cost low and a uniform relative error, we can thus use higher accuracy FMM near the data points and then progressively shift to lower order schemes as the distance increases.

The results of using higher order FMMs to minimize errors are illustrated on Fig. 15(d). They clearly show that the above higher order FMM has attenuated the divergence issues compared to the first order FMM. The divergence issues do not vanish completely because some discretization errors still occur but also because of the presence of landforms. To deal with the latter and the effects of the multiply-connected coastal domains, we further improve schemes next in Section 7.3.

7.3. Positive Definite covariance matrix for complex multiply-connected coastal regions

Apart from the inaccurate shortest path length, the covariance matrix may also become negative due to the presence of islands and coastlines. This is because the presence of islands and archipelagos stretches the direct Euclidean path, which can render the covariance matrix negative.

For example, consider the idealized multiply-connected domain with an island, shown on Fig. 16. This domain has 12 grid points marked as ocean points and 4 grid points marked as land points. The length of the shortest sea path is computed exactly to form the covariance matrix and so remove all discretization errors of the FMM/LSM. To do so, the positive-definite correlation function $Cor(r) = exp\left[-\frac{r^2}{2L^2}\right]$ with L = 2 is used. We find that the covariance matrix is not positive definite. The maximum eigenvalue for the covariance matrix is 6.3345 while the minimum is -0.0504. This idealized example clearly reveals that classic Euclidean-based covariance matrices for a complex multiply-connected region may not necessarily be positive definite. This is because the conditions of the Wiener-Khinchin and Bochner's theorems are not satisfied.

One could consider changing the coordinate system, for example curvilinear coordinates. For adequate coordinate choices, in the transformed space, the domain can then be simply connected and convex. However, the issue then is that the real distances among grid points become position dependent which violates another assumption of the Wiener-Khinchin and Bochner's theorems, see (Agarwal, 2009) for examples and more discussions.

Hence, other schemes have to be used to alleviate the divergence problems (Fig. 17(a)) due to the non-positive definite covariance matrix. They include:

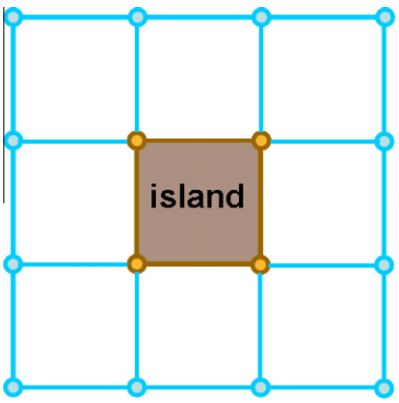

Fig. 16. Example of an idealized (multiply-connected) domain having an island.

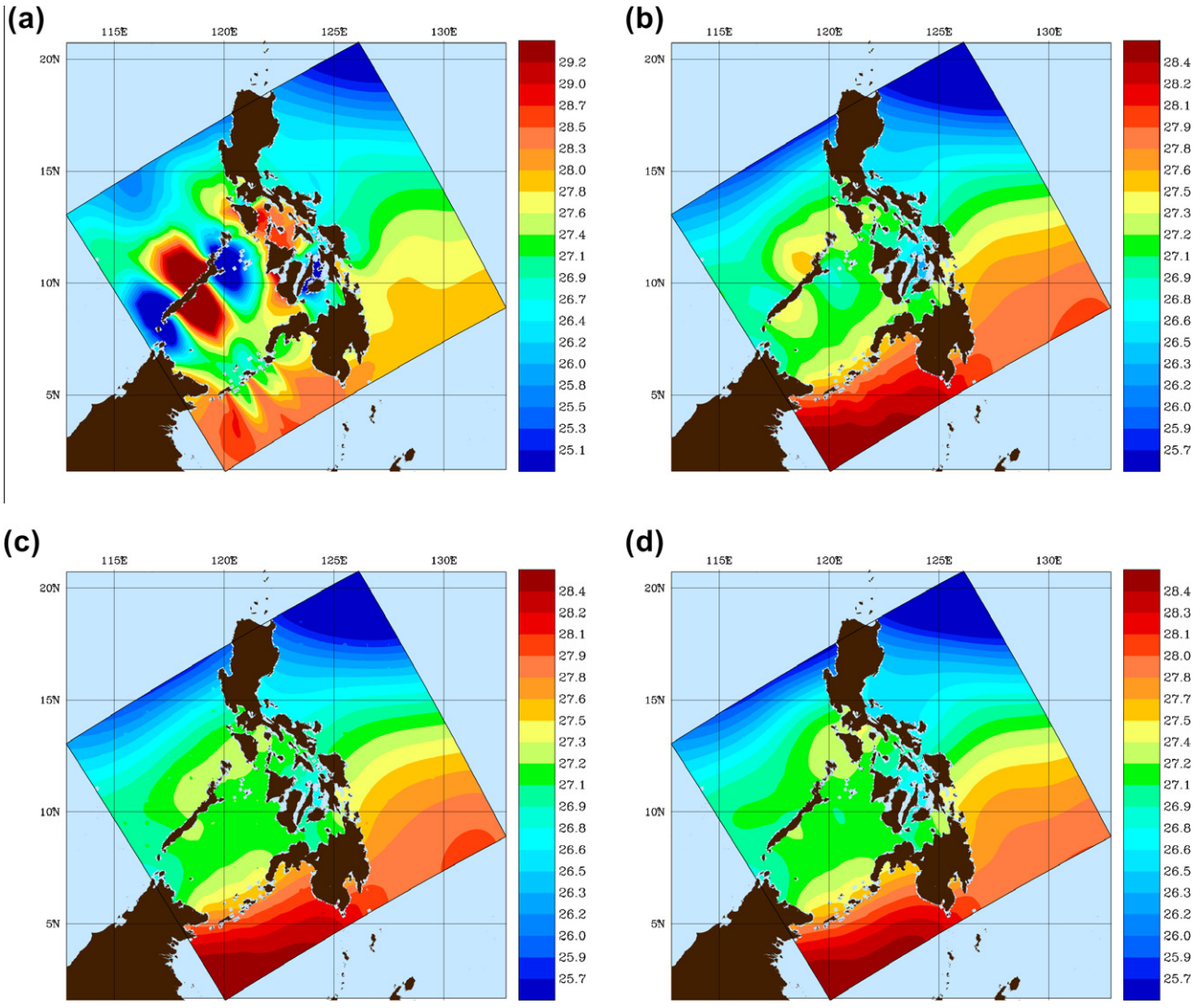

Fig. 17. Temperature (°C) OA Fields at the surface (0 m) (scales $L_0 = 1080$ km, $L_e = 360$ km) using the: (a) FMM, (b) FMM and removal of problematic data, (c) FMM and introducing process noise, (d) FMM and applying dominant singular value decomposition (SVD) of a priori covariance.

(**a**) **Discarding the problematic data:** Discarding the data that lead to negative values of $\mathbf{H_j Cor}(\mathbf{x},\mathbf{x})_{\mathbf{j-1}}\mathbf{H_j^T}$ would solve the issue and eliminate divergences in the resultant OA. However, this method is a not adequate since the information in the data is discarded entirely. The field map obtained by discarding the problematic data is shown in Fig. 17 (b). Clearly, the divergence problems are removed but loosing data is not acceptable.

(**b**) **Introducing process noise:** Adding a small process noise to the diagonal elements of the covariance matrix would help (Brown and Hwang, 1997), but it will lead to a degree of sub-optimality: the noise affects all of the problematic data. However, it is often a more acceptable scheme than discarding the data. We indeed find that introducing the process noise leads to less divergence problems, as shown in our example, see Fig. 17(c).

(**c**) **Dominant Singular Value Decomposition (SVD) of a priori covariance:** To construct the OA field maps, the full covariance matrix is not required. In fact, the full covariance matrix ($Cor(\mathbf{x},\mathbf{x})$) is expensive to compute and store, and it is therefore rarely computed. The necessary requirement for field maps is the covariance matrix among the grid and data points, i.e. $Cor(\mathbf{x},\mathbf{X})$. The divergence problems can be removed by first obtaining the singular value decomposition (SVD) of $Cor(\mathbf{x},\mathbf{X})$ and then retaining only the dominant singular values and setting the smaller singular values (e.g. less than 1% of the maximum singular value) to zero. This SVD procedure renders the covariance matrix non-negative definite, which was verified in multiple examples where a simulated map was used for a true ocean. Based on these results and on minimum error variance arguments, the dominant SVD method is the most acceptable one because it loses the least information contained in the data. Our example is shown on Fig. 17(d). We find that the field maps obtained using this dominant SVD of the a priori covariance is free from divergence problems. They are also similar to, but further improve, the fields obtained by introducing the process noise.

8. Summary and conclusions

New methodologies for the efficient mapping and dynamical inference of ocean fields from irregular data in complex multiply-connected domains were derived and utilized, and computational properties of these mapping schemes were studied. These new OA methods, which satisfy the coastline and bathymetry

constraints (e.g. there is no direct relationship across landforms), are based on estimating the length of the optimal sea path using either the Level Set Method (LSM) or the Fast Marching Method (FMM). The optimal sea path was geometrically defined: for purely horizontal OAs, it is the shortest sea distance in 2D, and for 3D OAs, it is the shortest sea distance in 3D, weighting the vertical or diapycnal distances more than the horizontal ones. Numerical schemes were derived and implemented, and their operation counts compared. Their properties and results were studied in complex domains, the Philippines Archipelago and Dabob Bay, in realistic situations. Both climatological and synoptic datasets were employed and estimates of temperature, salinity and chlorophyll fields were computed and discussed. We found that without these new OA methods, neither meaningful dynamical studies nor meaningful ocean simulations could be initiated.

Results were compared with those of a standard OA scheme (using across-landforms Euclidean distance in the analytical correlation function), of OA schemes based on other distance estimation methods and of OA schemes based on the use of stochastically forced PDEs (SPDEs). We showed that the FMM-based scheme is computationally cheaper than the LSM-based scheme and diffusion-based SPDE approach. We found that the field maps obtained using our FMM-based schemes were more robust than those obtained using SPDE schemes: fields did not require postprocessing (smoothing), i.e. they were devoid of any spurious gradients. Such spurious gradients in hydrographic maps lead to unrealistic geostrophic flows. The FMM and LSM were the most appropriate for estimating the optimal sea distances among other distance estimation schemes such as Dijkstra's optimization algorithm and the classic Bresenham-based line algorithm. The optimal distance computed using Dijkstra's algorithm is computationally expensive and inaccurate. Apart from being computationally expensive, the optimal distance computed using the Bresenham line algorithm is discontinuous. This results in the formation of numerical fronts with high field gradients. Such erroneous gradients do not occur when our FMM-based scheme is utilized.

Mathematical and computational properties of the new OA schemes were studied. The sequential processing of observations reduces the computational cost and also helps in understanding the impact of individual data. We found that the use of higher order FMMs increased the accuracy of the estimates of the length of shortest sea paths. The most efficient FMM schemes derived employed a variable order discretization, the order decaying as the distance between the data and model points increases. Accurate FMM distance estimates eliminate one of the sources of negative covariance matrices. The other source is the presence of islands or of other non-convex landforms. This is because the Wiener-Khinchin and Bochner theorems are valid only for correlation functions based on the Euclidean distance in convex simply-connected domains. Several approaches to overcome this issue were discussed. These include discarding problematic data points, introducing process noise, and reducing the covariance matrix by applying the dominant singular value decomposition (SVD). Among these, we showed that the latter use of the SVD to reduce the covariance matrix is the best solution.

We have also employed a FMM-based method to estimate the total velocity under geostrophic balance in complex multiply-connected domains. The FMM is used to compute the minimum vertical area between all pairs of islands. Such areas are neeeded to compute the transport streamfunction field that optimizes the inter-island transports and produces a smooth velocity field. The result is a mass-conserving geostrophic flow in balance with the hydrographic OA maps and with optimized inter-island transports. This method and the minimum vertical area estimates were necessary to obtain realistic velocity estimates in our Philippine Archipelago examples.

As part of our ongoing work, we have started to incorporate additional geometrical and non-homogeneous dynamical effects to our FMM-based OA scheme. An approach we have followed is to modify the scalar speed function in the Eikonal equation as a function of these geometrical properties and heterogeneous dynamics. In particular, we have used a bathymetry-dependent speed function to include depth effects in the correlation scales. To include heterogeneous scales due to the existence of fronts, we can first create an expected length scale field that is a function of space and direction, possibly using raw data only (Agarwal, 2009) or a feature model (Gangopadhyay and Robinson, 2002). We can then compute the optimal sea path as before, but select for correlation scale the smallest one found along that path. For example, if the optimal path crosses a front, the length scale in the across direction would then be the minimum cross-frontal scale. Analogous modification of the scalar speed function or the length scale can be used to incorporate other dynamical effects (e.g. conservation of potential vorticity). In the future, the ideas of optimal path length and our FMM/LSM-based scheme can be used to extend to complex coastal regions our 3D multivariate and multi-scale mapping of fields and of their dominant errors (Lermusiaux, 2002). Such schemes would be needed for ensemble initializations.

We expect a wide range of applications for our FMM-based OA schemes. Already when mapping relatively simple coastal domains, all constraints of landforms are accounted for. Constraints due to bathymetric features are also respected, even in deep ocean regions, from the simpler basins, plateaus and troughs to the more complex sills, ridges, seamounts and trenches: if the surface on which the OA is computed intersects bathymetry, our sea paths adequately go around it. These surfaces are general; they can be horizontal, terrain-following or density-based. Effects of distances perpendicular to these surfaces can be included in the speed function, to include 3D contributions in the correlation scales. Initial gridded conditions computed by the present FMM methods have thus enabled our simulations in varied regions, including the Taiwan region, New England shelf, Dabob Bay and Monterey Bay (Xu et al., 2008; Lermusiaux et al., 2010; Haley and Lermusiaux, 2010). Our new methods would also improve the widely-used gridded databases such as the World Ocean Atlas (WOA) since such maps were computed without explicitly accounting for coastline and bathymetry constraints.

Acknowledgments

We are very thankful to P.J. Haley and W.G. Leslie for very helpful inputs and multiple discussions, and to O.G. Logutov for sharing his SPDE mapping method. We are grateful to the whole PhilEx and PLUS teams for their fruitful collaborations. In particular, we thank crews, operators and support personnel of the Melville R/V (Prof. A.L. Gordon), gliders (Dr. Craig Lee, Dr. Bruce Howe, Dr. Marc Stewart) and kayaks (MIT LAMSS group and Prof. H. Schmidt) for their work and the critical data they provided. Finally, we thank the reviewers for their useful suggestions.

Appendix A. Objective analysis schemes of the 'Levitus Climatology'

The OA schemes used to map the 'Levitus Climatology' (Levitus, 1982; Locarnini et al., 2006; Antonov et al., 2006; Garcia et al., 2006a,b) originate from Cressman (1959) and Barnes (1964). The approach is based on adding "corrections", which are computed as a distance-weighted mean of all data point difference values, to the first-guess field. Initially, to reduce the computational time, the World Ocean Atlas 1994 (WOA94) used the Barnes (1973)

scheme which requires only a single "correction" to the first-guess field at each grid point in comparison to the successive correction method of Cressman (1959) and Barnes (1964). The most recent WOA98, WOA01 and WOA05 maps were completed employing a three-pass "correction" scheme, using the multi-pass analysis of Barnes (1994). The inputs to this global analysis scheme are differences among a first-guess field and the one-degree square means of the observed data values. An influence radius is then specified and a correction to the first-guess value at all grid points is computed as a distance-weighted mean of only the difference values that correspond to data points that lie within the area defined by the influence radius. Mathematically, the correction factor derived by Barnes (1964) is given by:

$$C_{i,j} = \frac{\sum_{s=1}^{d} W_s Q_s}{\sum_{s=1}^{d} W_s}, \tag{A.1}$$

where,

(i,j) – are coordinates of grid points;
$C_{i,j}$ – correction factor at the grid point coordinates (i,j);
d – the number of data points that fall within the area around point (i,j) defined by the influence radius;
Q_s – difference between the observed mean and the first-guess at the sth data point in the influence area;
$W_s = \exp(-Er^2/R^2)$ (for $r \leqslant R$; $W_s = 0$ for $r > R$) – the correlation weight;
r – distance between data and grid points;
R – influence radius; and,
$E = 4$.

At each grid point, the final analyzed gridded value $G_{i,j}$ is the sum of the first guess $F_{i,j}$ and the correction $C_{i,j}$. The expression is:

$$G_{i,j} = F_{i,j} + C_{i,j}. \tag{A.2}$$

If there is no data within the area defined by the influence radius, the correction is zero and the analyzed value is the first-guess. The analysis scheme is set up such that the inference radius can be varied at each iteration. To progressively analyze the smaller scale phenomena with each iteration, the analysis begins with a large inference radius which is decreased gradually with each iteration. Eq. (A.2) can also be expressed in a matrix–vector form:

$$\mathbf{G} = \mathbf{F} + [diag(\mathbf{W}\mathbf{e}_d)]^{-1}\mathbf{W}\mathbf{Q}, \tag{A.3}$$

where, if n and d denote the number of model-grid and data points, respectively, the analyzed field $\mathbf{G}$ and the first guess $\mathbf{F}$ are n-by-1, the correlation weight matrix $\mathbf{W}$ is n-by-d, the difference $\mathbf{Q}$ between the observed mean and first-guess at data points is d-by-1, and $\mathbf{e}_d$ is d-by-1 with unit entities. The operation diag ($\mathbf{v}$) creates a diagonal matrix i.e. it puts the vector $\mathbf{v}$ on the main diagonal.

In analogy to the Kalman Gain ($\mathbf{K}$) from the Gauss Markov criterion ($\mathbf{K} = Cor(\mathbf{x},\mathbf{X})[Cor(\mathbf{X},\mathbf{X}) + \mathbf{R}]^{-1}$), Eqs. (A.3) and (1) show that a similar Gain matrix ($K_L = [diag(\mathbf{W}\mathbf{e}_d)]^{-1}\mathbf{W}$) can be defined for the Levitus methodology. While the multi-scale OA approach in MSEAS is based on Gauss Markov estimation theory and successive scale-by-scale updates, the Levitus OA is based on computing the distance-weighted mean of all differences between the most recent first-guess field and the data mean within the inference radius and then repeat with a reduced inference radius. The main difference is that Gauss Markov estimation theory requires and uses prior error covariances for the data and the first-guess, while the Levitus OA requires radius of influence estimates and uses data averaging.

Appendix B. Fast Marching algorithm

The fast marching algorithm (Sethian, 1996; Sethian, 1999b) is:

1. Initialize
   (a) *Alive* points: Let A be the set of all grid points (i,j) on the starting position of the interface Γ; set $T_{ij} = 0$ for all points in A.
   (b) *Narrow Band* points: Let the *Narrow Band* be the set of all grid points (i,j) in the immediate neighborhood of A; set $T_{ij} = \frac{d}{F_{ij}}$ for all points in the *Narrow Band* where, d is the grid separation distance and F is the front speed (see Eq. (13)).
   (c) *Far Away* points: Let the *Far Away* region be the set of all remaining grid points (i,j); set $T_{ij} = \infty$ for all points in the *Far Away* region.
2. Marching Forward
   (a) Begin Loop: Let $(i_{\min},j_{\min})$ be the point in the *Narrow Band* with the smallest value for T.
   (b) Add the point $(i_{\min},j_{\min})$ to A; remove it from the *Narrow Band*.
   (c) Tag as neighbors any points $(i_{\min}-1,j_{\min})$, $(i_{\min}+1,j_{\min})$, $(i_{\min},j_{\min}-1)$, $(i_{\min},j_{\min}+1)$ that are either in the *Narrow Band* or the *Far Away* region. If the neighbor is in the *Far Away* region, remove it from that list and add it to the *Narrow Band*.
   (d) Recompute values of T at all neighbors in accordance with Eq. (14). Select the largest possible solution to the quadratic equation.
   (e) Return to the top of the loop.

Here are some properties of the fast marching algorithm. The smallest value in the *Narrow Band* is always correct. Other *Narrow Band* or *Far Away* points with larger values of T cannot affect the smallest value. Also, the process of recomputing T values at the neighboring points cannot give a value smaller than any of the accepted value at *Alive points*, since the correct solution is obtained by selecting the largest possible solution to the quadratic equation (Eq. (14)). Thus the algorithm marches forward by selecting the minimal T value in the *Narrow Band* and recomputing the values of T at all neighbors in accordance with Eq. (14).

The key to an efficient version of the algorithm lies in finding a fast way to locate the grid point in the *Narrow Band* with the minimum value for T. To do so, the heapsort algorithm (Williams, 1964; Sedgewick and Wayne, 2011) with backpointers is often implemented and it is the algorithm we used here. This sorting algorithm generates a "complete binary tree" with the property that the value at any given parent node is less than or equal to the value at its child node. Heap is represented sequentially by storing a parent node at the location k and its child at locations $2k$ and $2k+1$. The member having the smallest value is stored at the location $k = 1$.

All *Narrow Band* points are initially sorted in a heapsort. The fast marching algorithm works by first finding, and then removing, the member corresponding to the smallest T value from the *Narrow Band* which is followed by one sweep of DownHeap to ensure that the remaining elements satisfy the heap property. The DownHeap operation moves the element downwards in the heap until the new heap satisfies the heap properties. *Far Away* neighbors are added to the heap using the Insert operation which increases the heap size by one and brings the new element to its correct heap location using the UpHeap operation. The UpHeap operation moves the element upwards in the heap until the new heap satisfies the heap properties. The updated values at the neighbor points obtained from Eq. (14) are also brought to the correct heap location by performing the UpHeap operation.

Appendix C. Estimating the total velocity field under geostrophic balance by minimizing unknown inter-island transports

For mesoscale ocean flows, away from boundary layers, the dominant terms in the horizontal momentum equations are often the Coriolis force and the pressure gradient. Such a flow field, where a balance is struck between the Coriolis and pressure forces, is called geostrophic. The thermal wind equations are obtained for geostrophic flows by assuming that the vertical momentum equation is approximately given by hydrostatic balance. The thermal wind equations are:

$$-f\frac{\partial(\rho v)}{\partial z}=g\frac{\partial \rho}{\partial x} \quad \text{and} \quad f\frac{\partial(\rho u)}{\partial z}=g\frac{\partial \rho}{\partial y}, \tag{C.1}$$

where, ρ is the density, u and v are the horizontal fluid velocity in the zonal (x) and meridional (y) directions respectively, and $f = 2\Omega\, sin\phi$ is the Coriolis parameter at latitude ϕ for the spherical earth rotating at a rate of Ω. The thermal wind Eq. (C.1) when integrated in the vertical give:

$$\begin{aligned}\rho v(x,y,z,t) &= \frac{-g}{f}\int_{z_0}^{z}\frac{\partial \rho}{\partial x}dz+\rho v_0,\\ \rho u(x,y,z,t) &= \frac{g}{f}\int_{z_0}^{z}\frac{\partial \rho}{\partial y}dz+\rho u_0,\end{aligned} \tag{C.2}$$

where, z_0 is a level of reference where v_0, u_0 are assumed known (z_0 is referred to the level of no motion if v_0, $u_0 = 0$).

Flow estimation based on thermal wind balance (Eq. (C.2)) is a classical problem in oceanography (Wunsch, 1996). Historically, the main routine measurements were hydrographic: temperature, T, and salinity, S, at various depths. The equation of state for seawater then permits the estimation of density at a given pressure from these hydrographic data. Thus, with Eq. (C.2), the vertical shear of the geostrophic flow can be computed from hydrographic data alone and added to a velocity field of reference. This leads to mass-conserving estimates if the reference velocity field is conservative since the geostrophic shear already satisfies continuity. If reference or external barotropic velocities are provided at open boundaries, a Poisson equation can be formed for a transport streamfunction by taking the curl of this barotropic velocity. Solving for the transport streamfunction is then straightforward for domains without any islands. For complex coastal regions with islands, the same Poisson equation can be solved, imposing a fixed transport streamfunction value around each island. The result conserves mass by construction. Details are provided in App. 2.2 of Haley and Lermusiaux (2010) for both rigid-lid and free-surface primitive equations.

In the case with islands, a first-guess at the streamfunction along each island coast can be obtained by sinking the islands to a shallow depth, solving for the corresponding streamfunction and averaging its values along each island coast. However, we found that some of the resulting inter-island transports can be unrealistic, often much too large. Hence, a methodology was derived to correct for this (MSEAS, 2010). Specifically, the somewhat known inter-island transports are optimized (i.e. add a least-square penalty towards these values) and the unknown ones minimized. These optimized island transport streamfunctions are then used as Dirichlet boundary conditions in the Poisson equation. The result is a mass-conserving geostrophic flow in balance with the hydrographic OA maps and with optimized inter-island transports. This methodology was illustrated in Section 6.2.

Summarizing the inter-island transport optimization, the objective is to find a set of constant values for (Ψ) along the island coastlines that produce a suitably smooth initialization velocity field, e.g. with no unrealistically large velocities. In the unknown straits, the goal is to minimize the kinetic energy or the maximum absolute velocity. The working assumptions are:

1. Coastlines in the given domain can be divided into two distinct subsets:
   (a) Set A: N coastlines along which the transport streamfunction is unknown, $N \neq 0$.
   (b) Set B: M coastlines along which the transport streamfunction is known.
2. A first-guess Ψ_0 exists for the case with coasts in set B, but no coasts in set A, i.e. these coasts and their corresponding interiors are replaced by open ocean (e.g. island sunk to 10 m depth).
3. The difference between the first-guess Ψ_0 and the final solution Ψ is not extremely large. Otherwise, Ψ_0 would not be accurate enough.

Ψ_0 contains useful information such as the position of major currents relative to various coastlines and the effects of topography on the flow. Thus, Ψ_0 can be used to estimate Ψ along the other island coastlines by constructing an optimization functional for minimizing (in general optimizing) the inter-island transports subject to weak constraints. The optimization functional (E) is constructed as follows. Its general form is divided into a summation of three terms, given by:

$$E = E_1 + E_2 + E_3, \tag{C.3}$$

where, E_1 is the minimizing target for the transport between all pairs of the unknown (Set A) coasts, E_2 is the minimizing target for the transport between all pairs of unknown (Set A) and known (Set B) coasts and E_3 is the minimizing target for the transport between all pairs of the unknown (Set A) coasts and the open boundaries of the domain. The minimum of E is computed by solving a standard least square problem, i.e. by setting gradients with respect to the unknown Ψ values equal to zero. These streamfunction values, which smooth the velocity field, are then used as Dirichlet boundary conditions to the final Poisson equation.

The expressions for E_1, E_2 and E_3 are provided in MSEAS (2010). They require the use of appropriate weights: w_{nm} for the pair of islands denoted here by subscripts n and m. These weights are computed using a FMM scheme. Specifically, consider the stream function (Ψ) for a 2D horizontal flow. It is defined such that the flow velocity can be expressed as:

$$\vec{u} = (u, v) = -\frac{1}{H}\nabla \times \Psi\hat{k} \Rightarrow u = -\frac{1}{H}\frac{\partial \Psi}{\partial y}, \quad v = \frac{1}{H}\frac{\partial \Psi}{\partial x}. \tag{C.4}$$

Here, H is the ocean depth. The transport between a pair of islands having streamfunction ψ_1 and ψ_2 is given by:

$$\psi_2 - \psi_1 = \int_A \vec{u}\cdot\hat{n}dA, \tag{C.5}$$

where, A is the vertical area between the two islands and $\hat{n}$ is the unit vector normal to the vertical area. Eqs. (C.4) and (C.5) suggest that the appropriate weight function to optimize the velocity field should be $w_{nm} = 1/A_{nm}^2$, where, A_{nm} is the minimum vertical area along any path between the two islands n and m. Another heuristic choice of weight function can be $w_{nm} = 1/d_{nm}^2$, where the d_{nm}'s are mean depths. We found this choice only appropriate when the depth is almost uniform in between each pair of islands (n,m). In general, this is not the case and we thus needed to compute the minimum areas A_{nm}. Using the FMM, as described in Section 4.2, is a very convenient and efficient way to compute these A_{nm}'s. Simulations (Agarwal, 2009) have been performed with several other weight functions and they confirmed that the choice of weights $w_{nm} = 1/A_{nm}^2$ lead to the most accurate flow fields.

References

Agarwal, A., 2009. Statistical field estimation and scale estimation for complex coastal regions and archipelagos. SM Thesis. Massachusetts Institute of Technology.

Antonov, J.I., Levitus, S., Boyer, T.P., Conkright, M.E., O Brien, T.D., Stephens, C., 1998a. World Ocean Atlas 1998. In: NOAA Atlas NESDIS 27. Temperature of the Atlantic Ocean, vol. 1. US Government Printing Office, Washington, DC.

Antonov, J.I., Levitus, S., Boyer, T.P., Conkright, M.E., O Brien, T.D., Stephens, C., 1998b. World Ocean Atlas 1998. In: NOAA Atlas NESDIS 28. Temperature of the Pacific Ocean, vol. 2. US Government Printing Office, Washington, DC.

Antonov, J.I., Levitus, S., Boyer, T.P., Conkright, M.E., O Brien, T.D., Stephens, C., 1998c. World Ocean Atlas 1998. In: NOAA Atlas NESDIS 29. Temperature of the Indian Ocean, vol. 3. US Government Printing Office, Washington, DC.

Antonov, J.I., Locarnini, R.A., Boyer, T.P., Mishonov, A.V., Garcia, H.E., 2006. World Ocean Atlas 2005. In: Levitus, S. (Ed.), NOAA Atlas NESDIS 62, Salinity, vol. 2. US Government Printing Office, Washington, DC.

Balgovind, R., Dalcher, A., Ghil, M., Kalnay, E., 1983. A stochastic dynamic model for the spatial structure of forecast error statistics. Monthly Weather Review 111, 701–722.

Barnes, S.L., 1964. A technique for maximizing details in numerical weather map analysis. Journal of Applied Meteorology 3, 396–409.

Barnes, S.L., 1973. Mesoscale objective map analysis using weighted time series observations. NOAA Technical Memorandum, ERL NSSL-62.

Barnes, S.L., 1994. Applications of the Barnes objective analysis scheme, part III: tuning for minimum error. Journal of Atmospheric and Oceanic Technology 11, 1459–1479.

Bertsimas, D., Tsitsiklis, J.N., 1997. Introduction to Linear Optimization. Athena Scientific, Belmont, Massachusetts.

Boyer, T.P., Levitus, S., Antonov, J.I., Conkright, M.E., O Brien, T.D., Stephens, C., 1998a. World Ocean Atlas 1998. In: NOAA Atlas NESDIS 30. Temperature of the Atlantic Ocean, vol. 1. US Government Printing Office, Washington, DC.

Boyer, T.P., Levitus, S., Antonov, J.I., Conkright, M.E., O Brien, T.D., Stephens, C., 1998b. World Ocean Atlas 1998. In: NOAA Atlas NESDIS 31. Temperature of the Pacific Ocean, vol. 2. US Government Printing Office, Washington, DC.

Boyer, T.P., Levitus, S., Antonov, J.I., Conkright, M.E., O Brien, T.D., Stephens, C., 1998c. World Ocean Atlas 1998. In: NOAA Atlas NESDIS 32. Temperature of the Indian Ocean, vol. 3. US Government Printing Office, Washington, DC.

Boyer, T.P., Stephens, C., Antonov, J.I., Conkright, M.E., Locarnini, R.A., O Brien, T.D., Garcia, H.E., 2002. World Ocean Atlas 2001. In: Levitus, S. (Ed.), NOAA Atlas NESDIS 50, Salinity, vol. 2. US Government Printing Office, Washington, DC.

Brankart, J.-M., Brasseur, P., 1996. Optimal analysis of in situ data in the Western Mediterranean using statistics and cross-validation. Journal of Atmospheric and Oceanic Technology 16 (2).

Bresenham, J.E., 1965. Algorithm for computer control of a digital plotter. IBM Systems Journal 4 (1), 25–30.

Bretherton, F.P., Davis, R.E., Fandry, C., 1976. A technique for objective analysis and design of oceanographic instruments applied to MODE-73. Deep-Sea Research 23, 559–582.

Brown, R.G., Hwang, P.Y.C., 1997. Introduction to Random Signals and Applied Kalman Filtering. John Wiley & Sons, United Kingdom.

Burton, L., 2009. Modeling coupled physics and biology in ocean straits. SM Thesis. Massachusetts Institute of Technology.

Carter, E., Robinson, A., 1987. Analysis models for the estimation of oceanic fields. Journal of Atmospheric and Oceanic Technology 4 (1), 49–74.

Cho, Y., Shin, V., Oh, M., Lee, Y., 1996. Suboptimal continuous filtering based on the decomposition of the observation vector. Computers and Mathematics with applications 32 (4), 23–31.

Cossarini, G., Lermusiaux, P.F.J., Solidoro, C., 2009. The lagoon of Venice ecosystem: seasonal dynamics and environmental guidance with uncertainty analyses and error subspace data assimilation. Journal of Geophysical Research. doi:10.1029/2008JC005080.

Cressman, G.P., 1959. An operational objective analysis scheme. Monthly Weather Review 87, 329–340.

Daley, R., 1993. Atmospheric Data Analysis. Cambridge University Press.

Denman, K.L., Freeland, H.J., 1985. Correlation scales, objective mapping and a statistical test of geostrophy over the continental shelf. Journal of Marine Research 43, 517–539.

Derber, J.C., Bouttier, F., 1999. A reformulation of the background error covariance in the ECMWF global data assimilation system. Tellus A 51 (2), 195–221.

Dolloff, J., Lofy, B., Sussman, A., Taylor, C., 2006. Strictly positive definite correlation functions. Proceedings of SPIE, 6235.

Gamo, T., Kato, Y., Hasumoto, H., Kakiuchi, H., Momoshima, N., Takahata, N., Sano, Y., 2007. Geochemical implications for the mechanism of deep convection in a semi-closed tropical marginal basin: Sulu Sea. Deep-Sea Research II 54, 4–13.

Gandin, L.S., 1965. Objective analysis of meteorological fields. Israel Program for Scientific Translations.

Gangopadhyay, A., Robinson, A.R., 2002. Feature oriented regional modeling of oceanic fronts. Dynamics of Atmosphere and Oceans 36 (1-3), 201–232.

Garcia, H.E., Locarnini, R.A., Boyer, T.P., Antonov, J.I., 2006a. World Ocean Atlas 2005. In: Levitus, S. (Ed.), NOAA Atlas NESDIS 63, Dissolved Oxygen, Apparent Oxygen Utilization, and Oxygen Saturation, vol. 3. US Government Printing Office, Washington, DC.

Garcia, H.E., Locarnini, R.A., Boyer, T.P., Antonov, J.I., 2006b. World Ocean Atlas 2005. In: Levitus, S. (Ed.), NOAA Atlas NESDIS 64, Nutrients (Phosphate, Nitrate, Silicate), vol. 4. US Government Printing Office, Washington, DC.

Gordon, A.L., 2009. Philex: Regional cruise intensive observational period, leg 1 and 2 reports.

Gordon, A.L., Sprintall, J., Ffield, A., 2011. Regional oceanography of the Philippine Archipelago. Oceanography 24 (1), 14–27.

Haley, P.J., Lermusiaux, P.F.J., 2010. Multiscale two-way embedding schemes for free-surface primitive-equations in the Multidisciplinary Simulation, Estimation and Assimilation System. Ocean Dynamics, 1497–1537. doi:10.1007/s10236-010-0349-4 60.

Haley, P.J., Lermusiaux, P.F.J., Robinson, A.R., Leslie, W.G., Logoutov, O.G., Cossarini, G., Liang, X.S., Moreno, P., Ramp, S.R., Doyle, J.D., Bellingham, J., Chavez, F., Johnston, S., 2009. Forecasting and reanalysis in the Monterey Bay/California Current region for the Autonomous Ocean Sampling Network-II experiment. Deep Sea Research II, 68–86. doi:10.1016/j.dsr2.2008.08.010 56.

Hessler, G., 1984. Experiments with statistical objective analysis techniques for representing a coastal surface temperature field. Boundary Layer Meteorology 28, 375–389.

Lam, F.P., Haley, P.J., Jr., Janmaat, J., Lermusiaux, P.F.J., Leslie, W.G., Schouten, M.W., te Raa, L.A., Rixen, M., 2009. At-sea real-time coupled four-dimensional oceanographic and acoustic forecasts during Battlespace Preparation 2007. Special issue of the Journal of Marine Systems on "Coastal processes: challenges for monitoring and prediction, Drs. J.W. Book, Prof. M. Orlic and Michel Rixen (Guest Eds.), 78, 306–320. Doi: doi:10.1016/j.jmarsys.2009.01.029.

Lermusiaux, P.F.J., 1997. Error subspace data assimilation methods for ocean field estimation: theory, validation and applications. Ph.D. Thesis. Harvard University.

Lermusiaux, P.F.J., 1999. Data assimilation via error subspace statistical estimation, part II: Middle Atlantic Bight shelfbreak front simulations and ESSE validation. Monthly Weather Review 127, 1408–1432.

Lermusiaux, P.F.J., 2002. On the mapping of multivariate geophysical fields: sensitivities to size, scales, and dynamics. Journal of Atmospheric and Oceanic Technology 19, 1602–1637.

Lermusiaux, P.F.J., 2007. Adaptive modeling, adaptive data assimilation and adaptive sampling. Physica D 230, 172–196.

Lermusiaux, P.F.J., Anderson, D.G.M., Lozano, C.J., 2000. On the mapping of multivariate geophysical fields: error and variability subspace estimates. Quarterly Journal of the Royal Meteorological Society 126, 1387–1429.

Lermusiaux, P.F.J., Xu, J., Chen, C.F., Jan, S., Chiu, L.Y., Yang, Y.J., 2010. Coupled ocean-acoustic prediction of transmission loss in a continental shelfbreak region: predictive skill, uncertainty quantification and dynamical sensitivities. IEEE Transactions on Journal of Oceanic Engineering 35 (4), 895–916. doi:10.1109/JOE.2010.2068611.

Lermusiaux, P.F.J., Haley Jr., P.J., Leslie, W.G., Agarwal, A., Logoutov, O.G., Burton, L.J., 2011. Multiscale physical and biological dynamics in the Philippines Archipelago: Predictions and processes. Oceanography 24 (1), 70–89. doi:10.5670/oceanog.2011.05, Special PhilEx Issue.

Levitus, S., 1982. Climatological Atlas of the World Ocean. NOAA Professional Paper 13. US Government Printing Office, Washington, DC.

Levitus, S., Boyer, T.P., 1994. World Ocean Atlas 1994. In: NOAA Atlas NESDIS 4. Temperature, vol. 4. US Government Printing Office, Washington, DC.

Levitus, S., Burgett, R., Boyer, T.P., 1994. World Ocean Atlas 1994. In: NOAA Atlas NESDIS 3. Salinity, vol. 3. US Government Printing Office, Washington, DC.

Locarnini, R.A., Mishonov, A.V., Antonov, J.I., Boyer, T.P., Garcia, H.E., 2006. World Ocean Atlas 2005. In: Levitus, S. (Ed.), NOAA Atlas NESDIS 61, Temperature, vol. 1. US Government Printing Office, Washington DC.

Logoutov, O.G., 2008. A multigrid methodology for assimilation of measurements into regional tidal models. Ocean Dynamics, 441–460. doi:10.1007/s10236-008-0163-4 58.

Logoutov, O.G., Lermusiaux, P.F.J., 2008. Inverse barotropic tidal estimation for regional ocean applications. Ocean Modelling, 17–34. doi:10.1016/j.ocemod.2008.06.004 25.

Lynch, D.R., McGillicuddy, D.J., 2001. Objective analysis for coastal regimes. Continental Shelf Research 21, 1299–1315.

MSEAS, 2010. The Multidisciplinary Simulation, Estimation, and Assimilation System. Reports in Ocean Science and Engineering 6. Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge. <http://mseas.mit.edu/>, <http://mseas.mit.edu/codes>.

Osher, S., Sethian, J.A., 1988. Fronts propagating with curvature-dependent speed: algorithms based on Hamilton–Jacobi formulations. Journal of Computational Physics 79 (1), 12–49.

Papoulis, A., 1991. Probability, Random Variables and Stochastic Processes. McGraw-Hill.

Paris, C.B., Cowen, R.K., Lwiza, K.M.M., Wang, D.P., Olson, D.B., 2002. Multivariate objective analysis of the coastal circulation of Barbados, West Indies: implication for larval transport. Deep-Sea Research I 49, 1363–1386.

Parrish, D.F., Cohn, S.E., 1985. A Kalman filter for a two-dimensional shallow-water model: formulation and preliminary experiments. Office Note 304. US Department of Commerce, NOAA, NWS, NMC.

Pedlosky, J., 1987. Geophysical Fluid Dynamics. Springer.

Plackett, R.L., 1950. Some theorems in least squares. Biometrika 37 (1–2), 149–157.

Ruoy, E., Tourin, A., 1992. A viscosity solutions approach to shape from shading. SIAM Journal on Numerical Analysis 29, 867–884.

Sedgewick, R., Wayne, K., 2011. Algorithms. Pearson Education, Boston, MA.

Selvadurai, A.P.S., 2000. Partial Differential Equations in Mechanics. Springer.
Sethian, J.A., 1996. A Fast Marching Level Set Method for monotonically advancing fronts. Proceedings of the National Academy of Sciences 93 (4), 1591–1595.
Sethian, J.A., 1999a. Fast Marching Methods. SIAM Review 41 (2), 199–235.
Sethian, J.A., 1999b. Level Set Methods and Fast Marching Method. Cambridge University Press, Cambridge, United Kingdom.
Stacey, M.W., Pond, S., LeBlond, P.H., 1988. An objective analysis of the low-frequency currents in the Strait of Georgia. Atmosphere–Ocean 26 (1), 1–15.
Stephens, C., Antonov, J.I., Boyer, T.P., Conkright, M.E., Locarnini, R.A., O Brien, T.D., Garcia, H.E., 2002. World Ocean Atlas 2001. In: Levitus, S. (Ed.), NOAA Atlas NESDIS 49, Temperature, vol. 1. US Government Printing Office, Washington, DC.
Weaver, A., Courtier, P., 2001. Correlation modelling on the sphere using a generalized diffusion equation. Quarterly Journal of the Royal Meteorological Society 127, 1815–1846.
Williams, J.W.J., 1964. Algorithm 232 – heapsort. Communications of the ACM 7 (6), 347348.
Wunsch, C., 1996. The Ocean Circulation Inverse Problem. Cambridge University Press, Cambridge, United Kingdom.
Xu, J., Lermusiaux, P.F.J., Haley, P.J., Leslie, W.G., Logoutov, O.G., 2008. Spatial and temporal variations in acoustic propagation during the PLUSNet07 exercise in Dabob Bay. In: Proceedings of Meetings on Acoustics (POMA), 155th Meeting Acoustical Society of America. Doi: doi:10.1121/1.29880934.
Yaglom, A.M., 1987. Correlation Theory of Stationary and Related Random Functions I. Springer-Verlag.
Yaglom, A.M., 2004. An Introduction to the Theory of Stationary Random Functions, Dover Phoenix Editions. Dover.