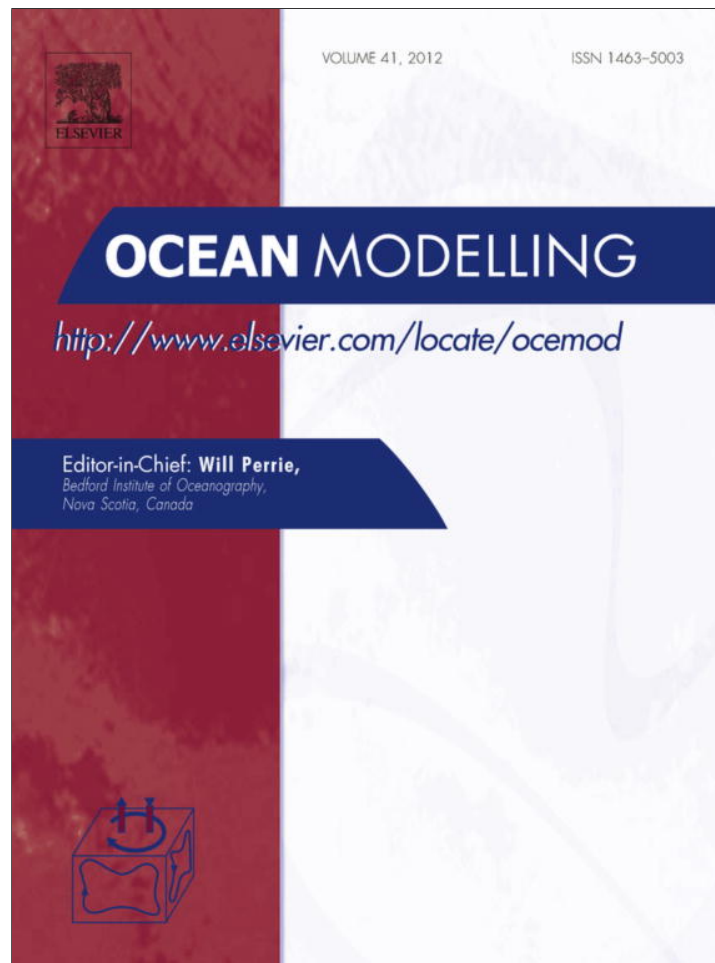


Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



(This is a sample cover image for this issue. The actual cover is not yet available at this time.)

This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>

Contents lists available at [SciVerse ScienceDirect](http://www.sciencedirect.com)

Ocean Modelling

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

Assimilation of glider and mooring data into a coastal ocean model

Emlyn M. Jones^{a,*}, Peter R. Oke^a, Farhan Rizwi^a, Lawrence M. Murray^b^a CSIRO Marine and Atmospheric Research, GPO Box 1538, Hobart, TAS 7001, Australia^b CSIRO Mathematical and Information Sciences, Private Bag 5, WA 6913, Australia

ARTICLE INFO

Article history:

Received 4 July 2011

Received in revised form 12 December 2011

Accepted 19 December 2011

Available online 1 February 2012

Keywords:

Data assimilation
Coastal ocean modelling
Tasmania
Glanders
SHOC

ABSTRACT

We have applied an ensemble optimal interpolation (EnOI) data assimilation system to a high resolution coastal ocean model of south-east Tasmania, Australia. The region is characterised by a complex coastline with water masses influenced by riverine input and the interaction between two offshore current systems. Using a large static ensemble to estimate the systems background error covariance, data from a coastal observing network of fixed moorings and a Slocum glider are assimilated into the model at daily intervals. We demonstrate that the EnOI algorithm can successfully correct a biased high resolution coastal model. In areas with dense observations, the assimilation scheme reduces the RMS difference between the model and independent GHRSSST observations by 90%, while the domain-wide RMS difference is reduced by a more modest 40%. Our findings show that errors introduced by surface forcing and boundary conditions can be identified and reduced by a relatively sparse observing array using an inexpensive ensemble-based data assimilation system.

Crown Copyright © 2012 Published by Elsevier Ltd. All rights reserved.

1. Introduction

Operational coastal ocean modelling has benefited from recent advances in global ocean modelling projects such as *Bluelink* (Oke et al., 2008) and high resolution meteorological products (e.g. Vincent et al., 2008), which supply open ocean and surface boundary conditions, respectively. A key reason for the recent success in eddy-resolving global ocean models has been the inception and dissemination of high quality observational data provided by networks such as the Global Ocean Observing System (<http://www.ioc-goos.org/>). Until recently, most coastal observing systems have lacked a centralised data collection and dissemination system. With the advent of regional observing systems like the Australian Integrated Marine Observing System – IMOS (<http://www.imos.org.au>), this has now changed. The challenge becomes one of integrating these available observations and improved forcing/boundary conditions into our efforts to model the coastal domain. Model-data fusion algorithms, often referred to as data assimilation, provide a robust framework in which to do this.

There are many examples of data assimilation (DA) techniques applied to ocean modelling on scales ranging from global to coastal. The data assimilation techniques that are currently in use within an ocean modelling framework can be divided into four broad categories:

1. Variational Methods, e.g. 3D/4D – Var,

2. static parametric ensemble methods, e.g. ensemble optimal interpolation (EnOI),
3. dynamic parametric ensemble methods, e.g. Kalman Filter/Smoothing (KF/S), and
4. dynamic non-parametric ensemble methods, e.g. Particle Filters.

The Variational (Kurapov et al., 2011; DiLorenzo et al., 2007) and EnOI (Oke et al., 2008) approaches tend to dominate the assimilation schemes, while the Kalman (Evensen, 1994; Bertino and Lisaeter, 2008) and Particle Filter (van Leeuwen, 2009) approaches are not so well represented, largely due to the computational burden of propagating forward an ensemble with 100–1000 members. A summary of the schemes used in GODAE, which largely represent global and regional models, is given in Cummings et al. (2009). Other approaches that do not fit into the aforementioned categories include the traditional OI approaches including the original MERCATOR SAW algorithm (Brasseur et al., 2005), and fast approximate approaches such as ESSE and emulators (Lermusiaux and Robinson, 1999; Margvelashvili and Campbell, 2012). The computational burden associated with carrying forward a dynamic ensemble or the iterative minimisation of a cost function can be circumvented through the use of a static background ensemble using the EnOI approach of Oke et al. (2008). This has been shown to be a computationally tractable algorithm and has been argued that it gives superior performance in some cases (Oke et al., 2010).

The application of DA systems in coastal/estuary domains could be considered in its infancy when compared with their application in the global ocean and the maturing field of numerical weather

* Corresponding author. Tel.: +61 3 62325483.

E-mail address: emlyn.jones@csiro.au (E.M. Jones).

prediction. In this study we have chosen to test an EnOI assimilation technique in the coastal region of south-east Tasmania, Australia (Fig. 1). The coastal and shelf regions of south-east Tasmania are influenced by a number of water-masses where there is a seasonal interaction between the extension of the East Australia Current (EAC) and the South Australian Current (SAC) (Schiller et al., 2008). The seasonal nature of this interaction is discussed in Cresswell (2000). The near-shore region has been classified as a Region of Fresh Water Influence (ROFI) due to the freshwater input from two major river systems (Herzfeld et al., 2010). To this end, the central region of Storm Bay contains the signatures of the EAC, SAC and the inshore ROFI. As with many coastal domains, this region is a challenge to model as any inaccuracies in either the forcing data or prescribed boundary conditions will strongly influence the internal dynamics of the model.

Unlike “blue-water” or open-ocean eddy resolving ocean general circulation models (OGCMs), where meso-scale chaotic behaviour (e.g. eddies) is the dominant source of error, the effects of friction in coastal ocean models tend to damp out this meso-scale chaotic behaviour. However, fine-scale features such as eddy shedding in the lee of headlands and the chaotic behaviour of upwelling and river plume filaments may be present, although in many cases difficult to observe. Observation platforms such as gliders (Rudnick et al., 2004) and AUV's can provide high resolution data in both time and space with coastal domains (Schofield et al., 2007). They are analogous to the profiling Argo floats used in OGCM DA systems. In coastal domains where the water depth exceeds 30 m, gliders are proving to be an invaluable observing platform to source observations of the water column for assimilation (Shulman et al., 2009; Dobricic et al., 2010).

Our fundamental premise for invoking a data assimilation scheme in a coastal ocean model is to correct for any errors introduced into the model due to the forcing data or prescribed boundary conditions, rather than the correction of chaotic behaviour. To the best of our knowledge, this is the first time that an EnOI algorithm has been applied to a high resolution (500–2000 m grid) coastal model. The purpose of this paper is to:

1. present an ensemble optimal interpolation (EnOI) scheme suitable for use in coastal ocean modelling,
2. investigate the correlation structure that the static ensemble has captured,
3. make use of both fixed and mobile in situ observation platforms,

4. assess the ability of the EnOI scheme to correct for biases in the model state that enter through forcing and boundary conditions.

This paper is organised into the following sections. The South East Tasmania (SETas) model and ensemble optimal interpolation (EnOI) algorithm is described in Section 2. The SETas EnOI system is then trialled in a three month case study (Section 3) followed by a discussion (Section 4) and conclusion (Section 5).

2. Model and algorithm description

2.1. SETas model

The south-east Tasmania (SETas) model is based on SHOC (Sparse Hydrodynamic Ocean Code), a finite difference, hydrodynamic, primitive equation model (Herzfeld, 2006). SHOC is based on the three-dimensional equations of momentum, continuity and conservation of heat and salt, employing the hydrostatic and Boussinesq assumptions. The equations of motion are discretized on a finite difference stencil corresponding to the Arakawa C grid. The model uses a curvilinear orthogonal grid in the horizontal, with a resolution of approximately 250 m in the estuaries and rivers, increasing to 2500 m at the open ocean boundaries. The vertical coordinate system uses a fixed staggered z coordinate grid, with a resolution of 1 m in the upper layers increasing to 10 m in the deeper layers. The z vertical system allows for wetting and drying of surface cells, useful for modelling regions such as tidal flats where large areas are periodically dry. SHOC has a free surface and uses mode splitting to separate the two-dimensional (2D) mode from the three-dimensional (3D) mode. The SETas model grid is shown in Fig. 2, with details included in Table 1.

2.2. SETas EnOI

Analysis fields for temperature (T) and salinity (S) are computed using the ensemble optimal interpolation (EnOI) analysis equations detailed in Oke et al. (2008):

$$w^a = w^b + K(w^o - Hw^b), \tag{1}$$

where w is the state vector, and the superscripts a , b and o denote the analysis, background and observed fields respectively. The state vector is given by:

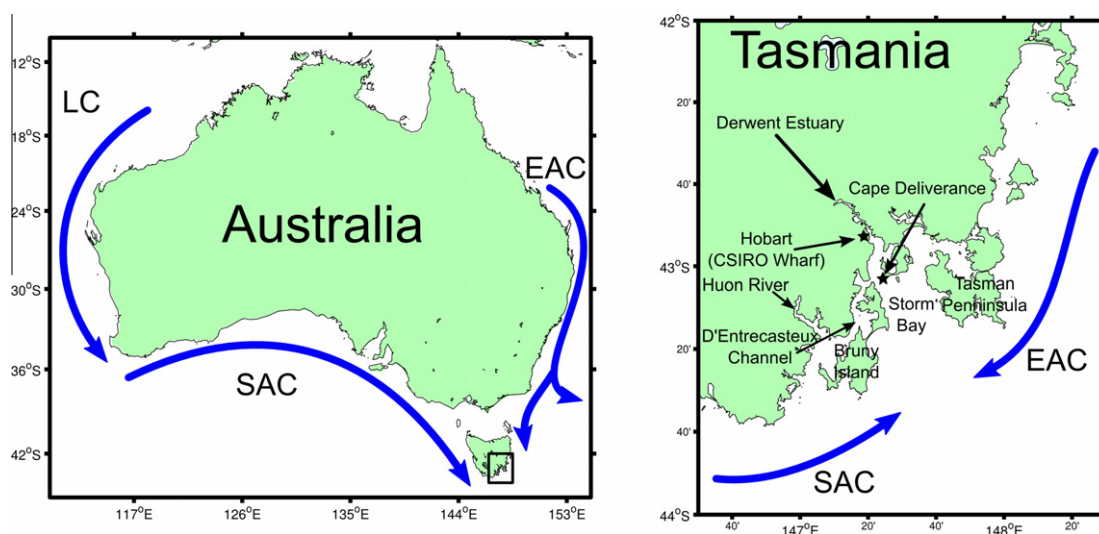


Fig. 1. The regional circulation around Australia (left panel) is dominated by the Leeuwin Current (LC) in the west, the East Australian Current (EAC) along the eastern seaboard and the South Australian Current in the south. The coastal ensemble optimal interpolation scheme is trialled in south-east Tasmania (right panel). This area is characterised as a mixing zone for the extensions of the EAC and SAC as well as the inshore region of freshwater influence (ROFI).

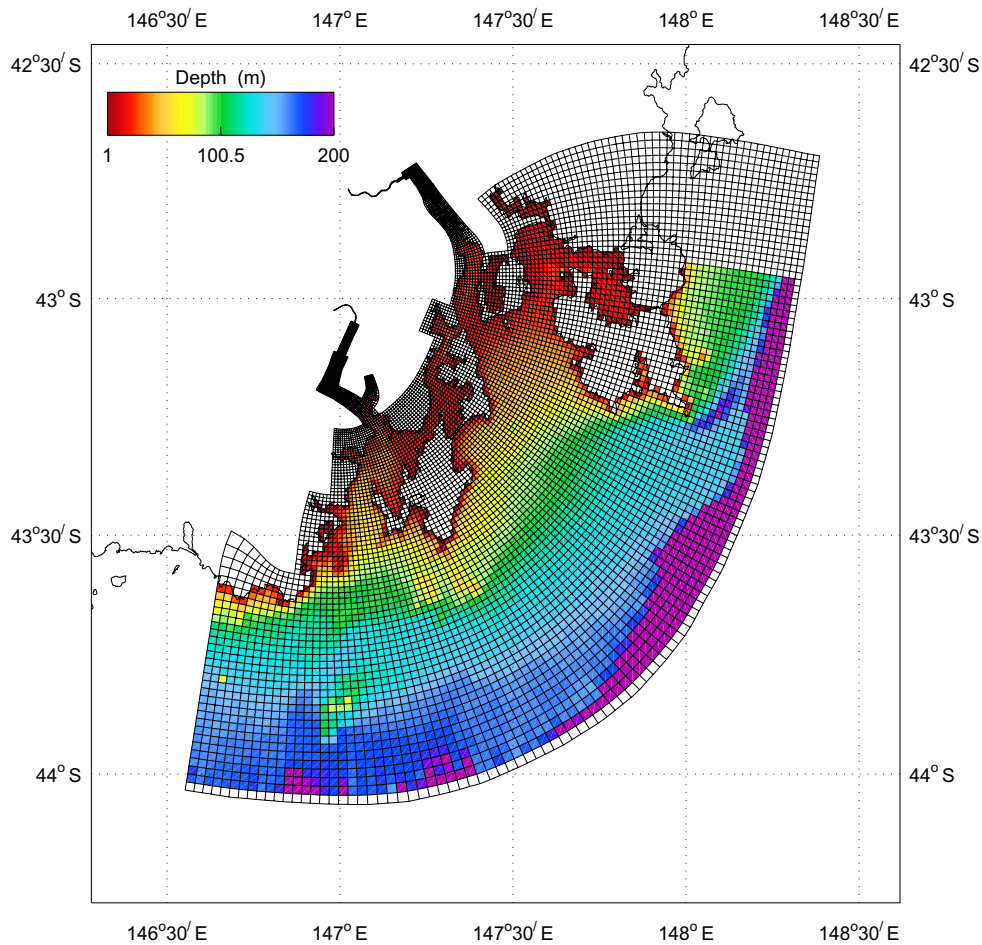


Fig. 2. SETas Model bathymetry and grid.

Table 1
 SETas model configuration.

Horizontal grid	Orthogonal curvilinear: 175 × 120 cells
Horizontal resolution	2500 m (Open Ocean) 250 m (Huon River)
Vertical grid	Geopotential vertical coordinate system: 21 layers
Vertical resolution	1–10 m
Minimum depth	1 m
Maximum depth	200 m
Atmospheric forcing	MesoLAPS
Open ocean boundary forcing	OceanMAPS
Head of estuary forcing	Gauged river flow
Time step (dt)	45 s
Bathymetry	Margvelashvili et al. (2009)
Vertical mixing	Mellor–Yamada 2.0
Bottom friction	Combination of linear and quadratic (Herzfeld and Waring, 2009)

$$w = [T \ S]^T, \quad (2)$$

where the superscript T denotes the matrix transpose. We could include sea-level (η) and the horizontal velocity vectors (u, v), but at this stage we do not assimilate or relax towards any sea-level or velocity fields (due to lack of observations). K is the gain matrix:

$$K = PH^T(HPH^T + R)^{-1}, \quad (3)$$

H is an operator that interpolates from the model grid to measurement location. R is the observation error covariance matrix and is in this case assumed to be diagonal. P is an estimate of the background error covariance (BEC) matrix:

$$P = \frac{AA^T}{n-1}, \quad (4)$$

where A is an ensemble of anomaly fields containing all the state variables from (2). If there are m wet cells in the model, A will be $2m$ rows (2, because we are only adjusting T and S state variables) by n columns, where n is the ensemble size

$$A = [w'_1 w'_2 w'_3 \dots w'_n], \quad (5)$$

the $'$ denotes the anomaly field of the state. The ensemble of anomaly fields was constructed from a long control run of the SETas model (see Section 3.1), where no data were assimilated, and spans the period of 1st of March 2008 to 31st of December, 2009. The anomaly fields used in (5) are intended to represent the dominant errors within the model (Oke et al., 2008). There appears to be little, if any, formal consensus in the formulation of these anomaly fields. In the EnOI formulation of Counillon and Bertino (2009), the ensemble is constructed using weekly output of the Gulf of Mexico model. Implicit in their ensemble construction is the introduction of seasonal variation, which is addressed through a scaling factor which can be introduced into (4), and imposes a degree of stationarity to the ensemble. Alternatively, Oke et al. (2008) constructs the anomaly fields by subtracting monthly means from seasonal means, again taken from a long integration of the model with no assimilation, which yield anomaly fields that contain length-scales associated with meso-scale chaotic behaviour. Oke et al. (2008) do not use a scaling factor as it is assumed that the ensemble is stationary.

The two time-scales chosen to construct the anomaly fields are therefore somewhat arbitrary. However, there are some logical

choices to use in the coastal ocean; we want to choose as short a timescale as possible to represent the “errors of the day”, yet this needs to be long enough to preclude the effects of tides. The natural choice then is to average over one tidal cycle which corresponds to approximately one day in this domain, \bar{w}_{1day} , where the over-bar denotes a time averaged value. The second timescale, should be characteristic of the timescales in which the errors are introduced. We hypothesise that these errors are introduced into the model over timescales that are shorter, or approximately equal to, that of the synoptic weather-band forcing; timescales that range between 5 and 12 days. We have chosen to average over 7 days, \bar{w}_{7day} . The control run of the SETas model can be divided into 91 consecutive 7 day windows (the first four windows are shown in Fig. 3a). We then take the 7 day average of the state over each 7 day window (\bar{w}_{7day}). A 1 day mean centred on each 7 day window is also calculated (\bar{w}_{1day}). A corresponding anomaly field for the state centred on each window is then calculated:

$$w' = \bar{w}_{1day} - \bar{w}_{7day}. \quad (6)$$

This is done for each of the 7 day windows from the control run which yields a 91 member ensemble of anomaly fields. This 91 member anomaly field, A , does not evolve during time and is therefore considered “static”, one of the key differences between the EnOI and EnKF algorithms. It should be noted that the shorter of these timescales used to construct the anomaly field is equal to the assimilation/forecast cycle used in the case study. If the assimilation/forecast cycle were run on a weekly cycle or longer, the choice of timescales may need to be reconsidered. A time series of anomalies at four locations indicates the anomaly fields have little or no discernible serial correlation (Fig. 3b) and are therefore considered stationary. We have chosen a 91 member ensemble as this is the number we were able to generate and handle within our computational constraints, and we regard it as a satisfactory size for the

application described in this paper. Lastly, unlike the Bluelink EnOI system (Oke et al., 2008), covariance localisation is not used as we anticipate domain wide correlations.

2.3. Initialisation

The assimilation cycle used in the SETas EnOI system involves running the model twice over a prescribed time period. This two-step procedure is used to avoid rapid changes in the model state (Bloom et al., 1996), which can have serious implications if the assimilating solution is used to drive a coupled biogeochemical model. The SETas model is not reinitialised to the analysis field within the assimilation cycle, but rather it is relaxed towards the analysis field using a prescribed time-varying relaxation constant, $\tau_r(t)$:

$$\frac{\partial w}{\partial t} = \frac{w - w^a}{\tau_r(t)}. \quad (7)$$

$\tau_r(t)$ is a quadratic function with units of time of the form:

$$\tau_r(t) = a(t - t_a)^2 + k, \quad (8)$$

where a is set to 20 days^{-1} , t_a is the analysis time and k is set to 15 min. The constants a and k can be used to vary the strength of the relaxation used. This relaxation procedure is applied to a time window spanning $t = t_a - 1 \text{ day}$ to $t = t_a$. The solution to (7) adds an increment to the state w at each time step, and is considered a form of “nudging”. These increments then form part of the solution to the primitive equations. A comprehensive study by Sandery et al. (2011), investigates the effects of this nudging term on various diagnostic variables and compares different forms of relaxation. We acknowledge that there is a need for further investigation into the choice of initialisation procedures and the impacts of varying

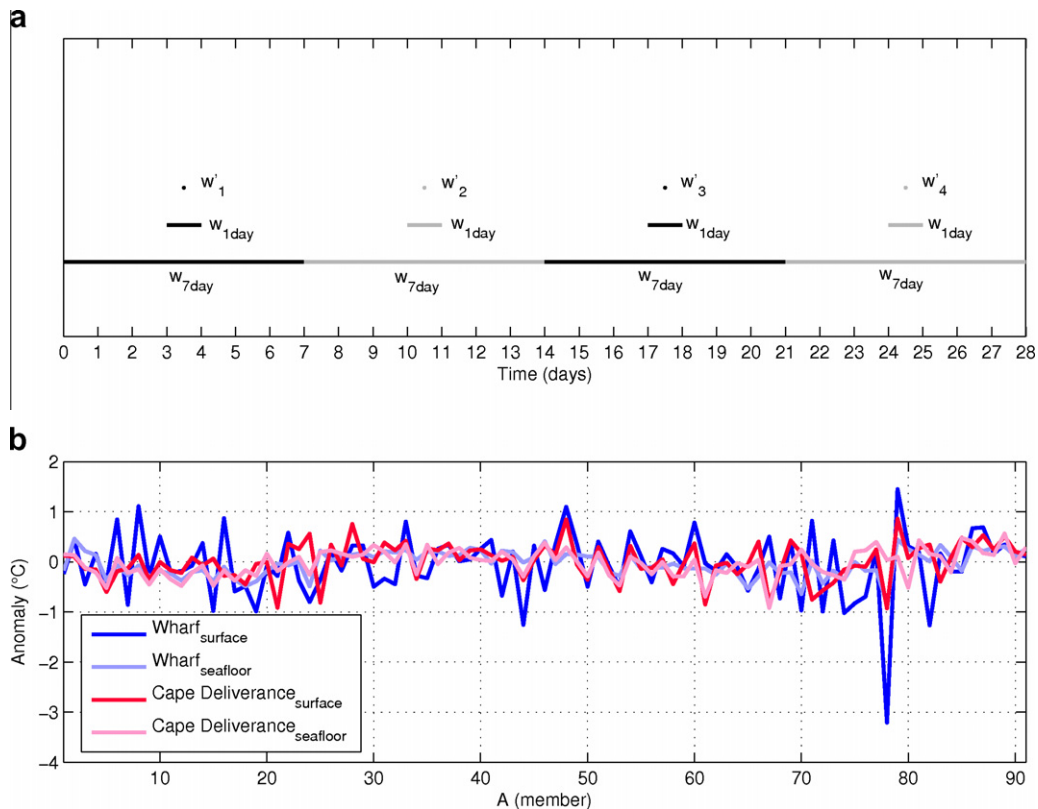


Fig. 3. An outline of (a) the method used to construct the anomaly fields of the state (w') and (b) a plot of the anomaly field at selected locations.

the strength of the relaxation, however this will be dealt with in a further study.

The assimilation cycle (presented graphically in Fig. 4) proceeds as follows:

1. The model is initialised at $t = 0$ with the state for the corresponding time of the control run.
2. A forecast is made for $t + 1$, the model state at $t + 1$ constitutes the model background (w^b).
3. An analysis field is generated by combining w^b with observations (w^o) following Eqs. (1)–(3).
4. The model is then reinitialised with the model state at time t and again integrated forward in time to $t + 1$, but now relaxed towards the analysis field.
5. The time increment is updated, $t \rightarrow t + 1$, and a new forecast is made. Steps 2–4 are repeated until the end of the time period of interest.

The time increment used in this analysis cycle is presently 1 day, which resembles the assimilation cycle that will be implemented within the operational modelling framework detailed in Margvelashvili et al. (2009).

2.4. Observations

Only in situ T and S data from moorings and gliders have been assimilated in this study. During the analysis step of the assimilation cycle, observations that fall within a 30 min window either side of the analysis time are used. Available data from each sensor are averaged in time (and space in the case of a glider) during the observation window, yielding a mean (μ_{obs}) and variance (σ_{obs}^2) for each sensor. This process is analogous to the “super-obing” presented in Oke et al. (2008), with the exception that observations from moorings are not averaged in space. Through the process of “super-obing” we are explicitly accounting for the unresolved sub-grid scale variability that will be present in the observational data, a topic that is thoroughly discussed in Oke and Sakov (2008). Of the available data, only 5% have been assimilated into the model, the rest being withheld for validation.

Data collected from mobile platforms (e.g. gliders) are handled in a slightly different way to the mooring data. When glider data are available, data points are extracted from the record in a similar way to the mooring data (e.g. data within a 30 min window of the analysis time). We have chosen to aggregate all of the glider data into a single horizontal position, that is the mean longitude and latitude of the record. The data are then “binned” into vertical levels that correspond to the vertical grid used in the model. This approach has been taken as spatial correlations are longer in the horizontal than vertical. Data are then averaged in each of these “bins” yielding a mean (μ_{obs}) and variance (σ_{obs}^2) for T and S at the centre of each corresponding vertical cell in the model. The variance within each of these vertical cells will capture the horizontal variance associated with the horizontal aggregation of the glider observations.

The model is interpolated onto the observation locations using the observation operator H . The observation error covariance ma-

trix, R , is diagonal, which assumes that observation errors are independent of each other. For each observation, the corresponding diagonal element is given by the sum of instrument/sensor error (σ_{error}^2) and the variance from the observation record (σ_{obs}^2), $R_{ij} = \sigma_{error}^2 + \sigma_{obs}^2$. In this case, σ_{error}^2 is $0.1 \text{ } ^\circ\text{C}^2$ for temperature and 0.05 PSU^2 for salinity. By explicitly including σ_{obs}^2 , we are accounting for the representation error (Oke and Sakov, 2008), which includes errors associated with advection, horizontal movement of the sensor and small-scale spatial variability.

3. Case study

3.1. Control run

The control run of the model is used to form the anomaly fields (Eq. (6)) and as a benchmark with which to compare the assimilating run. Comparisons between the control run and glider observations suggest that there is a $1 \text{ } ^\circ\text{C}$ over estimation of the temperature field in central storm bay in April (Fig. 8) and that this reduces during the onset of winter conditions. The salinity structure lacks the fine scale structure seen in the observations, however it closely resembles the observations to within 0.2 PSU. The exception is the inshore region of Storm Bay, which does not place the river plume in the correct location, hence it appears to be missing in Fig. 9.

3.2. Correlation structures

Using the ensemble-based background error covariances (BEC's) it is possible to derive the correlation structures from point measurements taken within the model domain. These spatial correlation structures give an indication of the footprint, or region of influence, of observations for each location. An observation of surface temperature at the CSIRO wharf results in a correlation structure (Fig. 5a) that suggests this observation carries information about the surface temperature both up- and down-stream of this point. A surface observation at the mouth of the Derwent Estuary at Cape Deliverance (Fig. 5b) yields information not only on the surrounding waters, but also for much of the D'Entrecasteaux Channel and nearby shallow Coastal Locations. A surface observation in central Storm Bay (Fig. 5c) contains information about the surrounding water to the base of the mixed layer (not shown). These are the typical structures that are contained in the BEC's that are implicit to the ensemble and are used to spread the observational data to regions of the domain where no observations exist.

The correlation structure depicted in Fig. 5a, suggests that two spatially separate river systems (the Derwent and Huon) share a reasonable degree of cross-correlation with values ranging between 0.5 and 0.7 for temperature and similar values for the salinity (not shown) fields. Intuitively, this feature makes sense because these river systems are affected by the same synoptic weather patterns, which strongly influence the baroclinic dynamics (Herzfeld et al., 2010).

A temperature observation taken in the surface layer of the model at Cape Deliverance (Fig. 5b), has a footprint that encompasses most of the sheltered inshore waters of the SETas domain. Again, most of these regions are influenced by the same synoptic weather system. An interesting phenomenon is the absence of any correlation adjacent to the southern capes on the Tasman Peninsula and Bruny Island (denoted by areas of purple shading in Fig. 5a and b). These areas have a correlation very close to 0, which are not statistically significant (discussed below). The regions surrounding the aforementioned capes are prone to relatively high current speeds due to the interaction of the tide with the regional circulation discussed earlier. As a result, wake-like formations may occur in the lee of these features, which could act to destroy any

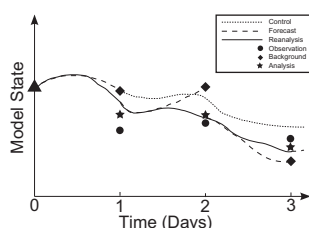


Fig. 4. A schematic outlining the SETas EnOI assimilation cycle.

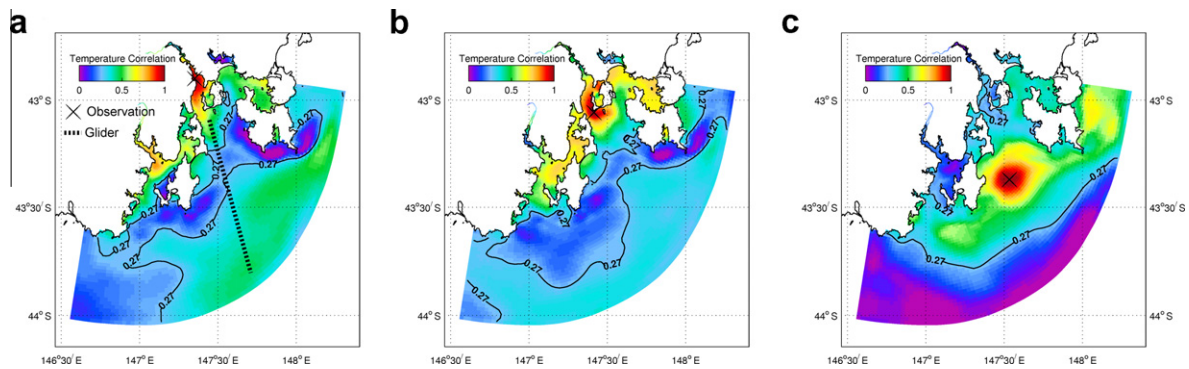


Fig. 5. An example of the ensemble-based cross-correlation structure for a temperature observation in the surface cell for (a) the CSIRO Wharf, (b) Cape Deliverance and (c) central Storm Bay. The statistically significant 0.27 correlation contour is shown in all three panels. Black crosses indicate the position of the observations and the dashed black line denotes the approximate path of the repeat Slocum glider transect.

correlation on the timescales that we are targeting with a static BEC. Based on a t -test with 89 degrees of freedom (recalling we have 91 ensemble members, $N - 2$ degrees of freedom), and a p -value of 0.01, any correlation with an absolute magnitude of below 0.27 is not statistically significant.

The observation in central Storm Bay (Fig. 5c) contains some information about these energetic areas, but the cross-correlation

is still weak. None of the observations in near shore regions or central Storm Bay yield any strong correlation signal at the boundary. Accordingly, we cannot expect coastal in situ moorings to correct for any biases associated with open ocean boundary conditions. Alternative sources of data will need to be considered, e.g. coastal gliders, remotely sensed SST, or if and when it is available, a remotely-sensed coastal sea level anomaly product.

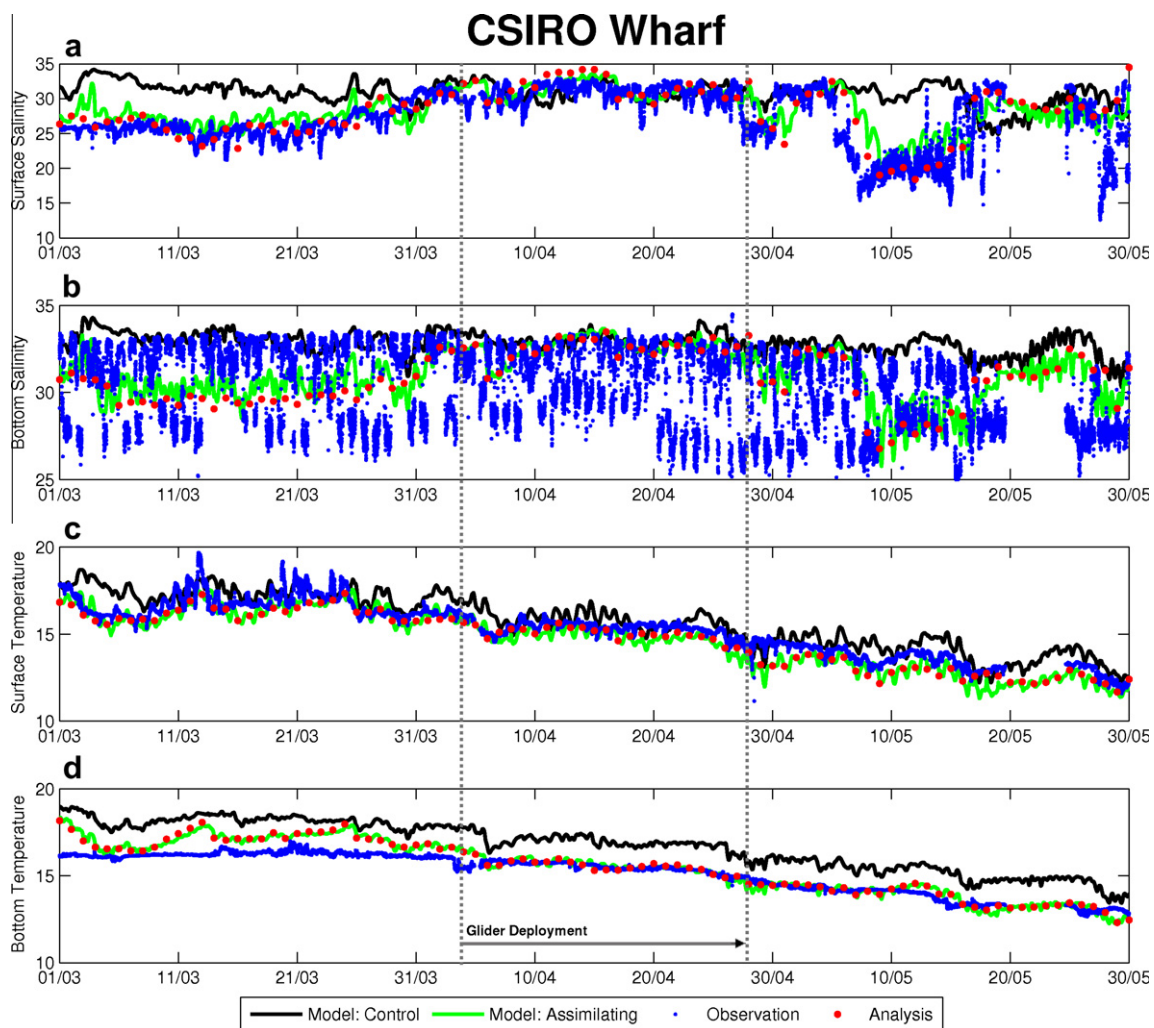


Fig. 6. An in situ time series of temperature and salinity collected at the CSIRO Wharf in Hobart. Model control run (black line), model assimilating run (green line), observations (blue dots), analyses (red dots). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The cross-correlation structures have plausible domain wide information. Also, the ensemble-based correlation fields are not particularly noisy, so we conclude that sampling error is not a major issue for this application. For these reasons covariance localisation has not been implemented. In any case, there are so few independent observations available that the rank of the ensemble typically exceeds the number of observations.

3.3. SETas case study

The SETas EnOI system was tested in the south-east Tasmanian (SETas) region using data collected between the 1st of March and 31st May 2009. During this period there were two real-time moorings located in coastal regions of the model domain (Fig. 1):

1. CSIRO Wharf (147.3376°E, 42.8853°S), *T* and *S* at surface (−2.0 m) and bottom (−10.5 m),
2. Cape Deliverance (147.4070°E, 43.0467°S), *T* and *S* at surface (−2.0 m) and bottom (−10.0 m).

A Slocum glider was deployed in Storm Bay for 25 days during April 2009. It traveled 450 km, dived to within 10 m of the bottom (50–160 m) and completed over 2000 casts in total (see Fig. 5 for approximate transect line). Data collected for assimilation were conductivity, temperature and depth.

3.3.1. Moored in situ time-series comparison

Moored in situ observations of *T* and *S* at two locations were assimilated into the model. The control run (no assimilation) is typically too warm and too salty throughout the water column (Figs. 6 and 7), and lacks the synoptic scale variability seen in the observations. This is highlighted by the inability of the control run to capture the freshwater flow event seen between the 6th and 22nd of May (Fig. 6a). The salinity within the Derwent Estuary is also not adequately captured in the control run, as there is no evidence of the tidal advection of a salinity front at the sensor as seen in the observations (Fig. 6a), largely due to the poor resolution within the Derwent River. The control run is able to capture the magnitude of the diurnal heating/cooling cycle in the surface temperature record (Fig. 6c), but there is some bias with the model typically too warm. The control run displayed similar behaviour at the mouth of the Derwent Estuary (Cape Deliverance, Fig. 7), and is again too warm and too salty. Both locations capture the seasonal cooling trend which is caused by the transition from autumn to winter conditions.

The SETas EnOI scheme is able to correct for the bias seen in the control run. While the assimilation scheme can correct for the synoptic scale freshwater fluxes, seen in the improvement in the salinity record between the 6th and 22nd of May, it is not able to reproduce the advection of the salinity front at the observation location over a tidal cycle. Rather, it is able to track the observa-

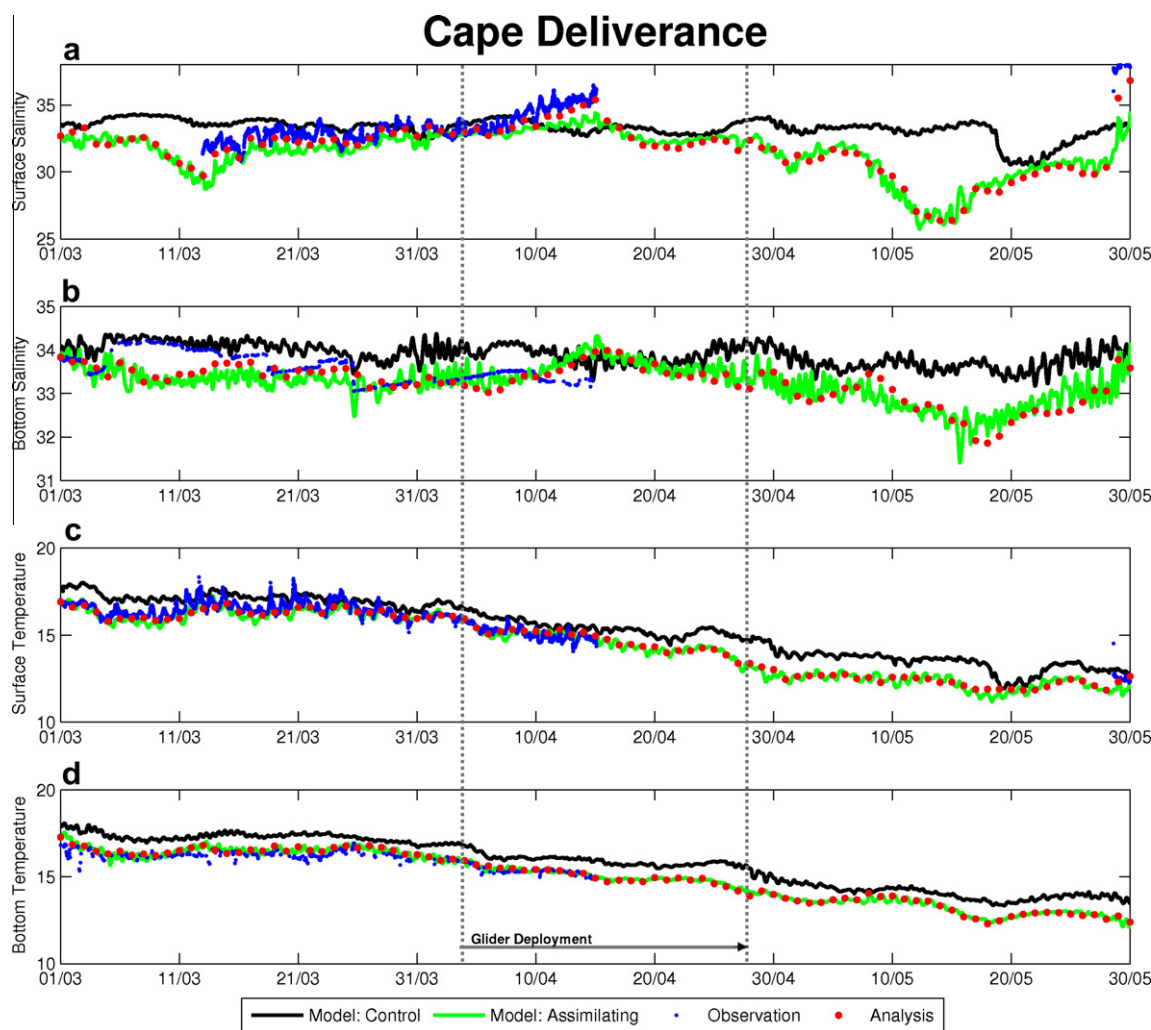


Fig. 7. An in situ time series of temperature and salinity collected at Cape Deliverance which is located at the mouth of the Derwent Estuary, downstream of Hobart. Model control run (black line), model assimilating run (green line), observations (blue dots), analyses (red dots). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

tions that are assimilated into the model during the assimilation window. This is discussed in more detail in Section 4. The modelled surface temperature captures the diurnal heating/cooling cycle, albeit at a reduced amplitude. This reduction in amplitude is an artifact of the relaxation scheme used to reconcile the model state with the analysis. The modelled bottom temperature has two distinct behaviours corresponding to the time period during the first 35 and the last 60 days. During the first period there is bias towards the over-prediction of the bottom temperature (Fig. 7d and Fig. 6d). However, during the second period, this bias is removed and the modelled bottom temperature tracks the observations well. The removal of this bias on the 5th of April, corresponds to the time when Glider data from Storm Bay are available for assimilation.

3.3.2. Mobile in situ profile comparison

The deployment of the Slocum Coastal Glider in Storm Bay, yielded a very high resolution (in both space and time) dataset. Observations from the Glider have identified four main watermasses:

1. A warm salty water mass that resides in the surface layer in the offshore region. This is likely to be derived from the extension of the EAC and/or the extension of the SAC (Schiller et al., 2008).
2. A cold and salty water mass that lies below the water mass previously described.

3. Derwent Estuary water that is typically fresh ($S < 30$) and has a temperature signature that is close to that of the overlying air. This water mass is often seen as a plume that exits the mouth of the Derwent Estuary at Cape Deliverance and flows into Storm Bay.
4. Waters residing in Storm Bay are a mixture of these three source water masses, that are mixed together with the properties slightly modified through evaporation, heating and cooling.

In periods of calm weather there is some stratification of the water column in central Storm Bay, whereas, during the passage of vigorous synoptic weather systems, the Bay becomes well mixed. When the control run is compared with the observed Glider data, the temperature structure in the Bay is too warm, by up to 3° (Fig. 8; centre panel), the freshwater plume from the Derwent Estuary is absent from the profiles (Fig. 9; centre panel), and the overall T and S structure of the Bay lacks both the fine scale detail seen in the observations and also the general placement of the dominant structures. The prescribed open ocean boundary conditions strongly influence the internal dynamics of coastal models. We suspect that this explains the inconsistencies between the observations and the control run.

The assimilating run of the model has reduced the temperature Root Mean Square Difference (RMSD) between the glider and model by 90% (Fig. 10), most of which is a reduction in bias. The dominant T and S structures are now in the correct place, the

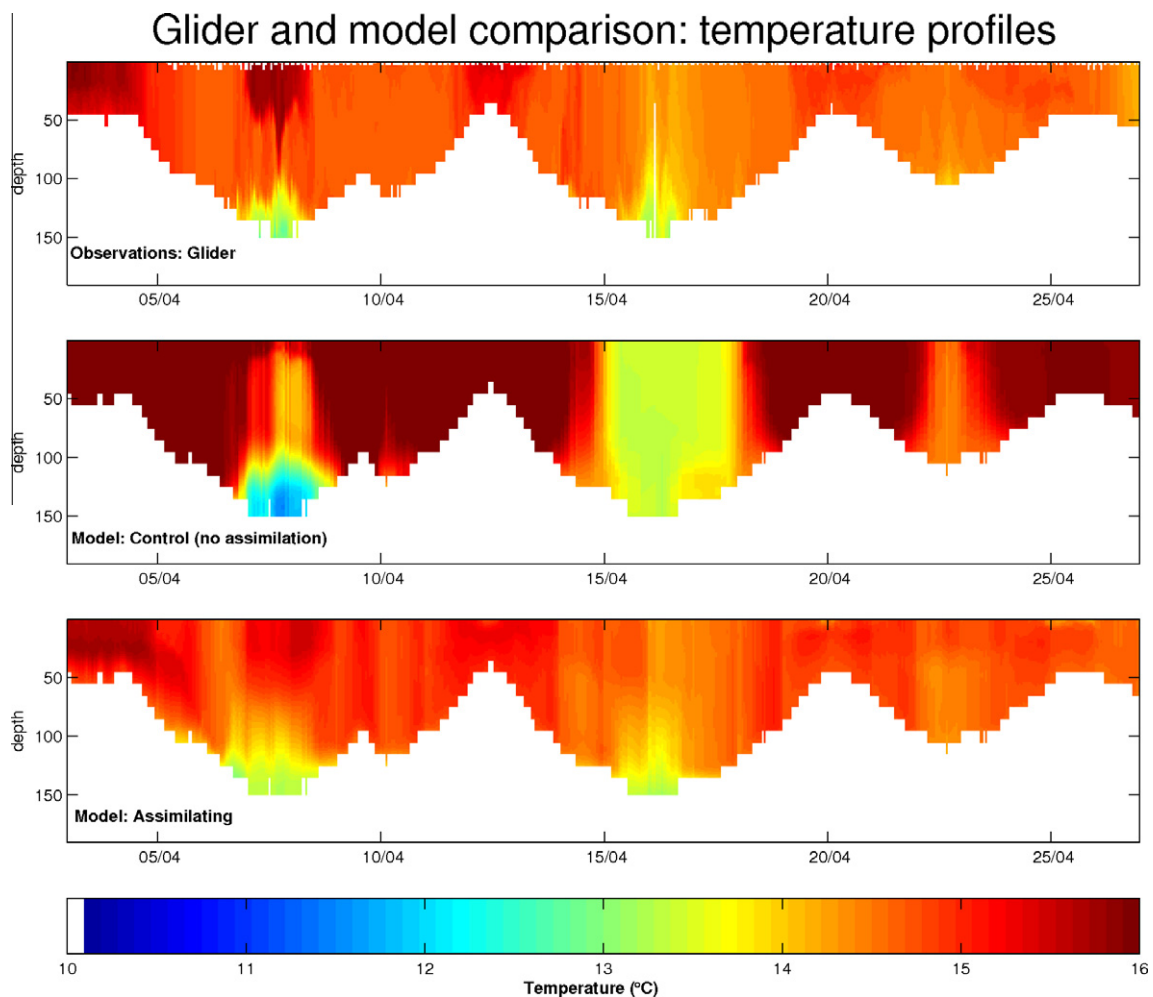


Fig. 8. Temperature profiles along a repeat glider transect through central Storm Bay (Fig. 1). Observed profile (top panel), the control model run (central panel) and the assimilating model run (lower panel).

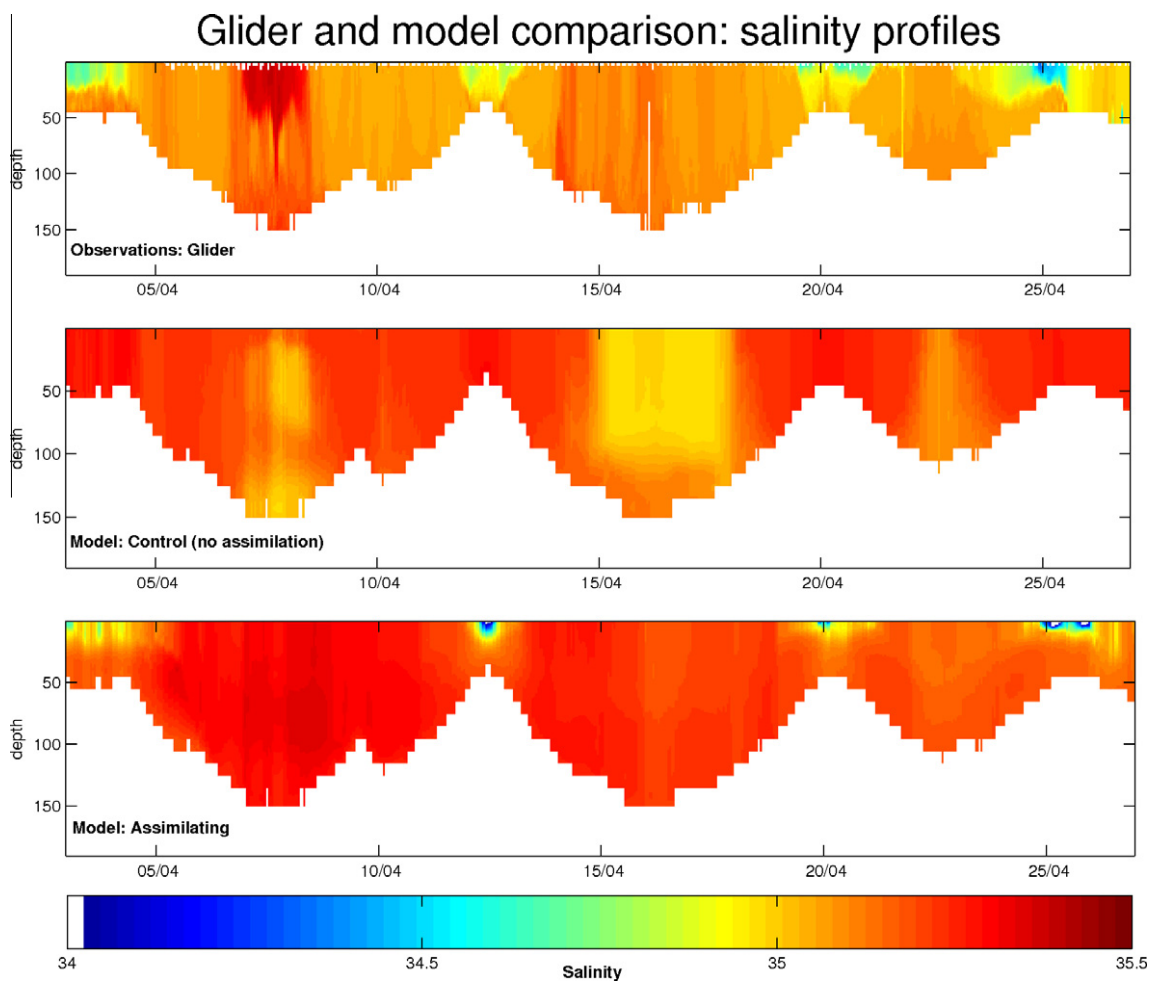


Fig. 9. Salinity profiles along a repeat glider transect through central Storm Bay (Fig. 1). Observed profile (top panel), the control model run (central panel) and the assimilating model run (lower panel).

temperature bias has been corrected and more importantly the vertical structure of the Bay is more consistent with observations. Additionally, there is now more fine scale structure within the assimilating run of the model. There is a more modest improvement in the salinity structure, the Derwent Estuary plume is now resolved which is likely due to the control run having less bias in the salinity record than seen in the temperature record.

3.4. Comparison with remotely-sensed SST data

An independent validation of the SETas EnOI scheme was undertaken using data obtained from the Group for High-Resolution Sea Surface Temperature (GHRSSST) level 2 dataset (Donlon et al., 2009). During the 3 month reanalysis period there were 3276 pixels that fell within the model domain which had an error of less than 0.3°C . A 3 month average RMSD was calculated for each model cell that had a minimum of 10 GHRSSST observations within a 1000 m radius of the cell centre. The spatial maps of the RMSD for SST in the control and assimilating runs of the model are shown in Fig. 11. The control run exhibits a strong cross-shore gradient in RMSD, with typical inshore values exceeding 1.4°C , reducing to 0.4°C near the offshore boundary. The RMSD in the assimilating model run displays a more complex distribution. In regions where observations carried substantial influence, e.g. central Storm Bay, the RMSD was reduced by between 40% and 90%, but there was little reduction along the offshore boundary. However, there was a substantial reduction in the RMSD to the south

and south-west of Bruny Island. Based on the reported circulation patterns, this area is “upstream” of Storm Bay, therefore observations from the Glider transect are being spread to the area, which result in the reduction of the mean RMSD. The domain wide average reduction in RMSD was 40% over the 3 month reanalysis period.

4. Discussion

Sequential data assimilation algorithms (including the EnKF and EnOI) are derived based on the assumption that the underlying model is unbiased. That is to say that for a sequential system to be optimal, the model should be unbiased. It is true that in this particular application the control run is biased in the temperature field, but so far as the assimilation algorithm is concerned, there only exists bias in the initial conditions. However, what is important is the degree of bias entering the model over an analysis cycle, i.e. Is the background biased? A time-series of spatially averaged increments for the temperature and salinity fields show that the algorithm is consistently removing heat from the model until day 7 (Fig. 12). Once this initial bias in the temperature field has been removed, the spatially averaged temperature increment fluctuates about a mean value of -0.05°C , with the 95% confidence interval of approximately $\pm 0.5^{\circ}\text{C}$. This means that for half the time, the increments are positive, not negative. If the background field were biased towards predicting warmer than observed temperatures, the spatially averaged increment would be consistently negative.

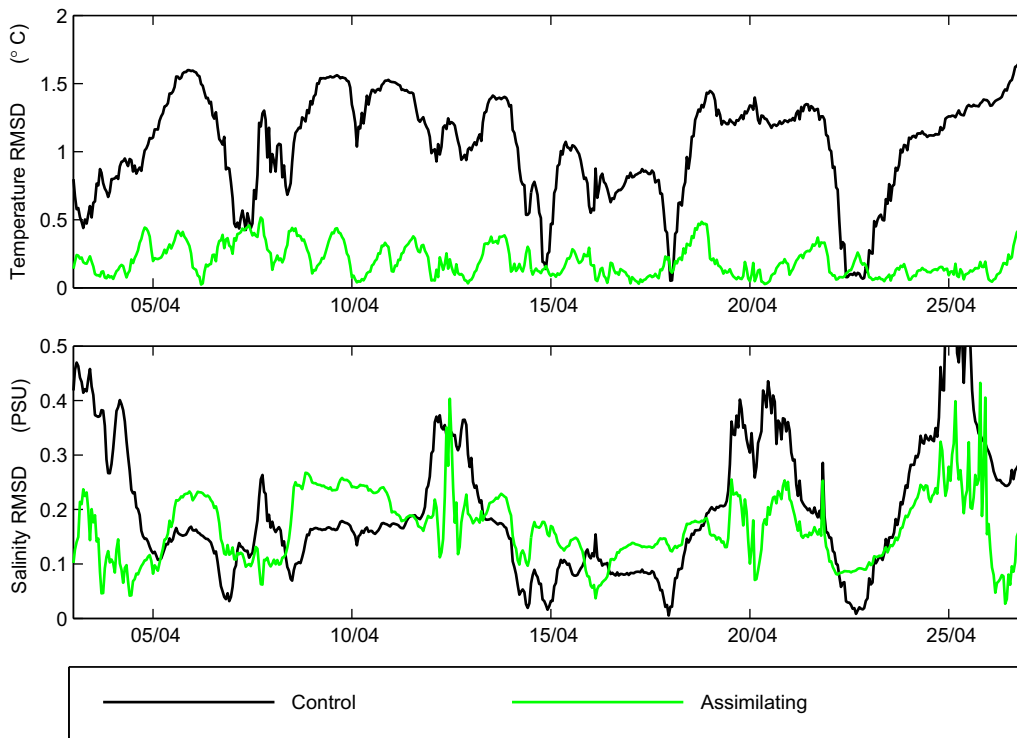


Fig. 10. A time series of the temperature (top panel) and salinity (lower panel) RMSD for each profile presented in Figs. 8 and 9 for the control run (black lines) and assimilating run (green line). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

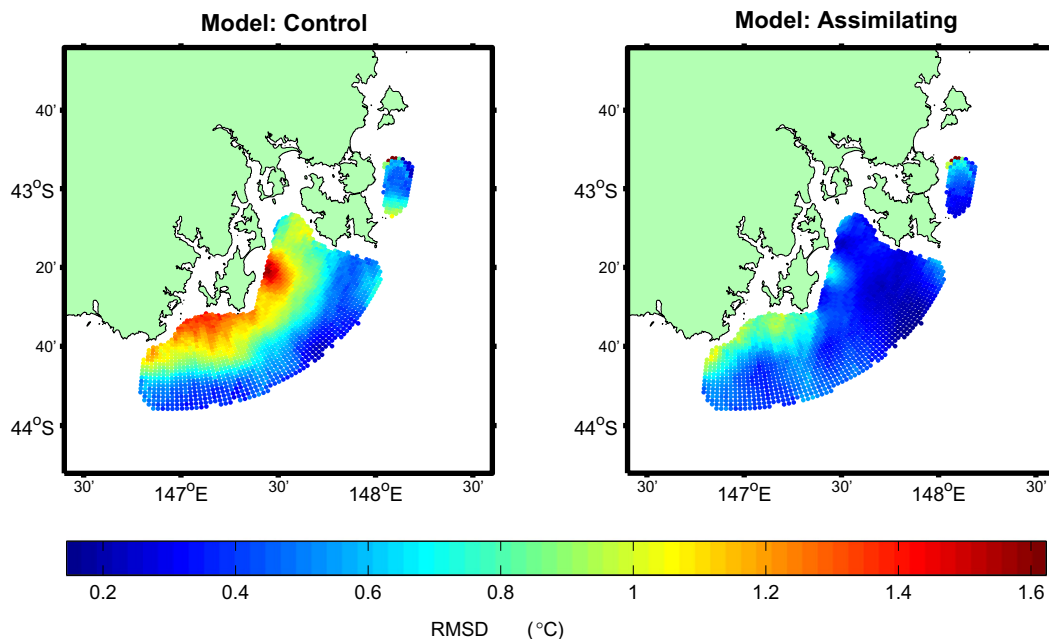


Fig. 11. The 3 month average RMSD between the observed SST and modelled SST from the control run (left) and the assimilating run (right).

Once the excess heat has been removed from the model, bias does not enter the background field over the duration of the assimilation cycle. There is less temporal variation in the spatially averaged salinity increments, with a spatial mean of 0 and a 95% confidence interval extending to ± 0.5 PSU.

The number of assimilation cycles per day was increased from 1 cycle per day to 4 cycles per day. The assimilating model was then initialised on the 1st of April and observations (both glider and

moorings) were assimilated for 15 days. It takes approximately 6–8 assimilation cycles (2 days) to remove the bias when the model is initialised from the control run (Fig. 13a and b), which is consistent with the number of cycles required for the longer assimilation cycle previously presented. Observations were assimilated for a further 13 days, then from this point onwards, no observations were assimilated and the model was integrated forward until it converged with the control run (Fig. 13).

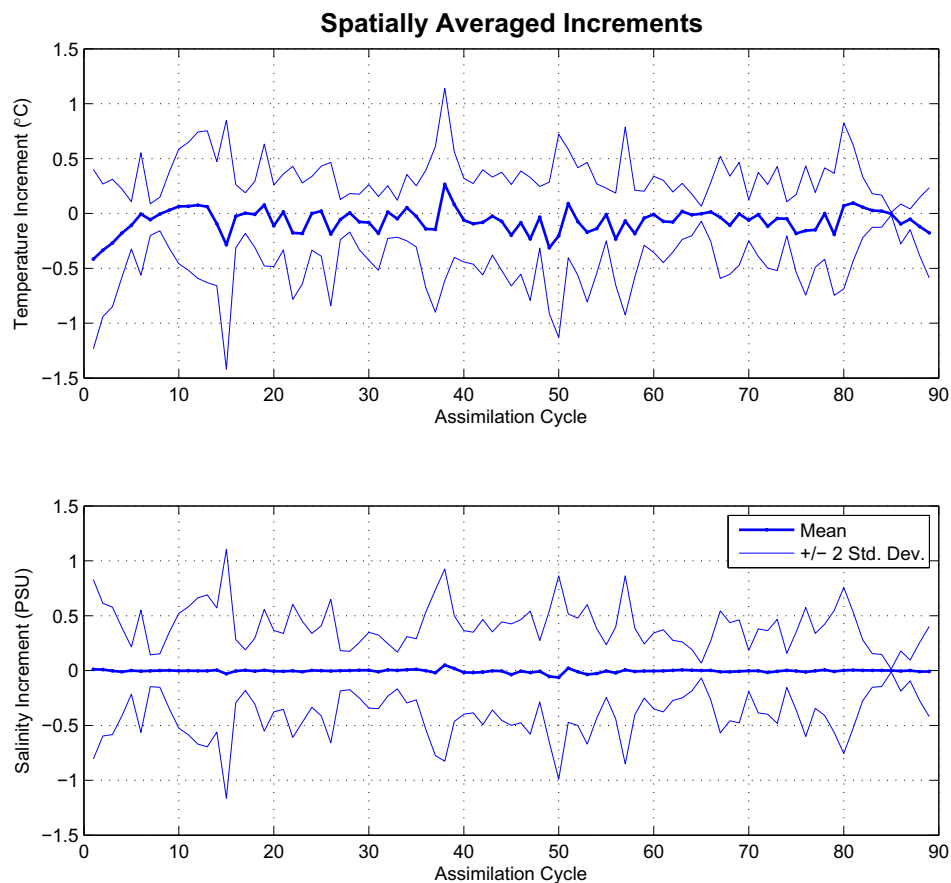


Fig. 12. Spatially averaged temperature (upper panel) and salinity (lower panel) increments and the 95% confidence intervals for the duration of the case-study period.

There are two timescales we need to consider when interpreting Fig. 13:

1. The time taken for the control and assimilating extension to converge;
2. the time taken for the assimilating extension to diverge from the observations.

The extension to the assimilating run takes approximately 75 days to converge with the control run. Both of these models are forced with the same open ocean, head of estuary and surface boundary conditions. The only differences between these runs are those in the initial conditions at day 15. We hypothesise that, due to the advective nature of the region, the time taken for these runs to converge should be approximately equal to the flushing time. This is in fact the case and agrees well with the flushing times published in Herzfeld et al. (2010).

The rate at which the assimilating extension then diverges from the observations indicates the time-scales that errors are introduced into the model. There are a number of events where the assimilating extension diverges from the observation, over periods of 7–14 days. These events are most noticeable in the temperature time series from the CSIRO wharf sensor (seafloor) in Fig. 13. This agrees with our hypothesis outlined in Section 2.2, that errors enter the model at timescales longer than the assimilation cycle used in Section 3.

Fortuitously, the control run converges with the observations approximately 90 days after initialisation. During this period there is a transition from summer to winter conditions. This result suggests that errors in the forcing data and/or model parametrisation, are substantially less during the winter months.

The time taken for the assimilating extension to converge with the control run is the same as the time taken for the control run to converge with the observations. Unfortunately, we do not have a dense dataset that spans the transition from winter to summer conditions and therefore cannot quantify any intra-seasonal variation in these time-scales. However, bearing these time-scales in mind, there are many possible sources of error that could contribute to this slow convergence between the control run and assimilating extension, and also the divergence between the assimilating extension and the observations. We suggest that the following three sources are likely candidates, as they directly influence the heat content of the model:

1. the global ocean model used to force the open ocean boundaries may be biased,
2. the net heat flux could be too high or seasonally biased,
3. the shortwave radiation is being incorrectly distributed throughout the water column.

A comparison between the OceanMAPS and the temperature sensors on the IMOS Maria Island reference station (located near the north-east boundary) shows no persistent bias in the open ocean boundary conditions (not shown). Apart from the GHRSSST data, there are no observations near the south-west boundary. However, the time-averaged RMSD along the southern and western boundary are 0.2 °C which is substantially less than the 1 °C bias seen in the interior of the domain. Furthermore, given OceanMAPS is a data-assimilating ocean model, persistent over estimation of the temperature is unlikely and has not been reported. Without a richer observational dataset to use for model calibration it is difficult to separate the relative contribution of errors in the net heat

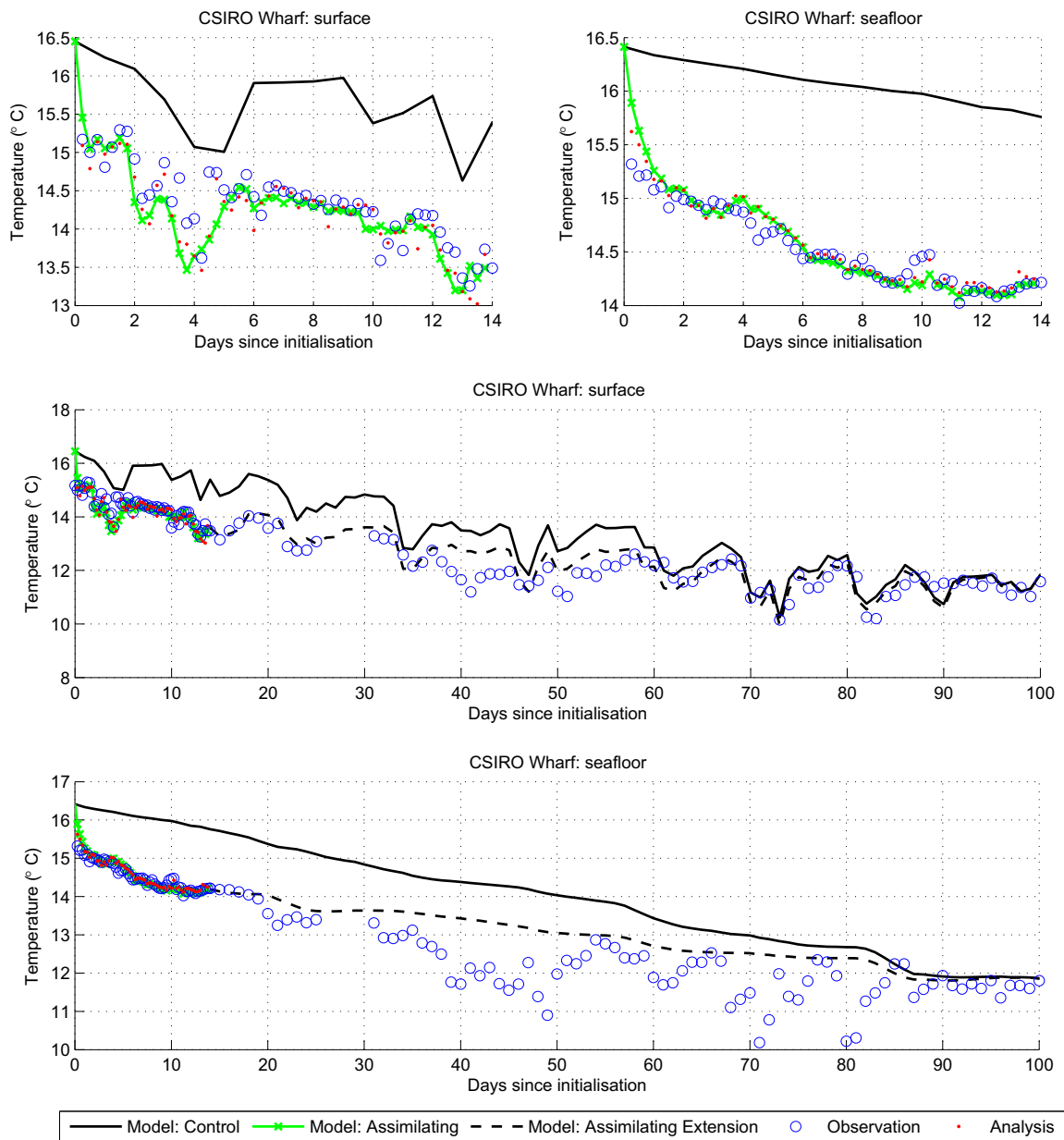


Fig. 13. A comparison between the analysis fields (red dots), the assimilating model (green line), control run (black line) and observations (blue circles) for the surface and seafloor sensors located at the CSIRO wharf. The top left and right panels offer an insight into the time taken to remove the initial bias from the control run. The extension of the assimilating run (dashed black line), where no observations were assimilated, is shown in the lower two panels. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

flux or parametrisation of the distribution of shortwave radiation. There is no evidence that the surface forcings derived from the MesolAPS product are seasonally biased (Vincent et al., 2008). If the bulk heat flux formulation were itself biased, then this would manifest as a persistent rather than seasonal bias. This suggests that the most likely source of the seasonal over-estimation of temperature is the shortwave parametrisation.

The parameters that control the distribution of shortwave radiation through the water column are transmission, absorption in the upper layer and bottom reflectance, all of which are invariant in space and time (Herzfeld and Waring, 2009). In coastal domains that experience periodic influxes of freshwater and have spatially varying bottom types in optically shallow water, the assumption of spatio-temporal invariance in these parameters can lead to errors. Therefore, with little *a priori* knowledge of how these parameters may vary in space and time, they cannot be mechanistically

modelled. For this reason we use the EnOI system to correct for these slowly evolving errors.

5. Conclusion

We have applied an efficient data assimilation scheme suitable for use in operational coastal hydrodynamic models. The use of a static BEC field has highlighted the “long distance” domain wide correlation structures inherent to coastal regions. We suspect that these structures are realistic and do not contain spurious long range correlation as seen in global ocean models (Oke et al., 2008), and exist due to the spatially coherent synoptic forcing through small coastal domains. For this reason, covariance localisation is not employed, and as a result, observations can carry a large “footprint” of influence.

The assimilation of in situ observations from gliders and moorings into the SETas forecast system was very successful. The assimilating model run realistically reproduces the observed variability, including the Derwent River freshwater plume in Storm Bay. We have identified domain wide correlation structures caused by the spatially coherent forcing data. In regions of dense observations, the RMSD between GHRST and the model was reduced by up to 90%, with a more modest 40% reduction in the domain wide RMSD. Furthermore, we have shown that this system can remove bias at initialisation within eight assimilation cycles. Once the initial bias has been removed, bias does not re-enter the model during the forecast period.

We have suggested the dominant source of error in the model relates to the parametrisation of shortwave radiation distribution through the water column. This error will manifest itself in all cases where there is little “*a priori*” knowledge of the optical and seafloor characteristics. However, as these errors are introduced at time scales much longer than the assimilation cycle, the EnOI algorithm is able to correct for them.

Acknowledgments

We would like to thank two anonymous reviewers for making suggestions that have substantially strengthened and improved the manuscript. This project was a collaboration between the Computational and Simulation Sciences: Transformational Capability Platform and INFORMD project funded by the Commonwealth Scientific and Industrial Research Organisation (CSIRO). Mooring data were kindly supplied by the Tasmanian ICT: TasMAN project (<http://www.csiro.au/science/TasMAN.html>), a CSIRO Wealth from Oceans Flagship project. We also wish to thank the Australian Bureau of Meteorology for supplying boundary conditions and the Integrated Marine Observing System (IMOS) for supplying the GHRST data. The authors would also like to acknowledge the help and support of John Andrewartha and Mike Herzfeld, both of CSIRO.

References

- Bertino, L., Lisaeter, K.A., 2008. The TOPAZ monitoring and prediction system for the Atlantic and Arctic Oceans. *Journal Of Operational Oceanography* 1 (2), 15–18.
- Bloom, S., Takacs, L., DaSilva, A., Ledvina, D., 1996. Data assimilation using incremental analysis updates. *Monthly Weather Review* 124 (6), 1256–1271.
- Brasseur, P., Bahurel, P., Bertino, L., Birol, F., Brankart, J.M., Ferry, N., Losa, S., Remy, E., Schroeter, J., Skachko, S., Testut, C.E., Tranchant, B., Van Leeuwen, P.J., Verron, J., 2005. Data assimilation for marine monitoring and prediction: the MERCATOR operational assimilation systems and the MERSEA developments. In: 4th WMO International Symposium on Assimilation of Observations in Meteorology and Oceanography, Prague, Czech Republic, April 18–22, 2005. *Quarterly Journal of the Royal Meteorological Society* 131 (613, Part C), 3561–3582.
- Counillon, F., Bertino, L., 2009. Ensemble optimal interpolation: multivariate properties in the Gulf of Mexico. *Tellus Series A – Dynamic Meteorology and Oceanography* 61, 296–308.
- Cresswell, G., 2000. Currents of the continental shelf and upper slope of Tasmania. *Papers and Proceedings of the Royal Society of Tasmania* 133 (3), 21–30.
- Cummings, J., Bertino, L., Brasseur, P., Fukumori, I., Kamachi, M., Martin, M.J., Mogensen, K., Oke, P., Testut, C.E., Verron, J., Weaver, A., 2009. Ocean Data Assimilation Systems for GODAE. *Oceanography* 22 (3, Sp. Iss. SI), 96–109.
- DiLorenzo, E., Moore, A.M., Arango, H.G., Cornuelle, B.D., Miller, A.J., Powell, B., Chua, B.S., Bennett, A.F., 2007. Weak and strong constraint data assimilation in the inverse Regional Ocean Modeling System (ROMS): development and application for a baroclinic coastal upwelling system. *Ocean Modelling* 16 (3–4), 160–187.
- Dobricic, S., Pinardi, N., Testor, P., Send, U., 2010. Impact of data assimilation of glider observations in the Ionian Sea (Eastern Mediterranean). *Dynamics of Atmospheres and Oceans* 50 (1), 78–92.
- Donlon, C.J., Casey, K.S., Robinson, I.S., Gentemann, C.L., Reynolds, R.W., Barton, I., Arino, O., Stark, J., Rayner, N., Leborgne, P., Poulter, D., Vazquez-Cuervo, J., Armstrong, E., Beggs, H., Llewellyn-Jones, D., Minnett, P.J., Merchant, C.J., Evans, R., 2009. The GODAE high-resolution sea surface temperature pilot project. *Oceanography* 22 (3, Sp. Iss. SI), 34–45.
- Evensen, G., 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research – Oceans* 99 (C5), 10143–10162.
- Herzfeld, M., 2006. An alternative coordinate system for solving finite difference ocean models. *Ocean Modelling* 14 (3–4), 174–196.
- Herzfeld, M., Andrewartha, J., Sakov, P., 2010. Modelling the physical oceanography of the D'Entrecasteaux Channel and the Huon Estuary, south-eastern Tasmania. *Marine and Freshwater Research* 61 (5), 568–586.
- Herzfeld, M., Waring, J.R., 2009. Sparse Hydrodynamic Ocean Code V1.0 User Manual. CSIRO Marine and Atmospheric Research, Hobart.
- Kurapov, A.L., Foley, D., Strub, P.T., Egbert, G.D., Allen, J.S., 2011. Variational assimilation of satellite observations in a coastal ocean model off Oregon. *Journal of Geophysical Research – Oceans*, 116.
- Lermusiaux, P.F.J., Robinson, A.R., 1999. Data assimilation via error subspace statistical estimation. Part I: Theory and schemes. *Monthly Weather Review* 127, 1385–1407.
- Margvelashvili, N., Campbell, E., 2012. Sequential data assimilation in fine-resolution models using error-subspace emulators: Theory and preliminary evaluation. *Journal of Marine Systems* 90 (1).
- Margvelashvili, N., Herzfeld, M., Wild-Allen, K., Andrewartha, J., Rizwi, F., Skerratt, J., 2009. Wealth from the Oceans project: Linking models to sensor networks – an operational modelling platform for INFORMD. Tech. Rep., CSIRO Marine and Atmospheric Research.
- Oke, P.R., Brassington, G.B., Griffin, D.A., Schiller, A., 2008. The BlueLink Ocean Data Assimilation System (BODAS). *Ocean Modelling* 21 (1–2), 46–70.
- Oke, P.R., Brassington, G.B., Griffin, D.A., Schiller, A., 2010. Ocean data assimilation: a case for ensemble optimal interpolation. *Workshop on Ensemble Prediction and Data Assimilation*, Melbourne, Australia, 2009. *Australian Meteorological and Oceanographic Journal* 59 (Sp. Iss. SI), 67–76.
- Oke, P.R., Sakov, P., 2008. Representation error of oceanic observations for data assimilation. *Journal of Atmospheric and Oceanic Technology* 25 (6), 1004–1017.
- Rudnick, D., Davis, R., Eriksen, C., Fratantoni, D., Perry, M., 2004. Underwater gliders for ocean research. *Marine Technology Society Journal* 38 (2), 73–84.
- Sandery, P., Brassington, G., Freeman, J., 2011. Adaptive nonlinear dynamical initialization. *JGR – Oceans*, 116.
- Schiller, A., Oke, P.R., Brassington, G., Entel, M., Fiedler, R., Griffin, D.A., Mansbridge, J.V., 2008. Eddy-resolving ocean circulation in the Asian–Australian region inferred from an ocean reanalysis effort. *Progress in Oceanography* 76 (3), 334–365.
- Schofield, O., Kohut, J., Aragon, D., Creed, L., Graver, J., Haldeman, C., Kerfoot, J., Roarty, H., Jones, C., Webb, D., Glenn, S., 2007. Slocum gliders: robust and ready. *Journal of Field Robotics* 24 (6), 473–485.
- Shulman, I., Rowley, C., Anderson, S., DeRada, S., Kindle, J., Martin, P., Doyle, J., Cummings, J., Ramp, S., Chavez, F., Fratantoni, D., Davis, R., 2009. Impact of glider data assimilation on the Monterey Bay model. *Deep-Sea Research Part II – Topical Studies in Oceanography* 56 (3–5), 188–198.
- van Leeuwen, P.J., 2009. Particle filtering in geophysical systems. *Monthly Weather Review* 137 (12), 4089–4114.
- Vincent, C.L., Bourke, W., Kepert, J.D., Chattopadhyay, M., Ma, Y., Steinle, P.J., Tingwell, C.I.W., 2008. Verification of a high-resolution mesoscale NWP system. *Australian Meteorological Magazine* 57 (3), 213–233.