Ensemble data assimilation for an eddy-resolving ocean model of the Australian region

By P. R. OKE[1][*], A. SCHILLER[1], D. A. GRIFFIN[1] and G. B. BRASSINGTON[2]
[1] *CSIRO Marine and Atmospheric Research, Hobart, Tasmania, Australia*
[2] *Bureau of Meteorology Research Centre, Melbourne, Victoria, Australia*

Summary

The Bluelink Ocean Data Assimilation System (BODAS) is an ensemble optimal interpolation system applied to a global ocean circulation model with 10 km resolution around Australia. BODAS derives estimates of forecast error covariances (FECs) from a stationary 72-member ensemble of intraseasonal model anomalies. The FECs are localised around each observation to reduce the negative effects of sampling error and to increase the rank of the ensemble. The FECs have characteristics that reflect the length-scales and the anisotropy of the ocean circulation in different regions. BODAS assimilates in situ and satellite-derived observations of temperature, salinity and sea-level anomaly. Results from a 13-year ocean reanalysis demonstrate that the reanalysed fields are often in very good agreement with withheld observations, and provide a good synoptic representation of the eddy field around Australia.

Keywords: Ocean prediction

1. Introduction

Bluelink is an Australian partnership between the Commonwealth Scientific and Industrial Research Organisation, the Bureau of Meteorology and the Royal Australian Navy. The primary objective of Bluelink is to develop a forecast system for the mesoscale ocean circulation in the Australian region. To this end, the Ocean Forecasting Australia Model (OFAM; Schiller et al. 2005), a global ocean general circulation model, has been configured and the Bluelink Ocean Data Assimilation System (BODAS) has been developed. BODAS is an ensemble optimal interpolation (EnOI) system, similar to that introduced by Oke et al. (2002) and Evensen (2003). The purpose of this paper is to describe recent developments in ocean data assimilation in the Australian ocean community. This paper includes a description of BODAS, examples of its ensemble-based estimates of FECs and some results from a global application of BODAS.

2. Model

The details of OFAM are described by Schiller et al. (2005). Briefly, OFAM is based on version 4.0 of the Modular Ocean Model (Griffies et al. 2004), with local enhancements including the hybrid mixed layer model described by Chen et al. (1994), an improved parameterisation for the penetration of solar radiation and vector optimisation for an NEC SX6. OFAM is intended to be used for reanalyses and short-range prediction. The horizontal grid has 1191 and 968 points in the zonal and meridional directions respectively; with 0.1° horizontal resolution around Australia, between 90°E and 180°, and between Antarctica and 16°N. Outside of this domain, the horizontal resolution decreases to 2° in the North Atlantic Ocean. OFAM has 47 vertical levels, with 10 m resolution down to 200 m depth. The topography for OFAM is a composite of a range of different

topography sources. Horizontal diffusion is modelled using isopycnal mixing and the parameterisation introduced by Gent and McWilliams (1990). Horizontal viscosity is resolution and state-dependent according to the Smagorinsky viscosity scheme described by Griffies and Hallberg (2000).

OFAM is initialised with a blend of climatologies from CARS2000 (CSIRO Atlas of Regional Seas; Ridgway et al. 2002) and Levitus (2001); and is forced at the surface using 6-hourly fluxes of momentum, heat and freshwater from ERA-40 (Kallberg et al. 2004) for 1992 to mid-2002; and using ECMWF 6-hourly forecasts from mid-2002 to 2004. A flux correction is also applied, restoring the sea surface temperature (SST) to a blend of Reynolds-SST (Reynolds and Smith 1994) and high-resolution satellite-derived observations over 30 days; and restoring surface salinity to monthly climatologies (Levitus 2001) over 30 days. OFAM is spun up without data assimilation for a 9 years simulation for the period 1994 to 2002. An analysis of the modelled trends of globally averaged sea-level indicates that the model reaches a state of quasi-equilibrium after 3 years of integration. Interior temperature and salinity properties require much longer to properly equilibrate. An important activity in the Bluelink project, is the Bluelink ReANalysis (BRAN). BRAN is a 13-year run with data assimilation covering the period 1992-2004. Observations that are assimilated include sea-level anomalies (SLA) from all altimeters (ERS 1 and 2, Topex/Poseidon, Geosat Follow-On, Jason and Envisat) and from a coastal tide gauge array around Australia; and temperature and salinity profiles from a range of field surveys including Argo (e.g., Argo Science Team 1998), the TAO array (McPhaden et al. 1998), XBTs and field surveys (e.g., WOCE). BRAN is the first comprehensive ocean reanalysis that is eddy-resolving around Australia.

3. Assimilation System

(a) Analysis Equations

Analyses of sea surface height η, temperature T, salinity S, and horizontal currents (u, v), are computed by solving the analysis equations,

$$\mathbf{w}^a = \mathbf{w}^f + \mathbf{K} \left(\mathbf{w}^o - \mathbf{H}\mathbf{w}^f \right) \tag{1}$$

$$\mathbf{K} = (\mathbf{C} \circ \mathbf{P}^f)\mathbf{H}^T \left(\mathbf{H}(\mathbf{C} \circ \mathbf{P}^f)\mathbf{H}^T + \mathbf{R} \right)^{-1}, \tag{2}$$

where $\qquad \mathbf{w} = [\eta \ \ T \ \ S \ \ u \ \ v]^T \tag{3}$

is the state vector; superscripts a, f and o denote analysis, forecast and observed respectively; $\mathbf{K}$ is the gain matrix; $\mathbf{C}$ is a correlation matrix; $\mathbf{H}$ is an operator that interpolates from the model grid to observation locations; $\mathbf{R}$ is the observation error covariance matrix; $\mathbf{P}^f$ is the FEC matrix; and the open circles denote a Schur, or Hadamard, product (an element by element matrix multiplication). This formulation of the analysis equations is the same as that presented by Houtekamer and Mitchell (2002).

The observation error covariances are assumed to be uncorrelated in space and time. Therefore $\mathbf{R}$ is a diagonal matrix. An estimate for the standard deviation of the error for each observation is given by

$$\epsilon^o = \sqrt{\epsilon_{instr}^2 + \epsilon_{age}^2} \tag{4}$$

where ϵ_{instr} is the standard deviation of the instrument error and

$$\epsilon_{age} = \beta \times RMS_{mod} \left(1 - e^{-0.5|t^a - t^o|/t_{ef}}\right), \tag{5}$$

is the standard deviation of the error associated with the age of an observation; β is a scalar (typically 3-5); RMS_{mod} is the spatially dependent root-mean-square of the model fields about a seasonal cycle during a non-assimilating model run; t^a is the analysis time; t^o is the time of the observation; and t_{ef} is an e-folding time scale (typically 3 days). Therefore, if an observation is made at the analysis time, $\epsilon_{age} = 0$; and as $|t^a - t^o|$ increases, ϵ_{age} approaches $\beta \times RMS_{mod}$, so that the influence of the observation on the analysis decreases.

Estimates of the FECs in (2) are given by

$$\mathbf{P}^f = \mathbf{A}\mathbf{A}^T/(n - 1), \tag{6}$$

where n is the ensemble size and

$$\mathbf{A} = \alpha \left[\mathbf{w}'_1 \quad \mathbf{w}'_2 \quad \cdots \quad \mathbf{w}'_n\right] \tag{7}$$

where α is a scalar that can be used to tune the magnitude of the covariances for a particular application; $\mathbf{w}'_i$ is the ith intraseasonal anomaly derived from a model run with no data assimilation. Each anomaly field consists of all prognostic model variables included in (3). These anomalies are calculated by removing a 91 day running mean from a daily mean field of a model run with no data assimilation. BODAS uses an ensemble size of 72, with one anomaly from every month of the last 6 years of a 9-year model run with no data assimilation. The parameter α has not been tuned for this application. A value of $\alpha = 0$, implies that the forecasts errors are zero, in which case, $\mathbf{K}$ and therefore the analysis increments are exactly zero. A value of $\alpha = 1$, implies that the forecast errors are the same size as the model's intraseasonal anomalies. For a model with some skill, this would be an over estimate. With these limits in mind, we regard $\alpha = 0.5$ as a reasonable starting point.

The FECs in (6) are localised in the horizontal around each observation in (2) using the correlation matrix $\mathbf{C}$. Elements of $\mathbf{C}$ are defined by the quasi-Gaussian function of Gaspari and Cohn (1992), after Houtekamer and Mitchell (2002). Localisation has been shown to reduce the effects of sampling error for applications of an ensemble Kalman Filter (e.g., Hamill et al. 2001). The localising correlation function in $\mathbf{C}$ forces the FECs to reduce to exactly zero, over L° from an observation location. The present implementation of BODAS uses a uniform radial distance L=8°. This has several ramifications for the system's performance. Firstly, the rank of the estimated FECs in $\mathbf{P}^f$ is increased significantly. Using an ensemble size of n, the rank of $\mathbf{P}^f$ is at most $n - 1$. By contrast, with localisation, there are over $XY/(2L)^2$ independent regions in the model domain, where X and Y are the zonal and meridional dimensions of the model. So, applying a 72-member ensemble to a global model ($X = 360$, $Y = 180$) and localising with $L = 8°$, the effective rank of $\mathbf{C} \circ \mathbf{P}^f$ could be $\mathcal{O}(10^4)$. This enables the assimilation system to determine analysis increments that fit the forecast innovations well. There are however, a few drawbacks to localisation. For example, analyses are not as dynamically balanced as they would be without localisation (Mitchell et al. 2002). Also, the inversion of the innovation covariance matrix becomes very expensive, since the techniques for computational efficiency described by Evensen (2003) aren't suitable when localisation is used. This makes the practical implementation of BODAS a challenge.

(*b*)   *Practical Implementation*

Throughout BRAN, observations are assimilated once every 3 days, using a time window of ± 5 days for η and ± 3 days for T and S observations. Therefore, there are typically over 10^6 individual observations available for assimilation in any application of BODAS to OFAM. Furthermore, the dimension of the model grid is approximately 2×10^8. These large dimensions, coupled with the increased rank of the analysis equations due to the localisation, makes the explicit solution to (1-2) very expensive and inaccurate. These challenges are overcome by preprocessing to reduce the number of observations that are assimilated to about 10^5. This is achieved by selecting only the most complete temperature and salinity profiles, assuming similar levels of quality, in regions where there are redundant profiles; and by computing super-observations for the altimeter observations so that the distribution of super-observations reflects the model grid, with more (less) observations in the high-resolution (low-resolution) part of the domain. Using standard error propagation techniques, an updated error estimate for the super-observations are calculated and used in **R**. In addition to these savings, the dimension of the model state is reduced by a factor of 4, by only computing an analysis on every 2*nd* horizontal grid point, before interpolating to the full grid.

For a typical application to OFAM, the model grid is divided into 40 zonal bands. The analysis in each zonal band is computed independently, on a separate processor. Each zonal band has a halo of 8° to the north and south, from which an observation can influence the analysis. The size of the halo is chosen to be consistent with the localising length-scale L. This ensures that adjacent domains are entirely independent and that analyses in adjacent domains are seamless. Within each zonal band, the domain is further divided into 10 sub-domains, each with an 8° halo in all directions. An analysis is computed for each sub-domain independently. Using 40 processors, on an NEC SX6 machine, an analysis is computed for all model variables at all grid points in 20-30 minutes of wall-clock time, depending on the number of observations assimilated.

4.   RESULTS

(*a*)   *Covariances*

Two examples of normalised FECs (i.e., correlations) between sea-level from a reference location and sea-level in the surrounding region for different coastal regions are shown in Fig. 1. This figure shows the original correlations, without localisation, and the localised correlations. The localised correlations show the field of influence of an observation of sea-level during the assimilation process. The example off the coast Java (Fig. 1a-b) demonstrates that the ensemble-based covariances reflect the length-scales and the anisotropy of the ocean circulation near the coast; with longer-decorrelation length-scales in the along-shore direction and shorter decorrelation length-scales in the across-shore direction. In this region of the ocean, the dominant currents are wind-driven, and are in the along-shore direction. Coastal trapped waves also typically propagate to the east off Java, representing a coastal extension of the tropical wave guide. This example also demonstrates that the ensemble-based statistics do not artificially extrapolate information from one ocean basin, across land barriers into an adjacent ocean basin, unlike most standard optimal interpolation systems that use Gaussian correlation statistics. Figure 1c-d show an example off western Tasmania. In this region, the Zeehan Current flows southeastward along the continental shelf edge of
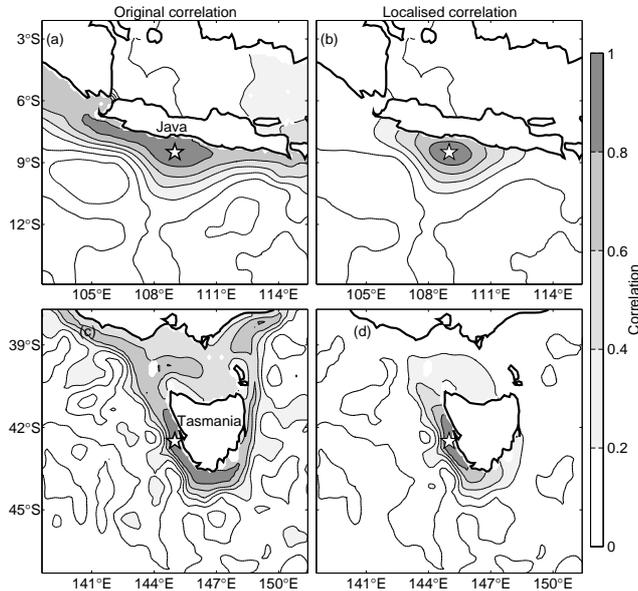
Figure 1.   The original (left) and localised (right) ensemble-based correlation field between sea-level at a reference location, denoted by the star, and sea-level in the surrounding region for an example off the coast of Java (top) and off Western Tasmania (bottom).

western Bass Strait and western Tasmania. In winter the Zeehan Current rounds southern Tasmania and proceeds as far north as about 42°S (Cresswell 2000). These characteristics are reflected in the original correlation field, showing highly correlated sea-level aligned with the continental shelf around southern Tasmania. For both of these examples, the localised correlations have somewhat shorter length-scales than the original. Also, much of the noise in the correlations away from the observation location is much less after localisation.

(b)   Reanalyses

One aspect of ocean reanalysis that remains unclear is whether the integrated global ocean observing system, combined with a relatively simple data assimilation system can constrain the mesoscale features of a high-resolution ocean general circulation model. To start to address this question, we provide a series of comparisons between analyses from BRAN and surface drifter paths from observations obtained from the WOCE Surface Velocity Program. For these comparisons, we focus on three very different regions around Australia. Namely, the Tasman Sea, the most energetic region around Australia; a region off the coast of Java, where the circulation is dominated by upwelling and variability associated with the Indonesian Throughflow; and off the coast of Western Australia, where the Leeuwin Current typically sheds a very complex field of mesoscale eddies (Feng et al. 2005). The time periods we have chosen here correspond to periods when several surface drifters remained in the corresponding regions for an extended period; and when their paths highlight some of the well known phenomena of each region. None of the drifter observations are assimilated and the modelled SST is only weakly restored to 12-day composites of observed SST.
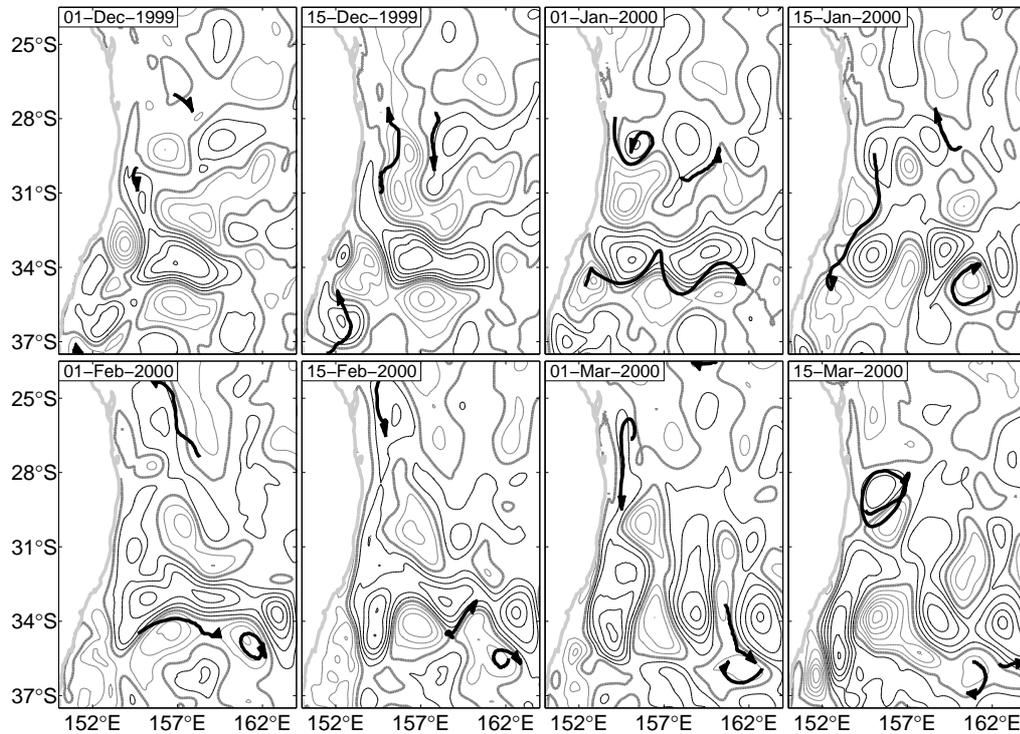
Figure 2. A sequence of maps of analysed SLA (15-day averages) from BRAN and surface drifter tracks for the 15 day period centered at the specified date in early 2000. The contour interval is 0.1 m s^{-1}; negative is half tone; zero is bold half tone.

The comparison in the Tasman Sea is shown in Fig. 2 for December 1999 to March 2000. This is the time of year when the East Australian Current (EAC) is very strong and when its eddy field is typically very rich. Fig. 2 shows a sequence of 15-day averaged SLA, spaced approximately 2 weeks apart, with drifter paths overlaid for the 15-day averaging period for each panel. These comparisons show excellent qualitative agreement between the drifter paths and the analysed SLA. Specifically, this sequence demonstrates that the path of the East Australian Current (EAC), after it separates from the coast ($\approx 31°$S) is well reproduced in BRAN. Further, it provides a good illustration of the variability in the Tasman Sea and shows that even some of the small-scale cold-core eddies are correctly depicted in BRAN. This demonstrates that for this period of BRAN, the eddy-field in the Tasman Sea is well represented by the reanalysis. By contrast, the equivalent SLA fields from the model-only run, with no data assimilation, demonstrate no correspondence between the eddy field and the drifter paths.

A comparison off the coast of Java is shown in Fig. 3, for a period in late-1997. This figure shows a comparison between daily-averaged SST fields from BRAN and 3-day SST composites from AVHRR. Observed drifter paths are overlaid for the 7-day time interval centered on the date of the individual SST maps. The drifter paths seem to correspond well with the SST fields in BRAN. The period covered in Fig. 3 corresponds to a period of strong upwelling off Java. The SST fields in BRAN show narrow filaments of cold, upwelled water penetrating

offshore. This penetration seems to correspond to one of the drifter paths on November 29. Also evident in this sequence of maps, is the warm water from the Indonesian Archipelago flowing through Sunda and Lombok Straits. There is generally good agreement between the modelled and observed SST, although the observed SST doesn't show the same level, or complexity, of upwelling off Java. The lack of a strong upwelling signal in the observed SST is likely to be due to the horizontal advection of warm Indonesian throughflow water that neutralises the surface upwelling signature (Du et al. 2005).

A qualitative assessment of the variability off Western Australia is shown in Fig. 4, for a period in late-1998. This figure shows daily-averaged SST fields from BRAN and 3-day composite SST fields from AVHRR. Observed drifter paths are overlaid for the 15-day time interval centered on the date of the individual SST maps. From these examples, it is clear that at least two specific features of the drifter paths are well represented in BRAN; firstly, the development and persistence of a large warm core eddy at about 29°S; and secondly, the generation of a much smaller cold-core feature at about 114°E. The paths of the other surface drifters in the region also correspond quite nicely with the mesoscale variability in BRAN. There is good correspondence between the observed and modelled mesoscale SST variability in this example, however, there is a warm bias in BRAN.

The qualitative comparisons presented in Fig. 2-4 are for selected periods when the circulation in BRAN corresponds well to the drifter observations. There are periods during BRAN when similar comparisons are not as favourable. However, in general, the agreement between the drifter paths and the circulation represented by BRAN in the high-resolution region of the model (90°E-180°E, south of 16°N) is typically quite good. This is demonstrated, in part, by a statistical comparison of the 6-hourly interpolated velocities of surface drifters and the daily-averaged fields from BRAN. Observations from the surface drifters include SST and the zonal and meridional component of the horizontal currents at 15 m depth (u and v respectively). We compute statistics for the period between the start of 1997 and the end of 2002 for the high-resolution region of the model, a total of 500 drifters. The drifter velocity estimates are the 6-hourly estimates from the GLD DAC krigging process, so we exclude estimates below 0.1 m s^{-1}, which are probably mostly due to data gaps. The comparisons are between daily averaged fields from BRAN and 6-hourly instantaneous observations from drifters. The cross-correlations between these independent observations and the corresponding fields in BRAN are around 0.6, 0.45 and 0.6 for u, v, SST respectively. These correlations are all statistically significant at the 99% confidence level. The root-mean-squared error is 0.24 m s^{-1}, 0.21 m s^{-1} and 1.4°C for u, v and SST respectively; and the mean bias is less than 1 cm s^{-1} for u and v and is 0.85°C for SST, with BRAN producing surface temperatures that are too warm.

Throughout BRAN, there is a warm bias (note the different colour axis in Fig. 3 and 4). This bias results from two deficiencies in the model; firstly, the albedo was inadvertently omitted when the surface fluxes were applied; and secondly, the surface heat fluxes were too warm during the spin-up phase of the run resulting in initial conditions that were too warm. Additionally, we find that the mixed layer depths in BRAN are generally much too deep. Subsequent to the completion of BRAN, the problems with heat fluxes were corrected and the mixed layer depths were improved by adjusting parameters in the mixed layer scheme.
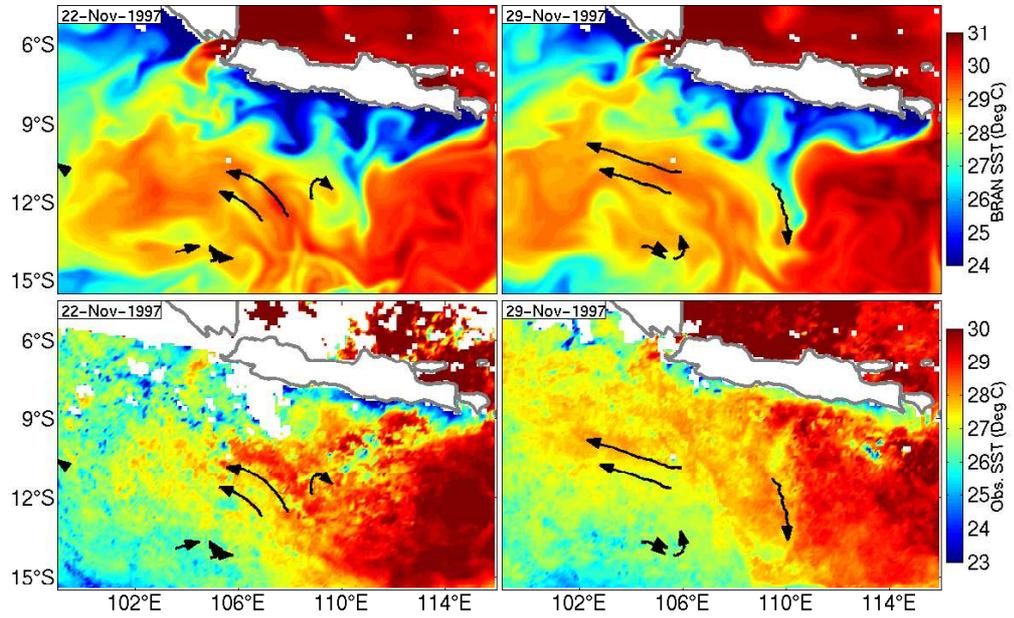
Figure 3. A sequence of maps off Java, showing daily averaged SST from BRAN (top) and from a 3-day composite from AVHRR (bottom) and surface drifter tracks for the 7 day period centered at the specified date.
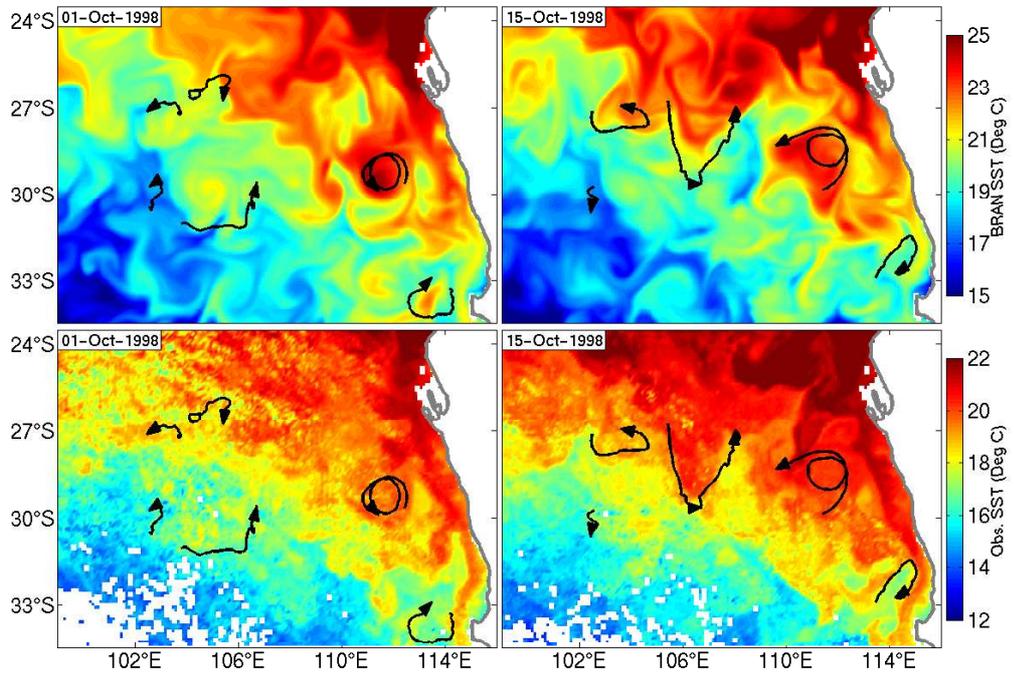
Figure 4. As for Fig. 3, except off Western Australia with surface drifter tracks for the 15 day period centered at the specified date.
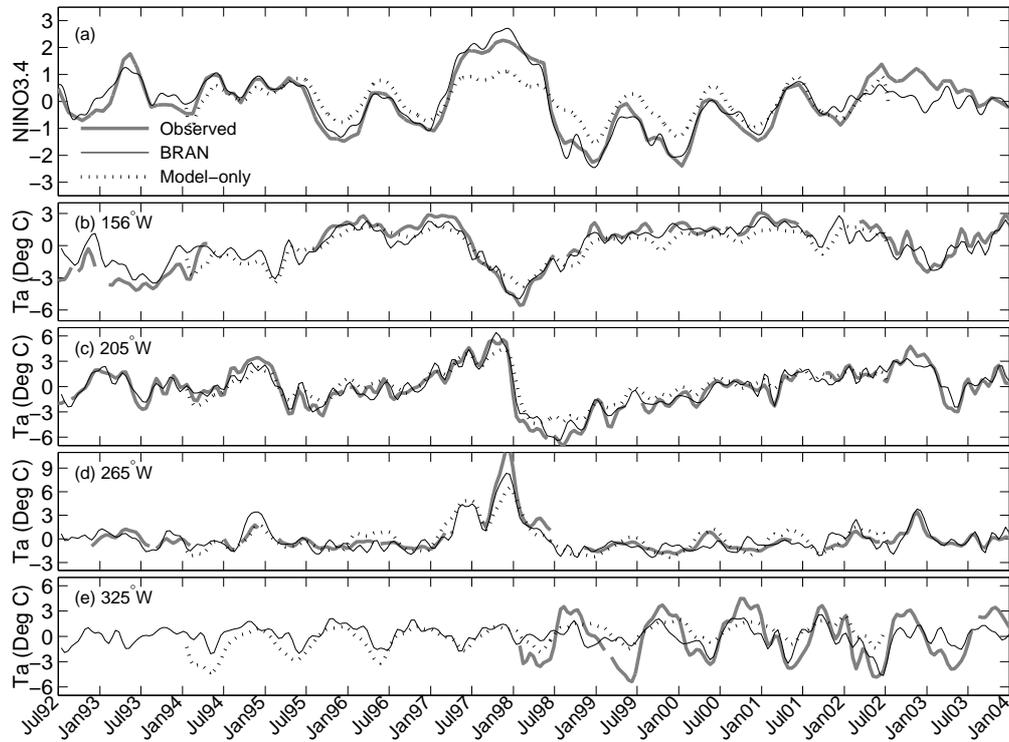
Figure 5. Comparisons between the low-pass filtered observed (half tone), BRAN (solid) and model-only (dashed) (a) NINO3.4 index, and (b-e) temperature at 100 m along the equator.

TABLE 1.   COMPARISONS BETWEEN THE MODEL-ONLY RUN, BRAN AND OBSERVATIONS OF TEMPERATURE AND 100 M DEPTH ALONG THE EQUATOR IN THE PACIFIC AND ATLANTIC OCEANS

Longitude	Δx	Mean (°C)			Standard Deviation (°C)			RMSE (°C)	
		Model	BRAN	Obs.	Model	BRAN	Obs.	Model	BRAN
156°E	0.1°	28.68	27.85	26.89	1.50	1.84	2.19	1.70	1.61
205°E	0.9°	25.52	25.20	24.25	1.95	2.36	2.57	2.11	1.63
265°E	0.9°	18.83	17.48	15.18	1.67	1.90	1.98	3.86	2.63
325°E	2.0°	24.89	24.67	21.11	1.04	1.70	2.85	4.67	4.34

Δx refers to the model's zonal resolution; RMSE is the root-mean-squared error; RMSEs, means and standard deviations are calculated for the periods when all three sources have data available.

An assessment of the temperature variability in the equatorial Pacific and Atlantic Oceans in BRAN is provided in Fig. 5 and Tab. 1, where the observed, reanalysed and model-only times series of NINO3.4 index are presented along with a series of comparisons of the low-pass filtered temperature anomalies from 100 m depth. Importantly, both BRAN and the model-only run are relaxed to SST over the same times scale (30 days). The comparisons of the NINO3.4 index demonstrate that BRAN gives a good representation of basin-scale phenomenon, such as El Nino. By contrast, and despite using SST observations in the same way as BRAN, the model-only run appears to give a poorer representation of El Nino. Anomalies of the sub-surface temperature fields along the equator are well represented in both BRAN and the model-only run. The magnitude of the

temperature variability, represented in Table 1 by the standard deviations, is more realistic in BRAN and similarly the root-mean-squared error (RMSE) is smaller in BRAN than it is in the model-only run. The temperature comparisons are presented for four different longitudes along the equator. Of these longitudes, the model's zonal resolution is highest at 156°E and decreases to the east, to a maximum of 2° at 325°E. The comparisons in Fig. 5 and Table 1 demonstrate that in both of the runs described here, the quantitative agreement with observations is worse in the coarser resolution parts of the domain, as we expect. However, we note that there is still reasonable qualitative agreement between the observed and modelled fields in the coarse regions of the domain.

5.   CONCLUDING REMARKS

This paper describes recent developments in ocean data assimilation in Australia. This has involved the development of BODAS, an ensemble optimal interpolation system, and its application to OFAM, a global ocean general circulation model. A 13-year ocean reanalysis, BRAN, has been performed, providing the first comprehensive ocean reanalysis with high resolution in the Asian-Australian region for the period 1992-2004. Comparisons with independent observations demonstrate that BRAN typically provides a good representation of the mesoscale ocean variability around Australia much of the time. This suggests that the integrated global ocean observing system, in combination with a relatively simple ocean data assimilation system, is able to reproduce the mesoscale eddy-fields in a high-resolution global ocean circulation model. Ongoing developments involve the generalisation of BODAS so that it can be easily applied to any ocean model on a range of different horizontal and vertical grids. In addition to OFAM, other applications to which BODAS has been applied range from high resolution coastal models for short-range prediction to coarse resolution global models for seasonal prediction. An ensemble Kalman filter version of BODAS has also been developed and is being applied to various regional applications. The next step in the Bluelink project is to modify this system for operational implementation at the Australian Bureau of Meteorology. The operational system will vary from BRAN in a number of important ways that may reduce the performance from what is demonstrated in this article. These variations include the use of near real-time observations which are reduced in both quantity and quality to delayed mode observations; the use of surface fluxes from operational atmospheric forecasts rather than reanalyses; and the system configuration changes that are required to perform robustly. A significant effort towards this end is currently underway.

REFERENCES

Argo Science Team   1998   On the design and implementation of Argo: An initial plan for a global array of profiling floats. International CLIVAR Project Office Rep. 21, GODAE Rep. 5, GODAE Project Office, Melbourne, Australia, 32 pp.

Chen, D., Rothstein, L. M. and Busalacchi, A. J.   1994   A hybrid vertical mixing scheme and its application to tropical ocean models. *J. Phys. Oceanogr.,* **24,** 2156-2179.

Cresswell, G.   2000   Currents of the Continental Shelf and Upper Slope of Tasmania. *Proc. Royal Soc. Of Tasmania,* **133** 23-30.

Du, Y., Qu, T., Meyers, G., Masumoto, Y., Sasaki, H.   2005   Seasonal heat budget in the mixed layer of the southeastern tropical Indian Ocean in a high-resolution ocean general circulation model. *J. Geophys. Res.,* **110,** doi:10.1029/2004JC002845.

Evensen, G.   2003   The Ensemble Kalman Filter: theoretical formulation and practical implementation. *Ocean Dyn.,* **53,** 343-367.

Feng, M., Wijffels, S., Godfrey, J. S. and Meyers, G.   2005   Do eddies play a role in the momentum balance of the Leeuwin Current? *J. Phys. Oceanogr.,* in press

Gaspari, G. and Cohn, S. E.   1999   Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.,* **125,** 723-757

Gent, P. R. and McWilliams, J. C.   1990   Isopycnal mixing in ocean circulation models. *J. Phys. Oceanogr.,* **20,** 150-155

Griffies, S. M. and Hallberg, R. W.   2000   Biharmonic friction with a Smagorinsky viscosity for use in large-scale eddy-permitting ocean models. *Mon. Weath. Rev.,* **128,** 2935-2946

Griffies, S. M., Harrison, M. J., Pacanowski, R. C. and Rosati, A.   2004   A technical guide to MOM4. GFDL ocean group technical Report No. 5, pp 339.

Hamill, T. M., Whitaker, J. S. and Snyder, C.   2001   Distance-dependent filtering of background error covariances in an ensemble Kalman Filter. *Mon. Weath. Rev.,* **129,** 2776-2790

Houtekamer, P. L. and Mitchell, H. L.   2002   A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Weath. Rev.,* **129,** 123-137

Kallberg, P., Simmons, A., Uppala, S. and Fuentes, M.   2004   The ERA-40 archive. Reading, UK, European Centre for Medium-range Weather Forecasts (ECMWF), ECMWF Re-Analysis Project (ERA). *ERA-40 Project Report Series,* **17,** 31p

Levitus, S.   2001   *World Ocean Database.* NOAA Professional Paper, **13**

McPhaden, M. J., and co-authors   1998   The Tropical Ocean Global Atmosphere (TOGA) observing system: a decade of progress. *J. Geophys. Res.,* **103,** 14,169-14,240

Mitchell, H. L., Houtekamer, P. L. and Pellerin, G.   2002   Ensemble size, balance, and model-error representation in an ensemble Kalman filter. *Mon. Weath. Rev.,* **130,** 2791-2808

Oke, P. R., Allen, J. S., Miller, R. N., Egbert, G. D. and Kosro, P. M.   2002   Assimilation of surface velocity data into a primitive equation coastal ocean model. *J. Geophys. Res. Ocean.,* **107(C9),** 3122

Reynolds, R. W., and Smith, T. M.   1994   Improved global sea surface temperature analyses using optimum interpolation. *J. Climate,* **7,** 929-948

Ridgway, K. R., Dunn, J. R. and Wilkin, J. L.   2002   Ocean interpolation by four-dimensional weighted least squaresApplication to the waters around Australasia. *J. Atmos. Oceanic. Technol.,* **19,** 1357-1375

Schiller, A., Brassington, G. B., Fiedler, R., Griffin, D. A., Mansbridge, J., Oke, P. R., Ridgeway, K., Smith, N. R.   2005   Eddy-resolving ocean circulation in the Asian-Australian region inferred from an ocean reanalysis effort. *manuscript submitted to J. Geophys. Res. Ocean.,*