

## Status and future of data assimilation in operational oceanography

M.J. Martin<sup>a\*</sup>, M. Balmaseda<sup>b</sup>, L. Bertino<sup>c</sup>, P. Brasseur<sup>d</sup>, G. Brassington<sup>e</sup>, J. Cummings<sup>f</sup>, Y. Fujii<sup>g</sup>, D.J. Lea<sup>a</sup>, J.-M. Lellouche<sup>h</sup>, K. Mogensen<sup>b</sup>, P.R. Oke<sup>i</sup>, G.C. Smith<sup>j</sup>, C.-E. Testut<sup>h</sup>, G.A. Waagbø<sup>k</sup>, J. Waters<sup>a</sup> and A.T. Weaver<sup>l</sup>

<sup>a</sup>Met Office, Exeter, UK; <sup>b</sup>ECMWF, Reading, UK; <sup>c</sup>NERSC, Bergen, Norway; <sup>d</sup>CNRS/LGGE, Grenoble, France; <sup>e</sup>Bureau of Meteorology, Melbourne, Australia; <sup>f</sup>NRL, Monterey, USA; <sup>g</sup>JMA/MRI, Tsukuba, Japan; <sup>h</sup>Mercator Ocean, Toulouse, France; <sup>i</sup>CSIRO, Hobart, Australia; <sup>j</sup>Environment Canada, Montreal, Canada; <sup>k</sup>Norwegian Meteorological Institute, Oslo, Norway; <sup>l</sup>CERFACS, Toulouse, France

The GODAE OceanView systems use various data assimilation algorithms, including 3DVar, EnOI, EnKF and the SEEK filter with a fixed basis, using different time windows. The main outputs of the operational data assimilation systems, the increments, have been compared for February 2014 in various regions. The eddy permitting systems' increments are similar in a number of the regions, indicating similar forecast errors are being corrected, while the eddy resolving systems represent smaller scale structures in the mid latitude regions investigated and appear to have smaller biases. Monthly average temperature increments show significant SST biases, particularly in the systems which assimilate swath satellite SST data, indicating systematic errors in the surface heat fluxes and the way in which they are propagated vertically by the ocean models. On going developments to the data assimilation systems include improvements to the specification of error covariances, improving assimilation of data near the equator, and understanding the effect of assimilation on the Atlantic Meridional Overturning Circulation. Longer term developments are expected to include the implementation of more advanced algorithms to make use of flow dependent error covariance information. Assimilation of new data sources over the coming years, such as wide swath altimetry, is also expected to improve the accuracy of ocean state estimates and forecasts provided by the GODAE OceanView systems.

### Introduction

The combination of observational data with model fields through data assimilation techniques is a crucial step in the production of accurate operational ocean forecasts and reanalyses. These state estimates rely on the robust and timely delivery of observational information from *in situ* and remote platforms (Clarke et al. 2015; Le Traon et al. 2015).

Accurate initial estimates of the mesoscale ocean state provide the basis from which forecasts of the ocean temperature, salinity and current structure, as well as sea-ice variables, can be made using the eddy-permitting and eddy-resolving GODAE OceanView (GOV) modelling systems (Tonani et al. 2015).

The data-assimilation derived state estimates provided by GOV systems are used in a wide range of applications. Their primary uses in near-real time are the initialisation of short-range ocean forecasts (out to a few days), and to initialise forecasts for longer time-ranges out to a few seasons using coupled ocean-atmosphere models. The same data assimilation systems are also used to produce reanalyses covering many years for various purposes including the calibration of coupled seasonal forecasts, and for understanding how the ocean has changed over the past decades (Balmaseda et al. 2015).

Data assimilation techniques are described in detail in the literature (Daley 1991; Evensen 2009; Zaron 2011) and the aim of this paper is not to describe all the various methods. Rather, we aim to give an overview of the data assimilation systems used in current state-of-the-art operational ocean forecasting systems. This builds on the description of the status of the GODAE data assimilation systems as of 2009 (Cummings et al. 2009).

The current status of operational data assimilation systems is described in the next section, including the observations assimilated, the types of algorithms used, the estimation and modelling of the error covariances, the initialisation methods, and other operational aspects. Following sections provide examples of the outputs of the data assimilation systems, describe some of the outstanding issues which are currently being worked on with the operational systems, and provide some indications as to the longer-term developments expected. A summary and conclusions are provided at the end.

### Current operational ocean data assimilation systems

Implementing data assimilative forecasting systems operationally imposes a number of constraints on the system design: the system must be robust and able to produce

\*Corresponding author. Email: [matthew.martin@metoffice.gov.uk](mailto:matthew.martin@metoffice.gov.uk)

analysis and forecast products within a set time using the computing resources available. The focus of this paper is on the schemes which have been implemented operationally, with these operational constraints, for basin-scale and global applications. A discussion of the coastal and shelf-seas systems is given in Kerafalou et al. 2015. The research done outside the operational centres is often very useful in determining future upgrades to the operational systems, and some areas of on-going and future research carried out in operational and research centres are discussed in later sections. Here we give an overview of various characteristics of existing operational data assimilation systems. The main objectives of these systems and their performance characteristics are described in other papers in this special issue (Hernandez et al. 2015).

### ***Assimilated observations***

The observations assimilated into each of the GOV forecasting systems in near-real time are detailed in Table 1. A variety of approaches is taken for use of sea surface temperature (SST) data with some systems directly assimilating swath data from the Group for High Resolution SST (GHRSSST; Bluelink, GOFS and FOAM) and other systems assimilating or relaxing to objective analysis products from GHRSSST. Most of the systems also assimilate *in situ* SST data available over the Global Telecommunications System (GTS) from ships, moored and drifting buoys. The GOFS system additionally assimilates Coastal-Marine Automated Network (C-MAN) *in situ* data available around the North American coast.

Along-track altimeter sea level anomaly (SLA) data are obtained from Aviso/CLS by most systems (directly or through MyOcean), Bluelink obtain the data from the Radar Altimeter Database System (RADS) and GOFS use data processed by the US Naval Oceanographic Office (NAVOCAENO). Almost all the systems assimilate the Jason-2, Cryosat-2 and Altika data. In order to assimilate the SLA data, a mean dynamic topography (MDT) is required, and various approaches are taken for this. Some systems (Bluelink, GOFS and TOPAZ) use a mean sea surface height field from a free model run as their MDT. Others, such as ECMWF, uses the MDT from a previous ocean reanalysis that assimilates temperature and salinity data. This approach means that the sea surface height observations (SLA+MDT) are not biased compared to the underlying model and the assimilation will only work to change the location of mesoscale features rather than adjusting the large-scale structure of the model dynamics. On the other hand, some systems use an observation-based MDT (CONCEPTS, FOAM, Mercator, and MOVE) which means that the large-scale dynamics will be brought into line with the observations. This approach can however cause problems if the modelled and observed MDT are significantly different to each other as the assimilation will constantly adjust the

model fields (including the subsurface density structure) (Lea et al. 2008). Some systems mitigate against this by applying a bias correction approach either off-line (Mercator) or on-line as part of the data assimilation (FOAM).

Profile temperature and salinity data are obtained either over the GTS or through an intermediate processing centre, CORIOLIS (for Mercator and CONCEPTS), with the Bluelink system accessing data through both these routes as well as from the US GODAE server, and generating a combined data-set. A similar set of platforms collecting profile temperature and salinity data are used by all the systems including Argo floats, XBTs (for temperature), moored buoys, gliders and marine mammals, with the TOPAZ system additionally assimilating ice-tethered profile data.

Sea-ice concentration data from satellites are assimilated only in the CONCEPTS, GOFS, FOAM, North Pacific (NP) MOVE and TOPAZ systems, with ECMWF prescribing (rather than assimilating) the sea-ice concentration based on observations. Most of these systems assimilate data from the SSM/I and/or SSMIS satellite instruments, with the CONCEPTS system also making use of sea-ice concentration data from Canadian Ice Service (CIS) charts, Radarsat and ship reports. TOPAZ is the only system to assimilate sea-ice drift data (estimated through cross-correlations of successive satellite images) operationally.

### ***Assimilation methods***

A summary of the grid specifications, the algorithm and the initialisation procedure used by each system is provided in Table 2, together with a reference for each system. The model resolution of the systems described here varies between about  $1/12^\circ$  for GOFS and Mercator HR (the higher resolution Mercator system),  $1/8^\circ$  for TOPAZ,  $1/4^\circ$  for CONCEPTS, FOAM and Mercator LR (the lower resolution Mercator system), and about  $1^\circ$  for the ECMWF and global MOVE systems. Bluelink runs on a  $1/10^\circ$  grid in the Australian region, with much lower resolution elsewhere. NP MOVE also runs on a  $1/10^\circ$  grid in the western north Pacific region. Most of the systems use the same grid to perform the data assimilation as the model grid. However, Bluelink, CONCEPTS and Mercator use a reduced resolution analysis grid to reduce the computational cost. GOFS uses the same horizontal grid for analysis as the model, but the vertical grid used is based on depth levels rather than the hybrid vertical coordinates of the HYCOM model.

A range of algorithms is used by the different operational systems including variational (3DVar) and ensemble schemes (Ensemble Optimal Interpolation, Ensemble Kalman Filter and a fixed-basis implementation of the SEEK filter). The variational schemes of GOFS (Cummings & Smedstad 2013), FOAM (Waters et al. 2014), ECMWF (Balmaseda et al. 2013) and MOVE

Table 1. Observation types assimilated into each GODAE OceanView system in near real time.

	Bluelink	CONCEPTS	ECMWF	FOAM	GOFS	Mercator	MOVE	TOPAZ
Institute (Country)	Bureau of Meteorology (Australia)	Environment Canada (Canada)	ECMWF (Europe)	Met Office (UK)	NAVOCEANO (USA)	Mercator (France)	JMA/MRI (Japan)	MET Norway (Norway)
Satellite SST	L2 GHRSSST (AVHRR, Wind SAT)	L4 analysis (CMC)	L4 analysis (OSTIA)	L2 GHRSSST (NOAA AVHRR, MetOp AVHRR)	L2 GHRSSST (NOAA AVHRR, MetOp AVHRR, GOESE, GOES W, MSG, MTSAT 2, NPP VIIRS) from NAVO	L4 analysis (NOAA AVHRR)	L4 analysis (COBE SST for global, GHRSSST L4 for regional).	L4 analysis (OSTIA)
<i>In situ</i> SST		Ships, drifting buoys, moored buoys, all from GTS		Ships, drifting buoys, moored buoys, all from GTS	Ships, drifting and moored buoys, CMAN stations, all from GTS		Ships, drifting buoys, moored buoys, all from GTS	
Altimeter SLA data	Jason 2, Cryosat 2, Altika, from RADS	Jason 2, Cryosat 2, Altika, from Aviso	Jason 2 from Aviso	Jason 2, Cryosat 2, Altika, from MyOcean	Jason 2, Altika, Cryosat 2, from NAVO	Jason 2, Cryosat 2, Altika, from MyOcean	Jason 2 IGDR/ GDR from Aviso	Jason 2, Cryosat 2, Altika, from MyOcean
MDT	Model mean (18 year average)	Rio and Hernandez, 2004	Model mean from a run assimilating only T/S data	CNES CLS09 with on line bias correction	Model mean	CNES CLS09, with off line adjustments	Monthly climatology generated using analysis system without model.	Model mean.
<i>In situ</i> profiles (T and S)	Argo, XBT, moored buoys, consolidated from GTS, US GODAE, and Coriolis	Argo, XBT, CTD, moored buoys, gliders, marine mammals from CORIOLIS	Argo, XBT, CTD, moored buoys, gliders, marine mammals, all from GTS	Argo, XBT, CTD, moored buoys, gliders, marine mammals, all from GTS	Argo, XBT, CTD, moored buoys, gliders, marine mammals, from GTS and US Navy sources	Argo, XBT, CTD, moored buoys, gliders, marine mammals from CORIOLIS	Argo, XBT, CTD, moored buoys, marine mammals, all from GTS	Argo, XBTs, moorings, icetethered profilers, all from My Ocean
Sea ice concentration		SSM/I, SSMIS, CIS charts, Radarsat, ship reports	Prescribed sea ice based on SSMIS, from EUMETSAT OSISAF, via OSTIA product.	SSMIS, from EUMETSAT OSISAF	SSM/I (F13, F14, F15), SSMIS (F16, F17, F18) from FNMOC		SSMIS (F16, F17, JMA Product) for NP system	SSM/I, from EUMETSAT OSISAF
Sea ice drift								OSI SAF at MET Norway

Table 2. Overview of characteristics of the data assimilation used in each GODAE OceanView system.

	Bluelink <sup>a</sup>	CONCEPTS <sup>b</sup>	ECMWF <sup>c</sup>	FOAM <sup>d</sup>	GIFS <sup>e</sup>	Mercator <sup>f</sup>	MOVE <sup>g</sup>	TOPAZ <sup>h</sup>
<i>Model configurations run operationally</i>	Near global 1/10°; 51 z levels. (scheduled for 2015); MOM4.	Global 1/4°; 50 z levels; NEMO.	Global 1° (0.3° meridionally at equator); 42 z levels; NEMO.	Global 1/4°; 75 z levels; NEMO.	Global 1/12.5°; 32 hybrid layers; HYCOM.	Two global systems, 1/4° and 1/12°; 50 z levels; NEMO.	Near Global 1° (0.3° meridionally at equator); 50 z levels; MRI. COM.	N. Atlantic & Arctic 1/8°; 28 hybrid layers; HYCOM.
	Australian region 1/10°; 51 z levels; MOM4.			N. Atlantic, Indian Ocean, Mediterranean, all 1/12°; 50 z levels; NEMO.		Atlantic (20°S 80°N) & Mediterranean 1/12°; 50 z levels; NEMO.	N. Pacific 1/2° (1/10° in Western N. Pacific); 54 z levels; MRI. COM.	
<i>Analysis grid resolution</i>	Every second model (horizontal) grid point.	Every third model (horizontal) grid point; more points within 150km of coast.	Same as model grid.	Same as model grid.	1/12.5° with 42 z levels	Every fourth model (horizontal) grid point; more points within 150km of coast.	Same as model grid.	Same as model grid.
<i>Algorithm</i>	EnOI	SEEK FGAT (with fixed basis) (ocean) and 3DVar FGAT (ice)	3DVar FGAT + bias correction	3DVar FGAT + bias correction	3DVar FGAT	SEEK FGAT (with fixed basis); 3DVar bias correction	3DVar (ocean) and a simple least square method (ice)	EnKF with 100 members.
<i>Initialisation technique</i>	Adaptive nudging of 3D T, S, u, v over one day.	Direct initialisation	IAU with constant weighting of SSH and 3D T, S, u,v over 10 days.	IAU with constant weighting of SSH and 3D T,S, u,v over one day.	IAU with constant weighting of T, S, U, V, layer pressure (for isopycnal layers) over 6 hours.	IAU with a non constant distribution function of SSH and 3D T, S, U, V over 7 days.	Ocean: IAU with constant weighting of 3D T, S over 10 days (Global)/ 5 days (NP). Ice: nudging	Direct initialisation

<sup>a</sup>Oke et al. 2013<sup>b</sup>Smith et al. 2014<sup>c</sup>Balmaseda et al. 2013<sup>d</sup>Waters et al. 2014<sup>e</sup>Cummings and Smedstad 2013<sup>f</sup>Lellouche et al. 2013<sup>g</sup>Usui et al. 2006<sup>h</sup>Sakov et al. 2012

(Usui et al. 2006) produce global minimisations to a cost function with efficient modelling of the background error covariances. They rely on *a priori* specification of physical relationships to transfer information between variables, and to determine the length-scales and variances used to weight and spread the information from the observations (described in more detail in the next section). In contrast, the ensemble schemes rely on statistical information from model runs in order to determine how observational information is spread by the assimilation. The Bluelink (Oke et al. 2013), CONCEPTS (Smith et al. 2014) and Mercator (Lellouche et al. 2013) ensemble systems use model data from previous long model integrations, while the TOPAZ (Sakov et al. 2012) system uses an up-to-date ensemble of model runs from the current forecast cycle. Most of the systems use a first-guess-at-appropriate-time (FGAT) scheme whereby the model counterpart of the observation is calculated close to the observation time.

All of the systems make updates to the full ocean state (three-dimensional temperature, salinity and velocity fields and SSH, or isopycnal layer pressure for GOFS) except for MOVE where only temperature and salinity fields are updated directly by the assimilation. For the other systems, corrections to the velocity fields are calculated using either dynamical or statistical relationships, rather than assimilating velocity data directly. The corrections (increments) produced by the data assimilation are used to initialise the forecasting systems in a number of ways, most of which are designed to reduce the generation of fast transient model responses from unbalanced increments. ECMWF, GOFS, FOAM, MOVE and Mercator use the Incremental Analysis Update (IAU) method (Bloom et al. 1996) which adds in a fraction of the increments to the model fields at each model time-step with the aim of reducing the shock to the model. The period over which these are added in varies from 6 hours in GOFS, to one day in FOAM, out to 5–10 days for Mercator, ECMWF and MOVE. All of these systems use a constant weighting for the IAU except for Mercator (Benkiran & Greiner 2008; Lellouche et al. 2013) who choose a distribution function which crosses the increments of two consecutive cycles in order to ensure continuity in the application of the increments from one cycle to the next. The Bluelink system uses an adaptive nudging scheme (Sandery et al. 2011) in which the model is nudged towards the analysis using a time-scale which depends on the magnitude of the increments. CONCEPTS and TOPAZ apply increments directly to the initial model state before the forecast is run.

Data assimilation methods generally assume that the model does not exhibit significant systematic errors. This is often not the case in practice, and some groups attempt to correct for model biases by using various schemes. For example, ECMWF employs a bias correction scheme which combines a correction to 3D temperature and salinity in the extra-tropics, with a correction to

the pressure gradient field in the tropics (Balmaseda 2007). This latter correction is also used in the FOAM system (Bell et al. 2004), while Mercator applies a general 3D temperature and salinity slowly-evolving large-scale bias correction globally. These bias correction fields are calculated by accumulating a running average of the data assimilation increments in the ECMWF and FOAM systems, with the Mercator biases calculated using a 3DVar approach based on innovations over the preceding months.

### *Error covariance estimation and modelling*

In the EnOI of Bluelink and the SEEK scheme as implemented at Mercator and CONCEPTS, a static ensemble of anomalies for three-dimensional temperature, salinity, currents and SSH is used to estimate the background error covariance matrix. A long integration of a free model is carried out and anomalies each day are calculated with respect to a running mean in order to estimate the 7-day scale error in the ocean state at a given period of the year. The anomaly calculation filters out temporal scales at low frequencies and retains the higher frequency signals (e.g. the mesoscale ocean structures) which are to be corrected by the data assimilation. A significant number of anomalies are kept from one analysis to the other, thus ensuring error covariance continuity. The covariances calculated using this technique are used to represent both the univariate and multivariate covariance relationships.

The EnKF scheme as implemented in TOPAZ uses a 100-member ensemble to represent the background error covariance of the current cycle. In order to maintain ensemble spread, a perturbation system works indirectly using dynamically consistent perturbations of the atmospheric forcing fields (Sakov et al. 2012). Inflation is also applied in the Deterministic EnKF by using a larger error variance of observations for updating the anomalies than for updating the mean, in order to maintain ensemble spread (Sakov et al. 2012). The Mercator system also inflates the error variances: at each assimilation cycle, an adaptive scheme calculates an optimal variance of the model forecast error maintaining the statistical equilibrium between the innovation, and the background and observation error covariances.

All of these ensemble-based schemes require localisation of the covariance information to prevent noisy signals at larger separation distances contaminating the increments, due to the relatively small size of the ensembles. Localisation of the impact of observations in TOPAZ is applied so that observations can only affect model variables within a radius of 300 km. Mercator uses a localisation technique which sets the covariances to zero beyond a distance defined as twice the local spatial correlation scales.



The 3DVar schemes used at ECMWF, FOAM and GOFS parameterise the background error covariance matrix by splitting the problem into the specification of error variances (the diagonal elements), and horizontal and vertical lengthscales associated with a functional representation for the correlations (the off-diagonal elements). The 3DVar scheme in MOVE parameterises it by splitting the problem into the specification of error variances, vertical correlations and horizontal length-scales. GOFS specifies the background error variances using an estimate based on model variability computed from a time history of differences between model forecasts separated by the update cycle interval, which includes the influence of the observations. ECMWF and FOAM use a parameterisation of the background temperature variances which depends on the vertical gradient of temperature in the background field; FOAM includes seasonally varying estimates at the surface based on output of previous assimilative runs. Using these various techniques, these systems (other than MOVE) account for some degree of flow-dependence in their background error variance specification.

FOAM and GOFS use the Rossby radius to define the horizontal length scales in order to represent the scales of mesoscale phenomena, while ECMWF uses  $2^\circ$  lengthscales with equatorial refinement so that the ocean features important for seasonal forecasting can be initialised, and MOVE splits the global domain into multiple (about 50 for global MOVE) sub-domains and specifies lengthscales in each one. The vertical length-scales in GOFS are specified based on the vertical density gradients, and in FOAM are based on the mixed-layer depth of the background field. In ECMWF they are specified based on the vertical grid-spacing of the model.

The univariate component of the background error correlations is modelled using a diffusion operator in ECMWF (Weaver & Courtier 2001) and FOAM (Mirouze & Weaver 2010) which approximates Gaussian correlation functions whereas GOFS uses a Second-Order Auto-Regressive (SOAR) function. The component is represented by the univariate nature of vertical coupled temperature-salinity (T-S) Empirical Orthogonal Functions (EOFs) from historical temperature and salinity profile data in MOVE (Fujii & Kamachi 2003). The multivariate components of the background errors (i.e. those transferring information from one variable to another) are specified in GOFS, ECMWF and FOAM using pre-determined physical relationships. In ECMWF and FOAM these transfer information from temperature to salinity using information from the model background water mass properties (Ricci et al. 2005), from density to sea surface height using the dynamic height relation, and from pressure to velocity using the geostrophic relation. GOFS transfers information from temperature and salinity to geopotential using the equation of state and hydrostatic relation, and then to velocity using the geostrophic relation. MOVE transfers information

between temperature and salinity through the coupled nature of the coupled T-S EOFs.

All of the systems except for MOVE represent the observation error covariances as a diagonal matrix, omitting any representation of error correlations between observations. In most systems these are specified *a priori* as either constant or seasonally varying estimates, and most systems specify spatially varying representativity errors (Oke & Sakov 2008). In GOFS they are specified based on the current state using a look-up table with profile adjustments based on vertical gradients. MOVE specifies the matrix based on the historical profile data, and approximates it using the first two vertical modes.

### Operational aspects

The delay in receiving an observation at the operational centres, compared to its measurement time (called timeliness here), depends on a number of factors including the transmission time, and the time taken to process and quality control the observation by the data producers. The timeliness of different observing systems puts constraints on the operational forecasting systems in terms of how close to real-time they can be run. SST and sea-ice data tend to be available relatively quickly (less than a day), altimeter SSH data are mostly available within two days, and profile data can take up to a few days to arrive, depending on the platform. Most of the systems therefore have two streams of analyses: one near-real time (NRT) analysis which provides up-to-date information in order to initialise forecasts; and one delayed analysis which provides less timely but more accurate information. The delayed analyses are used to initialise the next cycle of the NRT analyses so that information from late-receipt observations is still able to affect the forecasts. The precise way in which this set-up is implemented varies between the systems.

FOAM and GOFS are run on a daily basis assimilating all available data. Late receipt observations are assimilated in FOAM by going back two days each day with the ocean state at T-24h being used to initialise the next day's analysis (Blockley et al. 2014). GOFS assimilates late data by comparing those data to the model fields at the correct time (FGAT), and using those innovations in the current analysis. Bluelink runs each day using a time window which depends on the observation type: one day for SST; 7 days for *in situ* temperature and salinity; and 11 days for altimeter data. A multi-cycle ensemble is used in Bluelink whereby daily forecasts are made of four independent forecast members, each staggered by a day (Brassington 2013). Each forecast member is run on a different day, and an analysis is performed on that member every four days.

CONCEPTS, Mercator and TOPAZ all run the full assimilation system once a week with a 7-day time-window, with an additional analysis of the previous 7-days (T-14 to T-7) to enable assimilation of late-receipt

data carried out by CONCEPTS and Mercator. CONCEPTS additionally carries out a daily assimilation which includes only SST and seaice data, and daily forecasts are started from those analyses, while Mercator runs daily forecasts with fields updated from the last analysis by running the model with analysed surface fluxes. The TOPAZ analysis is carried out with a three day delay in order to allow most data to be assimilated.

ECMWF and global MOVE both use a 10-day time-window. ECMWF perform a full analysis in delayed mode every 10 days with a 6 day delay. The near-real time analyses are updated daily with a variable assimilation window starting from the last delayed mode analysis available. Global MOVE runs every 5 days and produces analyses for the previous 40 days by running four 10-day cycles for the near-real time mode, and running a ten-day cycle every 5 days for the delayed mode.

The time taken to run each system is between 20 minutes for the ECMWF analysis (the 10-day cycle of the delayed analysis, excluding the forecast, with the 5 ensemble members running in parallel), 30 minutes for global MOVE (excluding the forecast), one hour for the CONCEPTS analysis and forecast, approximately 2 hours for FOAM and GOFS to run the analysis and forecast, and 4.7 hours for the Mercator HR analysis and forecast.

### Demonstration of operational data assimilation schemes

In order to illustrate the outputs of the data assimilation schemes - as opposed to the forecast systems as a whole - the increments from each of the operational systems described above are compared for February 2014. The increments are useful to show as they illustrate the type of structures which each analysis system produces. These depend on the background model field and the observations assimilated (which also depend on the time window and the delay in running the operational system). They also depend on the error covariance structures specified in the data assimilation algorithm, and on the data assimilation algorithm itself.

As described in the previous section, the different systems run on different cycles, so it is impossible to show increments for exactly equivalent periods. For the plots in [Figures 1–5](#) the ECMWF fields are for the 10-day cycle 14th–23rd Feb, the global MOVE fields for the 10-day cycle 15th–24th Feb, the NP MOVE fields for the 5-day cycle 20–24th Feb, the Mercator and CONCEPTS fields are for the 7 day cycle 19th–25th Feb, and the Bluelink, FOAM and GOFS fields are the sum of the daily increments over the same 7 days as the Mercator/CONCEPTS fields. Due to the different periods over which the increments are valid, their magnitudes are different. A simple scaling of the fields based on the assimilation time windows is not appropriate as the growth of forecast

error is not linear (accumulating 10-days of daily increments does not give the same result as running a 10-day assimilation cycle, even with the same system). In order to plot comparable fields on the same colour-scale, we have normalised the spatial root-mean-square of each field to be one (K for temperature and m/s for currents); the factor multiplying the fields in order to achieve this is shown in brackets in each plot title. This is done globally for the level and variable plotted) except for the cross-section plots in which the factor is calculated based on the region plotted.

### Regional comparison of 5–10 day increments

Four regions are focussed on for this inter-comparison. The East Australian Current (EAC), Kuroshio and Gulf Stream regions are shown to illustrate how the eddy-permitting and eddy-resolving data assimilation systems represent changes in these important boundary current regions. Some of the systems are also used to initialise longer-range predictions so the tropical Pacific region, an important area for initialisation of seasonal forecasts, is also shown. In each region, only those systems which resolve or permit the dominant features in that region are shown. In particular, the ECMWF and global MOVE increments are only shown in the tropical Pacific region as they do not represent the mesoscale structures in the boundary current regions, Bluelink is only shown in the EAC region as it is much lower resolution in the other regions, TOPAZ is only shown in the Gulf Stream region, and NP MOVE is only shown in the Kuroshio region.

[Figure 1](#) shows spatial maps of the temperature increments in the EAC region. To investigate how these are related to the vertical structure of changes made by the assimilation systems, cross-sections are also shown at 35° S down to 500m depth. Despite using different models and assimilation schemes, there are some similarities in the increments produced by Mercator, Bluelink, CONCEPTS and FOAM in the region 30–40°S, west of about 160°E. The magnitude of the increments is different in each system but the location and scale of the increments are similar. The positive-negative-positive temperature increments tri-pole structure moving out from the east Australian coast is reflected in almost all the systems, although Mercator HR produces a negative anomaly close to the coast. Very few *in situ* temperature data are in the vicinity of these features during this period, so they must be due to the projection of altimeter SLA data through the multivariate aspects of the data assimilation systems. The depth of the maximum temperature increment is different in each system, with FOAM and GOFS producing a maximum at about 100 m, with Mercator, CONCEPTS and Bluelink having a broader vertical structure below the top 50 m with a maximum closer to 200 m depth for Mercator/CONCEPTS and 300 m for Bluelink.

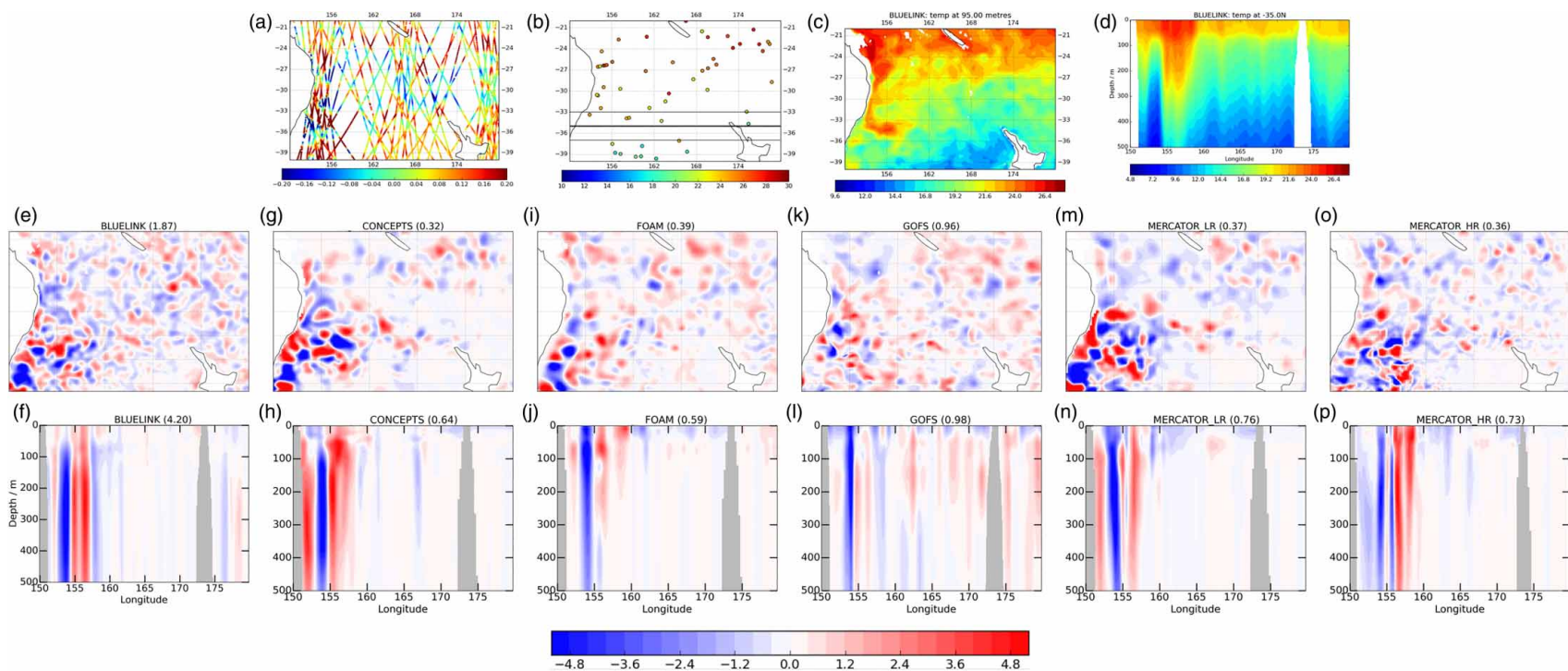


Figure 1. Normalised temperature increments in the EAC region for 19th–25th Feb 2014. (a) and (b) show the SLA and profile data coverage for reference. (c) (d) show the BlueLink analysis averaged over the same period. (e) (p): top row shows T increments at 100 m depth; bottom row shows T increments cross section along 35S between 150E–180E for each system. The numbers in brackets indicate the normalisation applied in the plot below in order to make the spatial root mean square of each field to be 1K.



These differences could be due to the different approaches for projecting the altimeter data into the sub-surface density field: GOFS and FOAM use flow-dependent parameterisations based on the current model background field as opposed to the static ensemble estimates used by Bluelink, Mercator and CONCEPTS.

The surface increments of speed (taking into account the increments to both components of velocity) made by each system are shown in Figure 2 for the equivalent region and period as Figure 1. The horizontal scales of the velocity increments in CONCEPTS, FOAM and Mercator LR are similar although the reduced resolution analysis grid used in Mercator LR results in less smooth structures. The Bluelink and GOFS velocity increments

appear to have smaller spatial scales than the other systems illustrating the difference from having a higher-resolution model able to resolve (rather than permit) mesoscale features at this latitude, and the ability of the assimilation schemes to represent those scales. The FOAM velocity increments are close to zero near the coast due to a scheme that increases the observation errors for all data types near coastlines.

Figure 3 shows temperature increments at about 100 m depth in the Kuroshio region. The GOFS and Mercator HR systems have smaller scale features than the other systems as would be expected due to their higher resolution. The systems have fewer common features in this region than they do in the EAC region, indicating that the forecast

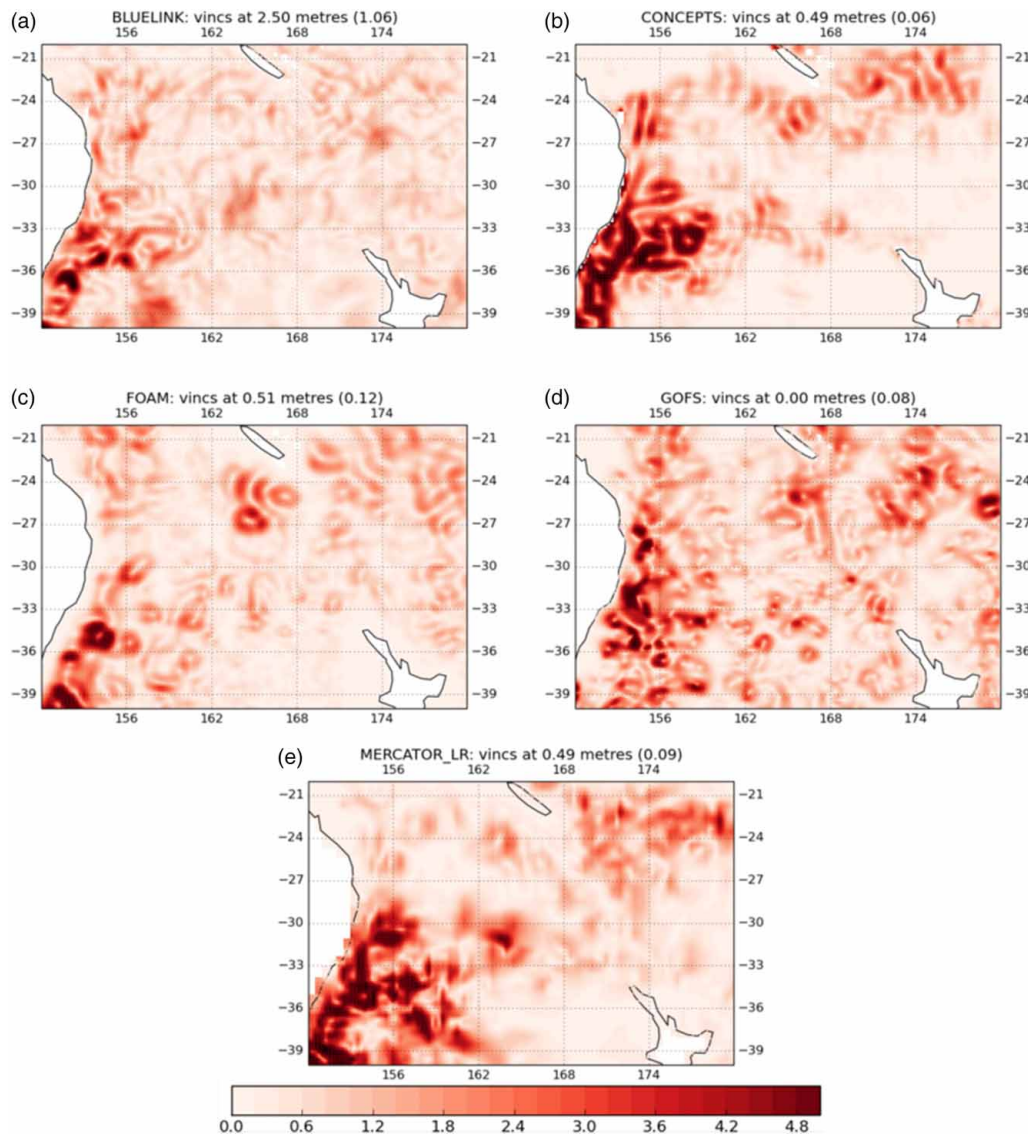


Figure 2. Normalised increments of speed (taking into account increments of both components of velocity) in the EAC region for 19th 25th Feb 2014 for the top analysis level. The numbers in brackets indicate the normalisation applied in the plot below in order to make the spatial root mean square of each field to be 1m/s.

errors during this period in this region are not the same in any of the systems. However, there is a mesoscale feature in the increments at about 154°E in the cross-section plots which all the systems except for CONCEPTS represent.

In the Gulf Stream region, shown in Figure 4, the CONCEPTS, FOAM, Mercator LR and TOPAZ systems all have a dipole structure near the separation of the Gulf Stream from the eastern US coast, indicating that the assimilation is acting to correct a model bias which is introducing too much lateral mixing. GOFS and, to a lesser extent Mercator HR, appear to have less of a dipole structure in this region, with GOFS having much less bias in the lateral mixing in this region. This demonstrates the improvements which might be expected from higher resolution models, and could be due to improved representation of lateral mixing in the GOFS implementation of the HYCOM model.

Figure 5 shows the temperature increments produced by each analysis system in the eastern tropical Pacific. At the Equator, where planetary waves propagate fast, it can be argued that the horizontal decorrelation scales depend more on the length of the assimilation window than on the ocean model resolution. From the horizontal plots it is clear that the longer-window and lower resolution seasonal forecasting systems of ECMWF and MOVE have much larger east-west scales close to the equator than the other analysis systems. CONCEPTS and Mercator have length-scales which appear to reduce more quickly poleward of the equator compared to FOAM and GOFS. The cross-section plots show that each system produces the largest increments in the region of the thermocline at around 180 m depth in the west of the region, and at about 20 m depth in the east. The larger horizontal scales of the ECMWF and MOVE systems are again clearly visible, with the other systems' increments containing more small-scale structures. The FOAM system produces increments with larger vertical extent in the western part of the domain than the other systems, probably due to the error covariance specification in that region.

### **Global comparison of monthly average increments**

The global monthly averaged temperature increments have been calculated for each system and are plotted in Figure 6 for the surface and at about 100 m depth. These are a useful diagnostic as they represent an estimate of forecast bias, although this is only the case in regions where observations are available to identify the bias, and observational bias can contaminate the signal.

FOAM and GOFS have a clear pattern at the surface with negative increments in the tropics and positive increments at mid-latitudes in the summer (southern) hemisphere. These two systems assimilate swath satellite SST data while the others relax to or assimilate objective SST analyses, which could be why the model biases are more

obvious in the increments from FOAM and GOFS. Systematic errors in the heat fluxes applied to the model would consistently heat/cool the surface layers and therefore result in these large average increments. The heat fluxes used in the GOFS system are adjusted to correct for atmospheric model forcing errors by adding heat across most of the tropical regions, scaled to minimise 5-day SST forecast errors. The correction results in the model being too warm at the update cycle interval (24 hours), which in turn causes the surface temperature increments to be negative on average. No such correction is carried out in the FOAM system, but the underlying heat fluxes presumably also exhibit this bias. Further investigations into these biases, relating them to heat flux differences, could be an informative follow-up to this work. For more information about the surface forcing used by each system see Tonani et al. 2015.

Almost all the systems have a bias in the cold tongue region of the eastern tropical Pacific which is much more obvious in the monthly averages than the increments from one week shown before. This period coincides with a developing negative SST anomaly and it appears that the assimilation is acting to increase the magnitude of the colder SSTs. At about 100m depth a cold anomaly centred at about 130°W on the equator is being put in by the data assimilation in the ECMWF, FOAM, GOFS and Mercator systems. The western boundary current regions are clearly visible in almost all the systems, with the eddy resolving and permitting systems correcting for persistent errors in the position of mesoscale structures and the lower resolution systems correcting for larger scale forecast errors.

### **Some outstanding operational and research issues in ocean data assimilation**

A sample of outstanding operational and research questions which are currently being worked on by various groups are described in this section. This is not an exhaustive list due to space limitations, but is meant to illustrate some important developments.

#### **Data timeliness issues**

The requirement described earlier in this paper for the operational systems to be re-run in order to make use of late-receipt data is due to the lack of timeliness in particular observing systems. As an example of this issue, the timeliness of receipt for *in situ* profile observation types received at the US Naval Research Laboratory (NRL) in Monterey is shown in Figure 7 for September–November 2012. These timeliness estimates are likely to be similar for other operational centres who receive their data over the GTS. Some data types are very timely while others arrive much later. Some data types like the profiles from sensors on marine

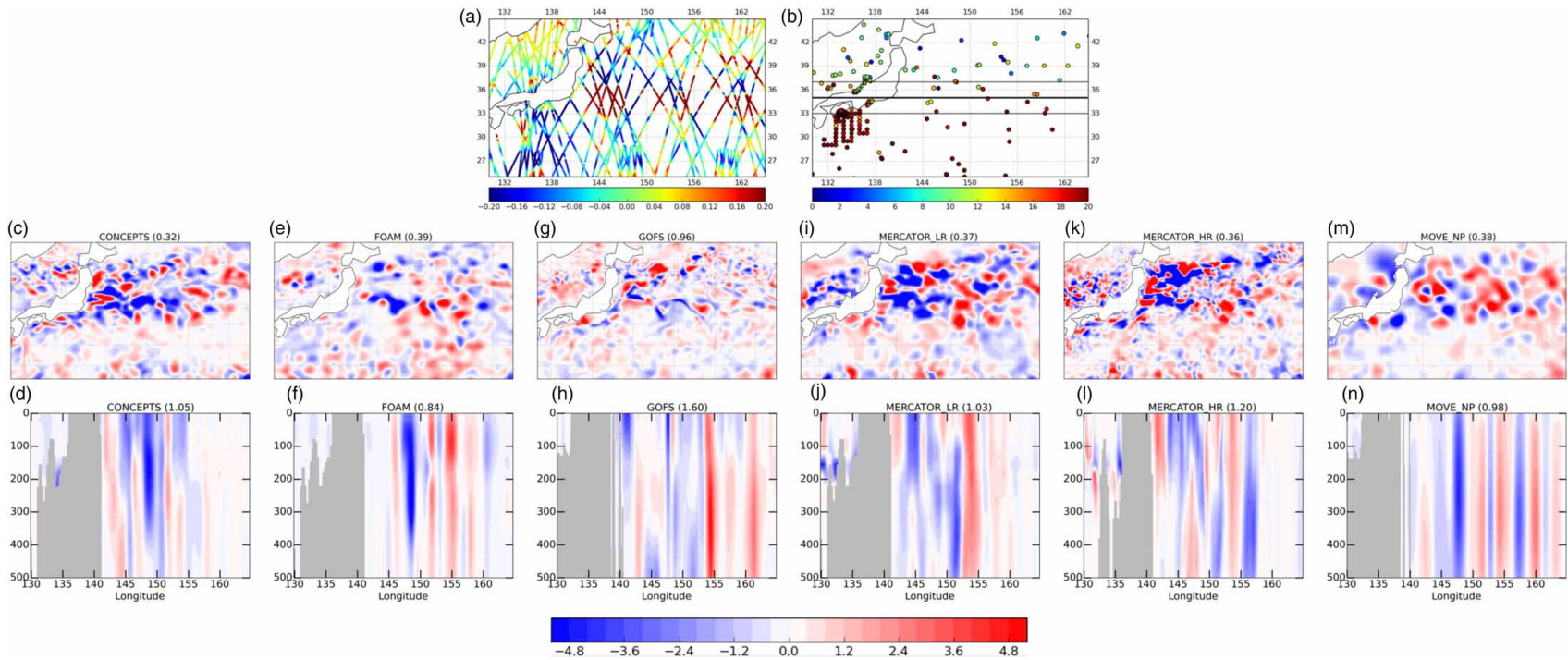


Figure 3. Normalised temperature increments in the Kuroshio region for 19th–25th Feb 2014 (except NP MOVE which is the analysis for 20th–24th Feb 2014). (a) and (b) show the SLA and profile data coverage for reference. (c)–(n): top row shows T increments at 100m depth; bottom row shows T increments cross section along 35N between 130E–165E for each system. The numbers in brackets indicate the normalisation applied in the plot below in order to make the spatial root mean square of each field to be 1K.

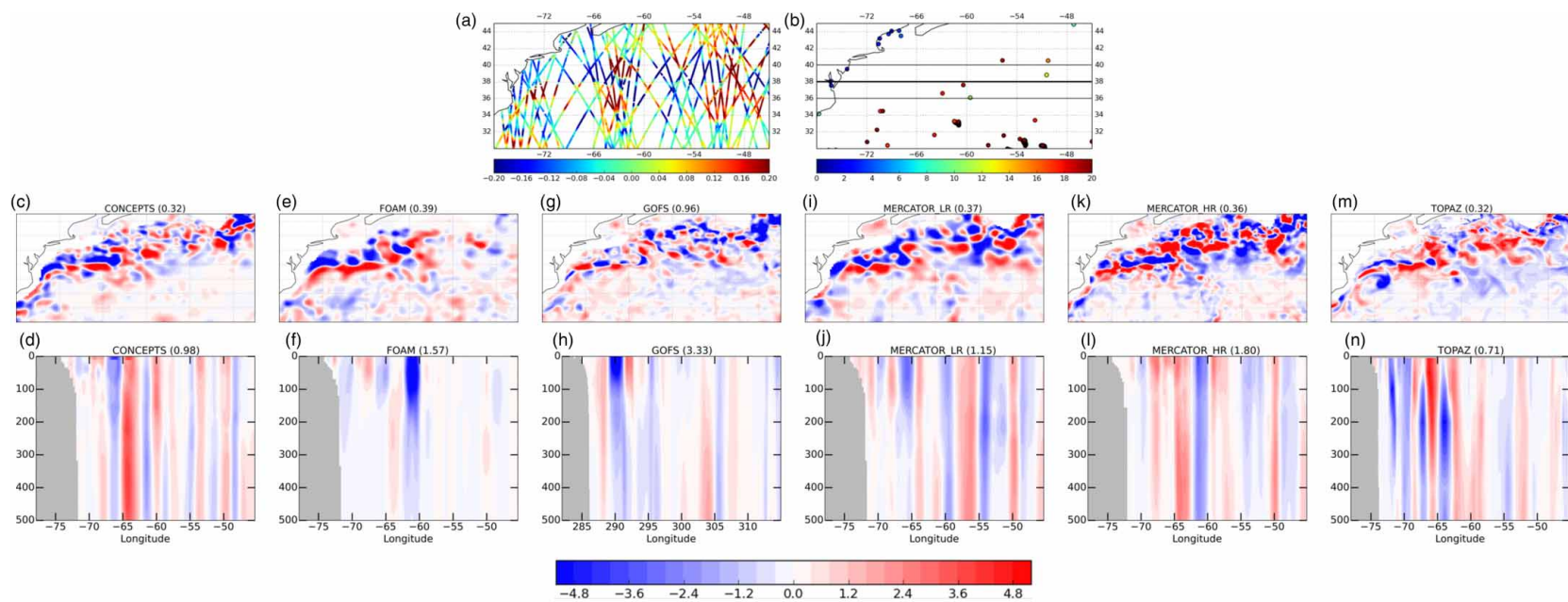


Figure 4. Normalised temperature increments in the Gulf Stream region for 19th-25th Feb 2014. (a) and (b) show the SLA and profile data coverage for reference. (c)–(n): top row shows T increments at 100m depth; bottom row shows T increments cross section along 38N between 78W–45W for each system. The numbers in brackets indicate the normalisation applied in the plot below in order to make the spatial root mean square of each field to be 1K.



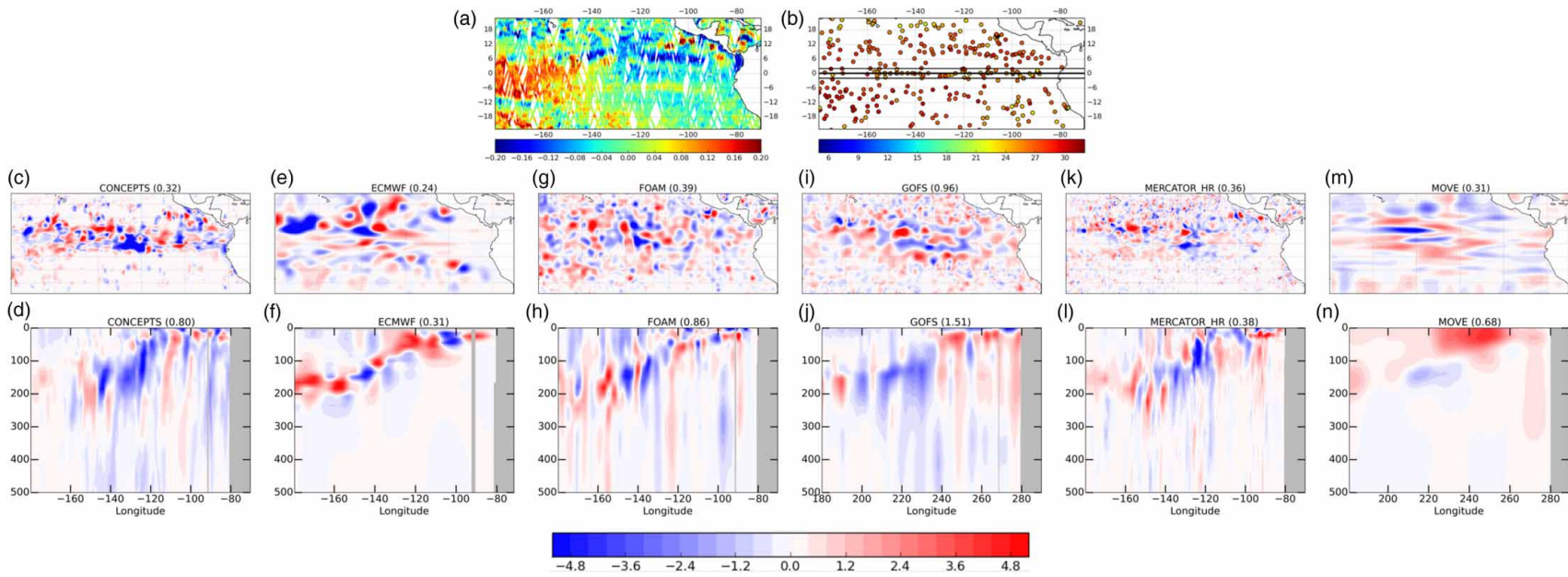


Figure 5. Normalised temperature increments in the Central and Eastern Equatorial Pacific region for 19th–25th Feb 2014 (16th–24th for global MOVE). (a) and (b) show the SLA and profile data coverage for reference. (c)–(n): top row shows T increments at 100m depth; bottom row shows T increments cross section along the equator between 180E–70W for each system. The numbers in brackets indicate the normalisation applied in the plot below in order to make the spatial root mean square of each field to be 1K.



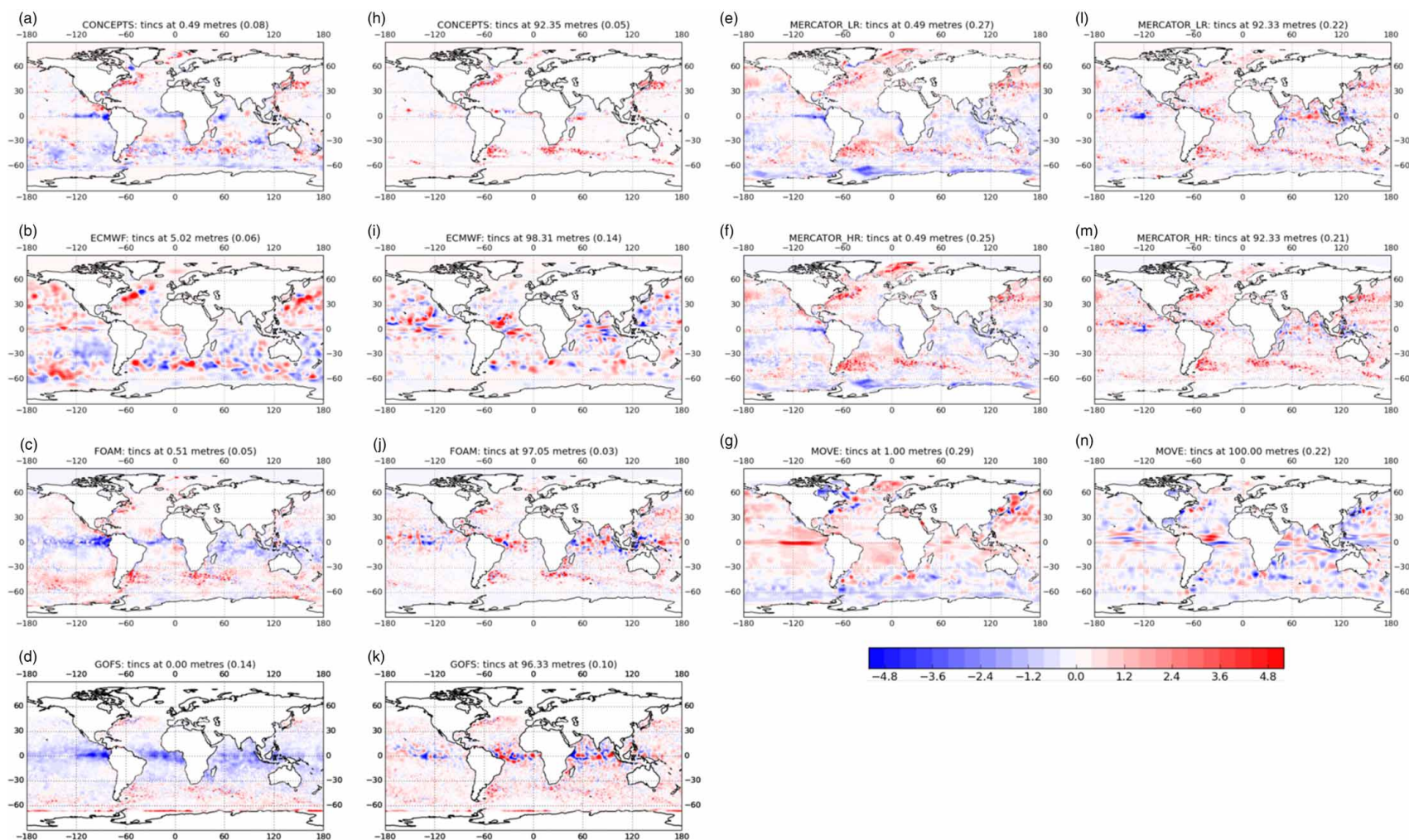


Figure 6. Normalised global temperature increments, averaged for February 2014. (a) (g) show T increments at the top model level; (h) (n) show T increments at 100 m depth. The numbers in brackets indicate the normalisation applied in the plot below in order to make the spatial root mean square of each field to be 1K.

mammals have very significant data impacts (Oke et al. 2015) and so are very valuable. Anything that can be done to speed up delivery of those data would enable better use to be made of them; ongoing work at the Sea Mammal Research Unit (SMRU) to investigate the possible use of Iridium communications for the marine mammal sensors may improve the situation in the future. It can also be seen from Figure 7 that the Argo data are not as timely as some of the other *in situ* platforms and, given the importance of Argo for operational ocean forecasting, anything which could be done to speed up the delivery of these data would lead to more use being made of them.

### ***Sensitivity of the Atlantic Meridional Overturning Circulation to data assimilation parameters***

One of the challenges for the systems which are used to generate ocean reanalyses such as the ECMWF ocean analysis system, ORAS4, is related to the assimilation of observations close to the coast, in particular in regions of strong western boundary currents (Stepanov et al. 2012). Figure 8 shows the time evolution of the Atlantic Meridional Overturning Circulation (AMOC) at 26°N in two different assimilation experiments, where the observation error standard deviations (OESD) by the coast have been

increased within a different distance from the coast (300 km and 800 km). The resulting AMOC shows a large sensitivity to this single parameter specifying the OESD of the coastal observations. When the coastal OESD is inflated within 800 km of the coast (DC 800 km, red line), the AMOC is substantially stronger than when it is inflated only within 300 km of the coast (DC 300 km). The spectrum of the inter-annual variability also changes in response to this parameter. This extreme sensitivity of large-scale climate indices highlights the importance of independent transport-derived estimations such as those from RAPID, which are essential to continue progress on data assimilation development.

### ***Improving assimilation near the equator***

Near the equator the dominant balance is between the sub-surface pressure gradient and the wind stress applied to the model. Assimilation of profile and altimeter data causes changes to the pressure gradient which results in imbalance to the unchanged wind stress. This could lead to the generation of spurious equatorial waves and vertical velocities. The pressure bias correction scheme used in FOAM (Bell et al. 2004) is designed to reduce the imbalance where it is consistent over the long term. However, we still see the

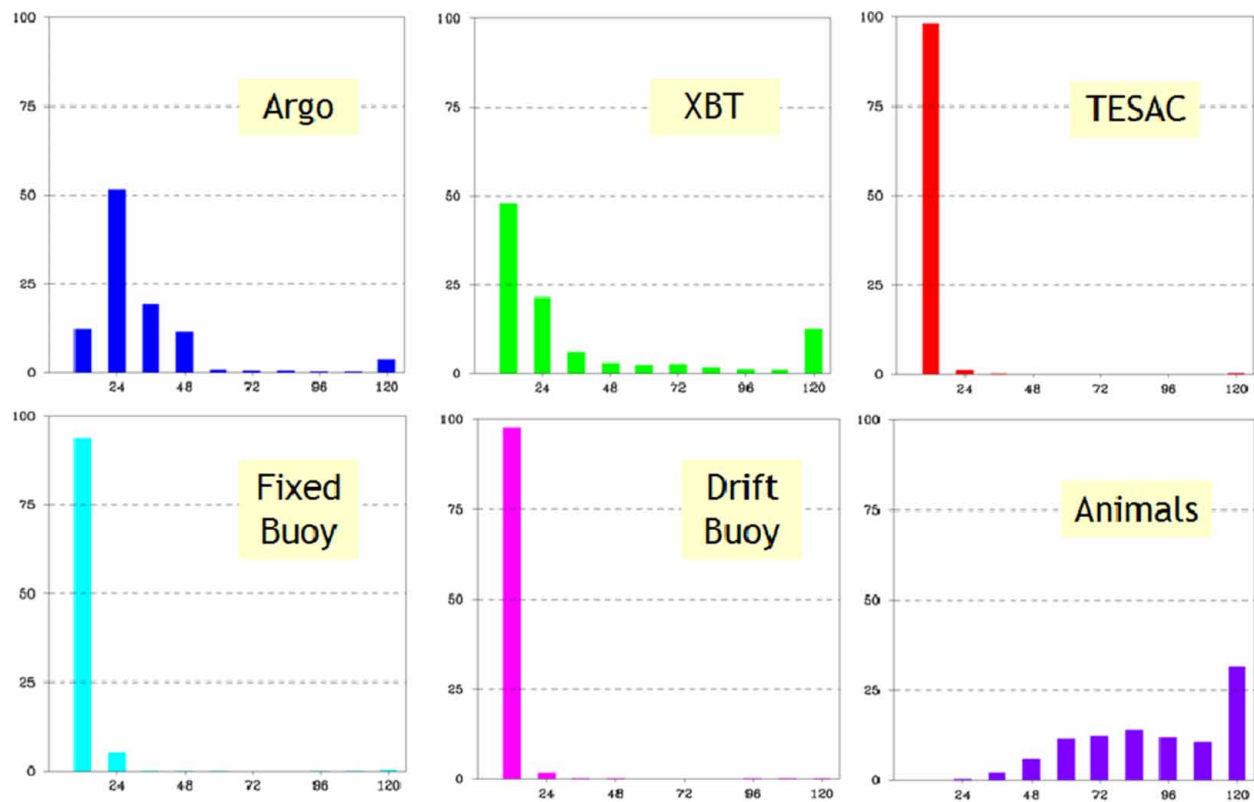


Figure 7. Differences between the time of receipt of an observation at NRL and its measurement time for profile data types. The results are presented in 12 hour bins out to 5 days behind the observed time, calculated over the period Sep–Nov 2012.

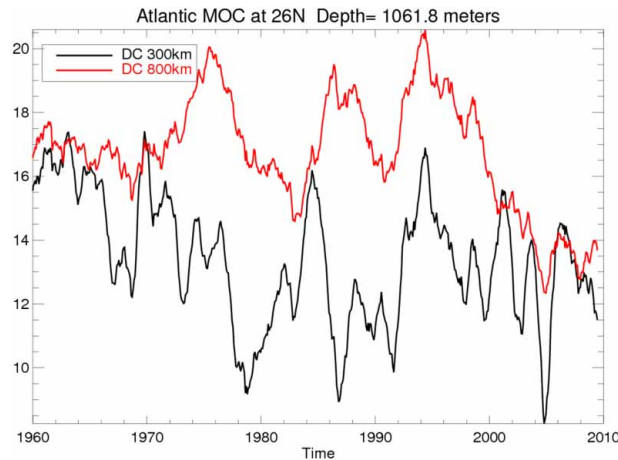


Figure 8. Atlantic Meridional Overturning circulation from two sensitivity assimilation experiments conducted with the same system as that of ORAS4, where the observation error is inflated for the observations within 800 Km (red) and 300 Km (black) of the coast.

impact of data assimilation on the equatorial currents and vertical velocities from shorter time-scale imbalances. In the pressure bias correction scheme the correction is calculated from accumulated temperature and salinity increments and can be considered a bias correction. The application of an instantaneous pressure correction is being investigated whereby the correction is calculated from the temperature and salinity increments from the current analysis step. In Figure 9 equatorial Atlantic sections of vertical velocity anomaly fields at the end of the 24 hour IAU step are presented. The anomalies are calculated as the IAU vertical velocities minus the vertical velocities from a free run starting from the same initial conditions. In these experiments temperature, salinity, SSH and SST observations are assimilated for one day. Figure 9 shows the vertical velocity anomalies with and without the instantaneous pressure correction. The vertical velocity anomaly is significantly reduced in the case with the instantaneous pressure correction. Work is ongoing to investigate how this correction should be ramped down away from the equator and how it should be applied with cycling of the data assimilation system.

#### **Improvement of high-frequency variability using 4DVar**

The Japan Meteorological Agency (JMA) plans to start operational use of a western North Pacific (WNP) four-dimensional variational (4DVar) data-assimilative model developed in the Meteorological Research Institute (MRI), as a part of a coastal monitoring and forecasting system in 2015. The grid spacing of the model is identical to the WNP part of NP MOVE: the horizontal resolution is  $0.1^\circ$  around Japan. This model applies a data assimilation scheme based on the 4DVar version of MOVE (Usui et al. submitted), (MOVE-4DVar). The model fields are

exploited for monitoring and relatively short-term ( $\sim 10$  days) ocean forecasting, as well as to initialize a finer-resolution (2 km) coastal model covering the area around the Seto Inland Sea (Kourafalou et al. 2015) using IAU.

The 4DVar scheme has the potential to improve reproducibility of high-frequency phenomena as demonstrated in Figure 10. Here, the assimilative model run with the MOVE- 4DVar scheme is compared with the run with the 3DVar version of MOVE (MOVE-3DVar), the current operational scheme. The length of each assimilation period is set to 10 days in both runs. High frequency variability of the Kuroshio path is reproduced with MOVE-4DVar, which induces a substantial improvement in the sea level variation at Hachijo Island (HJ) as illustrated in Figure 10(a). Figure 10(b) indicates that reproduction of the development of the cold eddy at  $138^\circ\text{E}$  causes the meandering of the Kuroshio path and the sea level rise close to the tide gauge observation at HJ in the 4DVar run. Improvements of high-frequency sea level variation appear not only in the open ocean area but also in the coastal areas, and increase the forecast accuracy (Usui et al. submitted). It also improves SSH fields in the coastal finer-resolution model initialized using the fields of the WNP assimilative model through IAU. These results indicate the need for further research into the use of more sophisticated data assimilation techniques for initialising high-frequency processes.

#### **Assessing error covariance specifications**

Differences between the observations and their model-estimated counterparts before and after assimilation (the innovations and analysis residuals, respectively) provide valuable information for assessing the performance of the data assimilation system. This information is also important for estimating error covariances, and for evaluating the



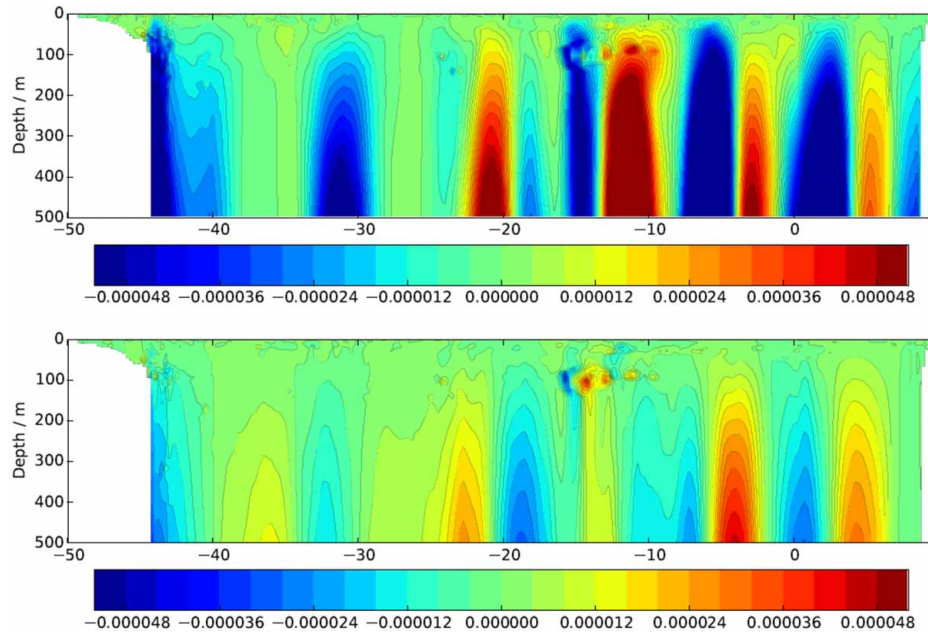


Figure 9. Vertical velocity differences (m/s) cross section along the equator in the Atlantic after 22.5 hours of IAU compared to a run which assimilates no data for, (a) a run with all data and (b) a run which uses an instantaneous pressure correction to balance the T/S increments.

consistency of the error covariances prescribed in the system (Desroziers et al. 2005). Figure 11 shows an example of a consistency check based on ‘Desroziers diagnostics’ applied to the ECMWF ORAS4 reanalysis (Balmaseda et al. 2013; Mogensen et al. 2012). Shown are 40-year time series of the globally averaged observation-space representations of the *specified* and *expected* temperature background error standard deviations (BESD) in the upper 50 m (blue and red curves, respectively).

The specified BESD are parameterized in terms of the vertical density stratification of the background state. Discrepancies between the specified and expected BESD are an indication of sub-optimality in the error covariance specifications. Seasonal variations in the specified and expected BESD are reasonably consistent. However, there is a noticeable decreasing trend in the expected BESD that is not present in the specified BESD. This trend roughly mirrors the increasing trend in the number of observations (black curve) and reflects the fact that the background state is becoming steadily more accurate, especially in the final decade as a result of the assimilation of Argo data. This important influence of the observation network on background error is not captured by the state-dependent parameterization used in ORAS4 and suggests that improved background error estimation methods are necessary. Methods which capture flow dependence in the background errors such as the EnKF or hybrid ensemble-variational methods are promising in this regard as discussed in the following section.

### Longer term perspectives for ocean data assimilation

A number of the operational groups are investigating improvements to the algorithms currently used. Some of the GOV systems provide initial conditions for seasonal forecasts for which purpose probabilistic forecasting is a necessity; short-range probabilistic forecasts could also improve the products for search and rescue, and oil spill applications. NRL, ECMWF and FOAM are developing the use of ensemble information within the variational framework using hybrid methods. The basic idea is to employ a perturbed model and perturbed data assimilation system to simulate the evolution of state errors in the system. The ensemble of states provides a flow-dependent estimate of analysis and forecast error, which can be used for initialising probabilistic forecasts and for estimating background error covariances on each data assimilation cycle. Due to the high cost of running high-resolution ocean models, only a small number of ensemble members will be affordable in practice. How to exploit limited ensemble information effectively and efficiently in the background error covariance formulation of variational assimilation is a particular challenge. One approach is to use the ensemble to estimate parameters, such as variances and correlation length scales, within existing error covariance models. Another approach is to use the ensemble to compute a sample estimate of the covariance matrix and to apply localization and filtering techniques to remove spurious covariances associated with sampling error. A third (hybrid) approach involves a weighted linear combination

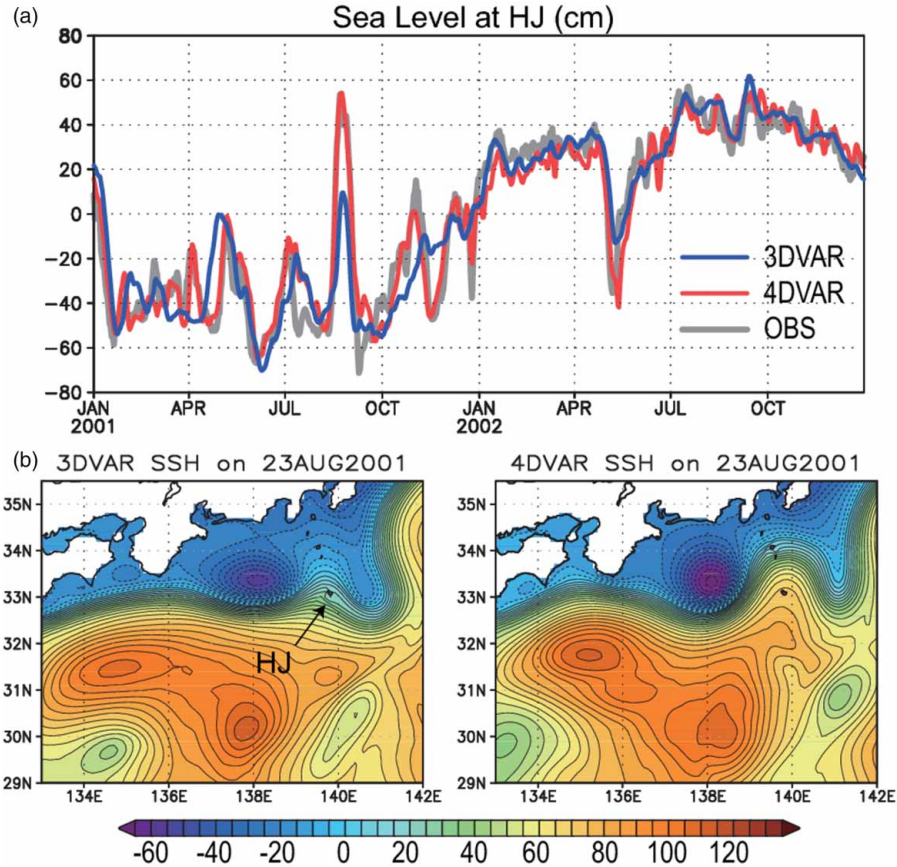


Figure 10. (a) Times series of the sea level at Hachijo Island (HJ) in the WNP assimilative model with MOVE 3DVar (blue line) and MOVE 4DVar (red line) schems and observed by the tide gauge (gray line). The correlation (RMS difference) between the time series from the simulation and the observation is 0.925 (14.05 cm) for MOVE 3DVar, and 0.968 (9.27 cm) for MOVE 4DVAR. (b) Horizontal distribution of SSH at 23 August 2001 in the model with MOVE 3DVar (left) and MOVE 4DVar (right). The position of HJ is indicated in the left panel.

of these two covariance formulations, with the aim of improving robustness and flexibility.

In order to produce the perturbations required to generate ensemble information, the deterministic forecast models can be transformed into stochastic models by including an explicit representation of model uncertainties in the

discretized equations (Brankart 2013). These uncertainties can be simulated using random processes representing unresolved fluctuations of the state variables. A recent version of the Nucleus for European Modelling of the Ocean (NEMO) code already includes an option for stochastic perturbations activated during the model

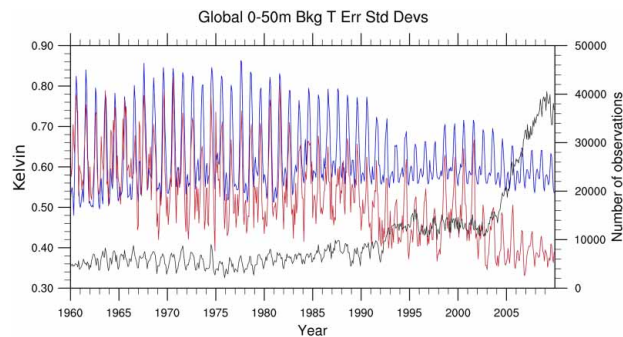


Figure 11. Time series of monthly averaged temperature background error standard deviations (BESD) in the ECMWF ORAS4 reanalysis: (1) the *specified* BESD in the assimilation system (blue curve); and (2) the *expected* BESD as determined using the Desroziers et al. (2005) method (red curve). The black curve shows the number of observations as function of time (right axis).



integration. Following this idea, stochastic ensemble models will be useful to propagate the relevant information about the ocean state for both probabilistic prediction purposes and data assimilation using “built-in” ensemble model errors.

Another way to improve the specification of errors in the variational framework is to use 4DVar whereby the error covariances are implicitly evolved by the model during the assimilation time window. JMA plans to start operational use of a WNP 4DVar assimilative model in 2015, as described in the previous section. Estimation of the tidal component in the system will be introduced in 5 years. JMA also aims to introduce the 4DVar scheme in the global ocean assimilative model in the seasonal forecasting system around 2020.

Mercator and CONCEPTS, who currently use a fixed-ensemble based approach, are planning to use ensemble information from the current cycle in the SEEK filter in order to make use of ‘errors of the day’. They also plan to develop schemes to improve the use of variables characterised by non-Gaussian distributions, which break the assumptions of the linear analysis. It also limits the quality of the multivariate extrapolation by introducing spurious effects. To overcome this short-coming in the statistical description, local anamorphic transformations restoring the Gaussian properties are being explored. The objective is to apply these methods to non-Gaussian variables such as sea-ice concentration (Brankart et al. 2012).

Keeping up with changes in the observing system (Le Traon et al. 2015; Clarke et al. 2015) is essential to maintain and improve the performance of operational systems. It is expected that wide-swath satellite altimetry data from Coastal and Ocean measurement Mission with Precise and Innovative Radar Altimeter (COMPIRA), and the Surface Water Ocean Topography (SWOT) missions, will offer effective information to better constrain the fine-scale ocean state in the GOV systems. Assimilating sea surface salinity based on measurements from satellites (e.g. Aquarius and SMOS) may also have the potential to improve near-surface salinity fields of GOV systems.

In the near future, it is likely that new satellite missions will deliver a mix of state variable measurements at pixel level, and structure information about the flow by means of images. Wide-swath altimetry such as SWOT, or ocean colour imager missions placed on geostationary orbits will deliver such information. A key challenge for ocean and coastal prediction systems is therefore to develop pattern assimilation methods that will provide additional constraints on the flow at higher resolution, and produce more accurate and physically consistent fields.

## Summary and conclusions

A description of the various operational ocean data assimilation systems contributing to GOV has been provided.

The components of the global ocean observing system available in near-real time are assimilated by all the systems with some differences in the way the data are accessed, and some differences in observation types for particular variables such as SST and sea-ice concentration. A variety of techniques is used to perform the data assimilation with different algorithms (3DVar, SEEK, EnOI, EnKF), different assimilation time-windows (from one day out to 10 days), and different procedures for applying the assimilation increments to the model.

An example of the outputs of the data assimilation has been provided by showing the changes calculated by each of the assimilation systems (increments) for February 2014. This illustrates the types of structures that the assimilation systems are correcting in the models, as well as the multi-variate nature of the assimilation systems and their ability to propagate information from sparse observations in space and time. In the EAC and Gulf Stream regions, a number of similar features are seen in the eddy-permitting GOV systems’ increments, indicating similar forecast errors are being corrected by each of the data assimilation systems. In the Kurshio region the forecast errors appear to be less systematic between the systems, although the scales of the structures are similar. The eddy resolving systems of GOFS and Mercator HR represent smaller-scale structures in the mid-latitude regions investigated and appear to have smaller biases, particularly in the Gulf Stream separation region. The ECMWF and MOVE systems, which have longer time-windows and lower resolution, have larger-scale increments everywhere.

Monthly average temperature increments show significant SST biases, particularly evident in GOFS and FOAM which assimilate swath satellite SST data, indicating systematic errors in the surface heat fluxes and the way in which they are propagated vertically by the ocean models. They also indicate that the ocean models are not fully representing a developing negative SST anomaly in the tropical Pacific during the period investigated, and the assimilation acts to correct this model bias.

The systems are continually being developed and some of the on-going developments have been described. Operational issues with observation timeliness are being addressed. Research issues currently being investigated include the assimilation of data to improve representation of the Atlantic Meridional Overturning Circulation, assessing the suitability of the error covariance specifications, and dealing with issues in the assimilation of data near the equator.

Longer term developments to the systems are expected to include the implementation of more advanced algorithms (such as hybrid variational-ensemble schemes, or 4DVar schemes), improvements to the error covariance representations and assimilation of higher resolution data such as might become available from wide-swath altimetry missions. Many of these challenges are common to all the

systems, which highlights the importance of a forum to discuss ideas and results, such as is provided by GODAE OceanView.

## References

- Balmaseda M, co authors. 2015. Intercomparing ocean estimates during the last decades: new approaches & results in the framework of the joint CLIVAR/GSOP GOV effort. *J Oper Oceanogr.* doi:10.1080/1755876X.2015.1022329
- Balmaseda MA, Dee D, Vidard A, Anderson DLT. 2007. A multivariate treatment of bias for sequential data assimilation: application to the tropical oceans. *Q J R Meteorol Soc.* 133:167–179, doi:10.1002/qj.12.
- Balmaseda MA, Mogensen K, Weaver AT. 2013. Evaluation of the ECMWF ocean reanalysis system ORAS4. *Q J R Meteorol Soc.* 139:1132–1161, doi:10.1002/qj.2063.
- Bell MJ, Martin MJ, Nichols NK. 2004. Assimilation of data into an ocean model with systematic errors near the equator. *Q J R Meteorol Soc.* 130:873–893, doi:10.1256/qj.02.109.
- Benkiran M, Greiner E. 2008. Impact of the incremental analysis updates on a real time system of the North Atlantic Ocean. *J Atmos Ocean Technol.* 25:2055–2073.
- Blockley EW, Martin MJ, McLaren AJ, Ryan AG, Waters J, Lea DJ, Mirouze I, Peterson KA, Sellar A, Storkey D. 2014. Recent development of the Met Office operational ocean forecasting system: an overview and assessment of the new Global FOAM forecasts. *Geosci Model Dev.* 7:2613–2638, doi:10.5194/gmd 7 2613 2014.
- Bloom SC, Takacs LL, da Silva AM, Ledvina D. 1996. Data assimilation using incremental analysis updates. *Mon Wea Rev.* 124:1256–1271.
- Brankart JM, Testut CE, Béal D, Doron M, Fontana C, Meinvielle M, Brasseur P, Verron J. 2012. Towards an improved description of ocean uncertainties: effect of local anamorphic transformations on spatial correlations. *Ocean Sci.* 8:121–142, doi:10.5194/os 8 121 2012.
- Brankart JM. 2013. Impact of uncertainties in the horizontal density gradient upon low resolution global ocean modeling. *Ocean Model.* 66:64–76.
- Brassington GB. 2013. Multicycle ensemble forecasting of sea surface temperature. *Geophys Res Lett.* 40:6191–6195, doi:10.1002/2013GL057752.
- Clarke C, co authors. 2015. Status & future of the in situ ocean observing system. *J Oper Oceanogr.* forthcoming
- Cummings J, Bertino L, Brasseur P, Fukumori I, Kamachi M, Martin MJ, Mogensen K, Oke P, Testut CE, Verron J, Weaver A. 2009. Ocean data assimilation systems for GODAE. *Oceanography.* 22:96–109.
- Cummings JA, Smedstad OM. 2013. Park S.K, Xu L. Variational data assimilation for the global ocean. Data assimilation for atmospheric, oceanic and hydrologic applications. II, doi:10.1007/978 3 642 35088 713.
- Daley R. 1991. Atmospheric data analysis. Cambridge University Press.
- Desroziers G, Berre L, Chapnik B, Poli P. 2005. Diagnosis of observation, background and analysis error statistics in observation space. *Q J R Meteorol Soc.* 131:3385–3396, doi:10.1256/qj.05.108.
- Evensen G. 2009. Data assimilation: the ensemble Kalman filter 2<sup>nd</sup> Edition. Springer.
- Fujii Y, Kamachi M. 2003. Three dimensional analysis of temperature and salinity in the equatorial Pacific using a variational method with vertical coupled temperature salinity EOF modes. *J Geophys Res.* 108:3297.
- Hernandez F, Blockley E, Brassington GB, Davidson F, Divakaran P, Drévillon M, Ishizaki S, Garcia Sotillo M, Hogan PJ, Lagemaa P, Levier B, Martin M, Mehra A, Mooers C, Ferry N, Ryan A, Regnier C, Sellar A, Smith GC, Sofianos S, Spindler T, Volpe G, Wilkin J, Zaron ED, Zhang A. 2014. Performance evaluations, near real time assessment of operational oceanography forecast products. *J Oper Oceanogr.* forthcoming.
- Kourafalou, co authors. 2015. Status and future of regional ocean forecasting systems. *J Oper Oceanogr.* forthcoming
- Le Traon P Y, co authors. 2015. Status & future of the remote sensing ocean observing system. *J Oper Oceanogr.* forthcoming
- Lea DJ, Drecourt J P, Haines K, Martin MJ. 2008. Ocean altimeter assimilation with observational and model bias correction. *Q J R Meteorol Soc.* 134:1761–1774, doi:10.1002/qj.320.
- Lellouche J. M, Le Galloudec O, Drévillon M, Régnier C, Greiner E, Garric G, Ferry N, Desportes C, Testut C. E, Bricaud C, Bourdallé Badie R, Tranchant B, Benkiran M, Drillet Y, Daudin A, De Nicola C. 2013. Evaluation of global monitoring and forecasting systems at Mercator Océan. *Ocean Sci.* 9:57–81, doi:10.5194/os 9 57 2013.
- Mirouze I, Weaver AT. 2010. Representation of correlation functions in variational assimilation using an implicit diffusion operator. *Q J R Meteorol Soc.* 136:1421–1443.
- Mogensen K, Alonso Balmaseda M, Weaver A. 2012. The NEMOVAR ocean data assimilation system as implemented in the ECMWF ocean analysis for System 4. ECMWF Tech Memo no 668, ECMWF, Reading, UK.
- Oke PR, co authors. 2015. Assessing the impact of observations on ocean forecasts and reanalyses: part 2: regional applications. *J Oper Oceanogr.* doi:10.1080/1755876X.2015.1022080
- Oke PR, Sakov P. 2008. Representation error of oceanic observations for data assimilation. *J Atmos Ocean Technol.* 25:1004–1017.
- Oke PR, Sakov P, Cahill ML, Dunn JR, Fiedler R, Griffin DA, Mansbridge JV, Ridgway KR, Schiller A. 2013. Towards a dynamically balanced eddy resolving ocean reanalysis: BRAN3. *Ocean Model.* 67:52–70.
- Ricci S, Weaver AT, Vialard J, Rogel P. 2005. Incorporating state dependent temperature salinity constraints in the background error covariance of variational ocean data assimilation. *Mon Wea Rev.* 133:317–338.
- Sakov P, Counillon F, Bertino L, Lisæter KA, Oke P, Korabely A. 2012. TOPAZ4: an ocean sea ice data assimilation system for the North Atlantic and Arctic. *Ocean Science.* 8:633–656.
- Sandery PA, Brassington GB, Freeman J. 2011. Adaptive nonlinear dynamical initialization. *J Geophys Res.* 116:C01021.
- Smith GC, Roy F, Reszka M, Surcel Colan D, He Z, Deacu D, Belanger J M, Skachko S, Liu Y, Dupont F, Lemieux J F, Beaudoin C, Tranchant B, Drévillon M, Garric G, Testut C E, Lellouche J M, Pellerin P, Ritchie H, Lu Y, Davidson F, Buehner M, Lajoie M, Caya A. 2014. Sea ice forecast verification in the Canadian Global Ice Ocean Prediction System. *Q J R Meteorol Soc.* submitted.
- Stepanov V, Haines K, Smith GC. 2012. Assimilation of RAPID array observations into an ocean model. *Q J R Meteorol Soc.* 138:2105–2117, doi:10.1002/qj.1945.
- Tonani M, co authors. 2015. Status and future of global and regional ocean prediction systems. *J Oper Oceanogr.* forthcoming
- Tsujino H, Hirabara M, Nakano H, Yasuda T, Motoi T, Yamanaka G. 2011. Simulating present climate of the global ocean ice system using the Meteorological Research Institute Community Ocean Model (MRI.COM): simulation

- characteristics and variability in the Pacific sector. *J Oceanogr.* 67:449–479, doi:10.1007/s10872-011-0050-3.
- Usui N, Fujii Y, Sakamoto K, Kamachi M. 2014. Improved representation of mesoscale and coastal variability with a four dimensional variational assimilation system. To be submitted to *Mon. Wea. Rev.*
- Usui N, Ishizaki S, Fujii Y, Tsujino H, Yasuda T, Kamachi M. 2006. Meteorological Research Institute Multivariate Ocean Variational Estimation (MOVE) System: some early results. *Adv Space Res.* 37:806–822.
- Waters J, Lea DJ, Martin MJ, Mirouze I, Weaver A, While J. 2014. Implementing a variational data assimilation system in an operational 1/4 degree global ocean model. *Q J R Meteorol Soc.* doi:10.1002/qj.2388.
- Weaver A, Courtier P. 2001. Correlation modelling on the sphere using a generalized diffusion equation. *Q J R Meteorol Soc.* 127:1815–1846, doi:10.1002/qj.49712757518.
- Zaron ED. 2011. Introduction to ocean data assimilation in operational oceanography in the 21st Century, Eds. Schiller, Brassington, p. 321–350.