Ocean Observing System Evaluation

Peter R. Oke\textsuperscript{1}, Magdalena A. Balmaseda\textsuperscript{2}, Mounir Benkiran\textsuperscript{3}, James A. Cummings\textsuperscript{4}, Eric Dombrowsky\textsuperscript{3}, Yosuke Fujii\textsuperscript{5}, Stephanie Guinehut\textsuperscript{6}, Gilles Larnicol\textsuperscript{6}, Pierre-Yves Le Traon\textsuperscript{7}, Matthew J. Martin\textsuperscript{8}

\textsuperscript{1}CAWCR/CSIRO and WfO National Research Flagship, Hobart, Tasmania, Australia
\textsuperscript{2}European Centre for Medium-Range Weather Forecasting, Reading, United Kingdom
\textsuperscript{3}Mercator Ocean, Ramonville Saint Agne, France
\textsuperscript{4}Oceanography Division, Naval Research Laboratory, Monterey, USA
\textsuperscript{5}Oceanographic Research Department, Meteorological Research Institute, Tsukuba, Japan
\textsuperscript{6}CLS-Space Oceanography Division, Toulouse, France
\textsuperscript{7}Ifremer, Technopole de Brest-Iroise, Plouzane, France
\textsuperscript{8}Met Office, Exeter, United Kingdom

Corresponding author: Peter Oke, CSIRO Marine and Atmospheric Research, GPO Box 1538, Hobart 7001, Tasmania, Australia (Peter.Oke@csiro.au)
Abstract

Global ocean forecast systems, developed under the Global Ocean Data Assimilation Experiment (GODAE), can be used to assess the impact of different components of the Global Ocean Observing System (GOOS). GODAE systems can be exploited to help identify observational gaps and to ultimately improve the efficiency and effectiveness of the GOOS for constraining ocean models for ocean prediction and reanalysis. Many tools are currently being used by the GODAE community to evaluate the GOOS. Observing System Experiments where different components of the GOOS are systematically withheld, can help quantify the extent to which the skill of a model depends on each observation type. Various other techniques, including observing system simulation experiments, adjoint- and ensemble-based approaches, can be used to aid the design and evaluation of ocean observing systems. A suite of examples using these methods to evaluate the GOOS from a GODAE perspective are presented in this paper. Also included in this paper is a proposed plan to move these activities towards routine monitoring of the GOOS using operational GODAE systems.
1. Introduction

The development of operational ocean forecast systems is a key initiative of the Global Ocean Data Assimilation Experiment (GODAE). All GODAE systems are underpinned by the Global Ocean Observing System (GOOS; www.ioc-goos.org) that is comprised of satellite altimetry, satellite sea surface temperature (SST) programs, delivered through the GODAE High Resolution SST effort (GHRSSST; www.ghrsst-pp.org), and in situ measurements from the Argo program, the tropical moored buoy, surface drifting buoys, XBT and tide gauge networks. Each of these observation programs are expensive and require a significant international effort to implement, maintain, process and disseminate. While many components of the GOOS are primarily intended for climate applications, their application to operational ocean forecast systems is important. In this paper, we present results from analyses that seek to assess the benefits of different observation types and arrays to realistic ocean forecast and reanalysis systems using Observing System Experiments (OSEs) and Observing System Simulation Experiments (OSSEs).

OSEs generally involve the systematic denial of different observation types from a data assimilating model in order to assess the degradation in quality of a forecast or analysis when that observation type is not used. Importantly, the impact of each observation depends on the details of the model, assimilation method and error estimates employed.

OSSEs often involve some sort of twin experiment, where a model is sampled in a way that resembles real observations, and those observations are assimilated into an alternative model. Similarly, ensemble- and adjoint-based methods for observing system design and assessment, often do not use real observations, but instead diagnose properties of a model to identify regions of high sensitivity and influence. These types of analyses, though idealised, may be used to
assess the impact of hypothetical observations that may not exist yet, and therefore contribute to the design of future observing systems.

The inaugural Ocean Observing Panel for Climate (OOPC) - GODAE meeting on OSSEs and OSEs was held at UNESCO/IOC in Paris, France in November 2007 (www.godae.org/OSSE-OSE.html). This was the first international meeting dedicated to the subject of observing system evaluation using GODAE systems. Many of the ideas and results presented in this paper are based on presentations from the OOPC-GODAE OSSE/OSE meeting. Other recent reviews that provide an assessment of the GOOS for constraining data assimilating ocean models include Oke et al. (2008; 2009) and Heimbach et al. (2009).

2. Observing System Experiments

A determination of the requirements of the GOOS for operational oceanography is the primary goal of the studies described in this section. Collectively, we seek to assess the importance of different observation types for meeting the needs of operational systems, including observation-based mapping systems, like that of CLS/Aviso, short-range prediction systems from GODAE partners (e.g., Bluelink, Mercator, NRL, UK Met Office, TOPAZ), and seasonal prediction (e.g., ECMWF, JMA, POAMA).

a. Number of altimeter missions

The Ssalto/Duacs center has led several studies aiming to identify the most appropriate satellite configuration to observe the mesoscale ocean. Focusing first on the Mediterranean Sea, and later on the global oceans, Pascual et al. (2006; 2007) have demonstrated the benefits of merging data from four altimeter missions to produce high resolution maps of sea level anomalies (SLA). For example, Pascual et al. (2006) show that in areas of intense variability, the root-mean-squared (RMS) differences between a classical configuration of two altimeters (Jason-1+ERS2/Envisat) and the scenario merging data from four altimeters can reach 10 cm for SLA and 400 cm^2/s^2 for
EKE (derive from SLA-based estimates of geostrophic velocities). This represents a significant percentage of the signal variance. At mid- and high-latitudes, previous studies have also shown a clear underestimation of EKE due to the under-estimated high frequency and high wavenumber signals produced when data from only two altimeters are used (Ducet et al. 2000; Le Traon and Dibarboure 2002; Brachet et al. 2004).

The impact of four altimeters is expected to be particularly important for operational forecast and analysis systems. Pascual et al. (2008) quantify the degradation in the quality of the altimeter products when Near-Real-Time (NRT) data are used compared to when Delayed-Time (DT) data are used. Three main sources of errors are identified in NRT data: the orbit is less accurate; the latency of data is a problem; and observation windows necessarily favour “old” data for NRT systems. Validation with independent in-situ data demonstrates the degradation of NRT maps compared to DT maps (Table 1). This shows that 4 altimeters in NRT are needed to get the same performance as 2 altimeters in DT. The statistics in
Table show comparisons between SLA from tide gauges and SLA maps. Table 1 includes results using an old and new DT data set and demonstrates the importance of continuous advances in the processing of altimeter data.

A series of OSEs, using the Mercator Ocean forecasting system in the North Atlantic Ocean and the Mediterranean Sea has been conducted by Benkiran et al. (2008) to evaluate the impact of data from multiple altimeter missions on the forecast skill over 7-days. This system assimilates along-track altimeter data, SST and in situ profiles using a multivariate OI scheme. Specifically, Benkiran et al. (2008) sought to assess the degradation in the forecast skill when the number of altimeters is varied. They performed several 6-month simulations in which they assimilated all available SST and T/S profiles and altimeter data from 0 to 4 altimeters (T/P, Jason-1, Envisat and GFO). The OSEs were conducted during the tandem T/P and Jason-1 missions in 2004-2005 when data from 4 altimeters were available. Figure 1 summarises their results, showing the degradation of the system skill, when data from 0, 1, 2 and 4 altimeters are assimilated. The estimated degradation, presented as a percentage of the observed variability, is relative to an OSE that assimilates data from 3 altimeters (Jason-1, Envisat and GFO; so positive degradation is worse and negative is better than 3 altimeters). When no altimeter data are assimilated, there is effectively no predictive skill at the mesoscale. Conversely, Figure 1 suggests that some skill is added to the Mercator system when data from 4 altimeters are assimilated, instead of just 3. These results are consistent with those presented by Pascual et al. (2006). Clearly, the addition of the first altimeter has the greatest impact on forecast skill – and there are diminishing returns from each additional altimeter. However, the benefits of additional altimeters are likely to be at smaller and smaller scales, as higher spatial and temporal resolution is resolved. We note that these small mesoscale features are important for many end-users of GODAE products (e.g., search and rescue, oil spill mitigation and so on). Benkiran et al. (2008) conducted their OSEs in
a real-time context. That is, they performed OSEs to produce nowcasts under realistic conditions, excluding missing data due to latency of data availability. They produce 7-day forecasts that are initialised with each nowcast and also hindcasts, using all available data. Table 2 summarises their results, showing that if only SST and in situ T/S are assimilated (i.e., no altimetry) the error is large (up to ~13cm RMS). Note also that to obtain error levels equivalent to the hindcast with only one altimeter, data from 4 (2) altimeters are needed to produce forecast (nowcasts) of equivalent skill under realistic conditions. These results are consistent with the conclusions of Pascual et al. (2008) who found that 4 altimeters in NRT is equivalent to 2 altimeters in DT.

Results from a series of OSEs designed to assess the impact of different numbers of altimeters using the UK Met Office system are presented in Figure 2. They use the 1/9° North Atlantic FOAM configuration together with an OI-based method of assimilation (Martin et al 2007), and run a series of three month integrations beginning in January 2006. The impact of different numbers of altimeters are assessed by comparing the modelled SLA with the assimilated along-track altimeter data, and comparing the modelled surface velocities with those derived from surface drifting buoys (which are not assimilated). These results quantify the improvements when 1, 2 and 3 altimeters are added to the assimilated observations. The addition of the first altimeter seems to have the most impact. The results are different for different regions; surface velocities in the north-east Atlantic are better than surface velocities in the north-west Atlantic. This indicates that the mesoscale dynamics in the north-east are better constrained by the altimeters than in the north-west. The difference in the quality of surface winds in different regions probably has a significant influence on the quality of the modelled surface velocities.

b. Impact of different data types

Using the Bluelink forecast system, Oke and Schiller (2007) performed a series of OSEs to compare the relative impact of Argo, SST and SLA observations on an eddy-resolving ocean
reanalysis. They systematically with-held altimeter, Argo and SST observations. Their results highlight the complimentary nature of the different observation types. For example, satellite SST observations are the only observation type considered that have the potential to constrain the circulation in shallow seas and over wide continental shelves; altimetry is the only observation type that even goes close to constraining the mesoscale ocean circulation; and Argo observations are the only observation type that constrains sub-surface temperature and salinity. Their results indicate that while there is some redundancy for representing broad-scale circulation, all observation types are required for constraining mesoscale circulation models.

The impact of the different components of the GOOS on ECMWF seasonal forecast system has been assessed through a series of OSE studies (Vidard et al. 2007; Balmaseda et al. 2007; Balmaseda and Anderson 2009). Vidard et al. (2007) focussed on the relative impact of the tropical in situ mooring arrays, XBTs and Argo observations for a period when Argo array was incomplete, and when altimeter data was not assimilated. Balmaseda et al. (2007), used a letter version of the system (Balmaseda et al. 2008) that assimilated both salinity and altimeter data are assimilated and showed the significant positive impact of Argo observation. In the most recent series of OSE experiments using the ECMWF system, Balmaseda and Anderson (2009) assess the relative contribution of Argo, altimeter and moorings to the skill of seasonal forecast through a series of OSEs. The results demonstrate that Argo, altimeter and mooring observations contribute to the improvement of the skill of seasonal forecasts of SST. For example, they demonstrate that assimilation of Argo observations are particular beneficial to SST forecasts in the eastern tropical Pacific, altimeter data are particularly beneficial to the central Pacific and the north subtropical Atlantic and that mooring data have a significant positive impact on forecast skill across the entire tropical Pacific. The positive impacts of Argo and mooring data on the
forecast skill of SST in seasonal forecasts are also confirmed in JMA's system (Fujii et al. 2008b).

A series of OSEs using the Global Observed Ocean Products (Larnicol et al. 2006) that combine remotely-sensed (SLA, SST) and in situ observations, using the method described by Guinehut et al. (2004), facilitates a quantitative assessment of the relative contributions from different components of the GOOS. Figure 3 shows the RMS errors of sub-surface temperature (T) and salinity (S) using this approach. This demonstrates that more than 40% of the temperature signal can be reconstructed at depth from remotely-sensed data using a simple statistical method and that the complementary use of in situ measurements (denoted combined fields in Figure 3) improves the estimation by an additional 10-20%.

3. Observing System Simulation Experiments

The potential impact of the assimilation of remotely sensed sea surface salinity (SSS) observation from SMOS or Aquarius on the forecast skill of the Mercator Ocean system has been assessed by Tranchant et al. (2008) through a series of OSSEs. They conclude that the level of observation error will have a critical impact on the value of this new observation type to GODAE systems. This is consistent with those of Brassington and Divakaran (2009) who assessed the theoretical impact of SSS observations on an ensemble-based data assimilation system.

Several different techniques have been used together with GODAE systems to contribute to the design of ocean observation programs. These include OSSEs that assess specific pre-determined design options (e.g., Guinehut et al. 2002; Schiller et al. 2004) and techniques that objectively generate “optimal” observation arrays. The latter includes Kalman filter techniques (e.g., Ballabera-Poy et al. 2007), ensemble approaches (e.g., Sakov and Oke 2008) and adjoint and representer-based methods (e.g., Vecchi and Harrison 2007; Fujii et al. 2008a; Le Henaff et al. 2009). Some of the studies referred to above have contributed to the design to assessment of the
Argo array; some have assessed the design of tropical mooring arrays; and others have identified regions that may help constrain model variability in western boundary currents. OSSE activities, while often somewhat theoretical, have contributed to discussions of the design of oceanographic observation programs.

4. Emerging Techniques

To date, observing system evaluation activities conducted under GODAE, and related programs, have typically employed conventional methods including OSEs and OSSEs, as described above. These activities have been designed to assess the limitations of the GOOS for GODAE applications (including forecast, reanalysis and analysis systems). These have typically involved OSEs that are performed several years after observations are collected (e.g., during periods when data from 4 altimeters were available and when the Argo program was still incomplete). However, we recognise that the GOOS is constantly changing. The significance of the completed OSEs is therefore increasingly irrelevant to the observational community. To have a real impact, the GODAE OceanView community is collectively shifting their efforts to transition their OSE/OSSE activities towards routine monitoring of the GOOS. Some initial steps have been taken to coordinate these activities. Specifically, agreement is sought on how GODAE partners can and should move towards routine monitoring of the GOOS; agreement on how this can be coordinated between the international groups; and a staged plan for moving these activities towards routine monitoring, so that the GODAE OceanView community can have a real impact on the ongoing design and assessment of the GOOS.

Emerging techniques under consideration by GODAE include analysis and forecast sensitivity experiments. These represent diagnostics from analysis and forecasts systems that are relatively inexpensive to compute. Analysis sensitivity experiments seek to quantify the impact of each
individual observation on an analysis (Cardinali et al. 2004). Similarly, an adjoint technique can quantify the sensitivity of a forecast to assimilated observations (Langland and Baker 2004).

Diagnostics derived from analysis sensitivity include the information content (IC) of each observation and the degrees of freedom of signal (DFS). These quantify the impact of each observation on an analysis, given the assumed errors, length-scales etc, in the data assimilation being used. A preliminary example of the IC and DFS for different observation types on the Bluelink reanalysis system (Oke et al. 2008) is given in Figure 4. Based on these results, it appears that both altimetry and SST observations are well used by the Bluelink system. However, information from the Argo data is either not extracted by the Bluelink system in an optimal way, or is somewhat redundant – possibly well represented by the other assimilated observations. At this stage of development, the former explanation seems most likely. By producing these, and other, diagnostics from a number of GODAE systems, it is anticipated that the true value of all observations for GODAE systems can be routinely monitored and quantified.

In turn, these evaluations could be fed back to the broader community for consideration.

5. Conclusions

The purposes of this paper is to summaries the OSE/OSSE activities conducted under GODAE, and related programs, to document some of the key results and to describe how these activities may progress under GODAE OceanView in the future.

One recurring result from different OSEs includes the apparent complimentary nature of different observation types (e.g., Guinehut et al. 2004; Larnicol et al. 2006; Oke and Schiller 2007; Balmaseda et al. 2008b). This means that none of the observation types in the GOOS is redundant. Each different observation type brings unique contributions to the GOOS and all observation types should be routinely assimilated by forecast and reanalysis products; and more importantly maintained by the international community.
Another result that is common to many studies is the necessity of assimilation of altimeter data to represent mesoscale variability (e.g., Oke and Schiller 2007; Martin et al. 2007; Pascual et al. 2008; Benkarin et al. 2008). Moreover, a couple of studies demonstrated that for NRT applications data from 4 altimeters is needed to obtain errors that are comparable to systems using 2 altimeters in delayed-mode (e.g., Pascual et al. 2008; Benkiran et al. 2008).

Several studies have demonstrated the importance of Argo observations. These include several OSE and OSSE studies based on analysis systems (e.g., Guinehut et al. 2004; Larnicol et al. 2006) and OSEs based on both short-range and seasonal prediction systems (e.g., Oke and Schiller 2007; Balmaseda et al. 2008b; Fujii et al. 2008b). Several of these studies specifically noted that Argo is the only observation platform that provides global-scale information for constraining salinity.

All GODAE forecast systems considered in this paper include SST observations as an essential core data set. Indeed, one could argue that in many coastal regions and shallow seas, SST is the only observation type that adequately monitors ocean properties. The consistent uptake of SST observations is a credit to the GHRSST program that provides high level quality controlled SST data in NRT.

The versatility of OSSEs and variational data assimilation techniques are also demonstrated in this paper, where it is shown that insight into observing strategies for resolving specific processes, like the Kuroshio meander (Fujii et al. 2008b), and specific time-scales of variability (e.g., Sakov and Oke 2008) can be gained. The impact of new observation types, like surface salinity observations, has also been assessed, with promising results (Tranchant et al. 2008; Brassington and Divakaran 2009).

We note that many groups from the NWP community routinely provide statistics on data impacts; in some cases - every day for every assimilation cycle. The methods discussed in
sections 2 and 3 of this paper (OSE and OSSEs) are very expensive - and as a result are not applied routinely. They are also, arguably, of limited value. For example, they will not automatically identify the impacts of changes in the Argo array – as the total number of Argo floats fluctuates and their spatial distribution changes. By contrast, as the NWP community have demonstrated, the routine application of computationally efficient methods, such as those referred to in section 4 can readily be applied to operational systems in NRT – and can potentially support the maintenance and development of the GOOS on an ongoing basis. Following the lead of the NWP community, during the new sustained phase of GODAE, so called GODAE OceanView, a coordinated effort is planned for OSE/OSSE activities to move towards the routine monitoring of the GOOS using GODAE systems.
Acknowledgement

The participants of the OOPC-GODAE meeting on OSSEs and OSEs, UNESCO/IOC, Paris, France in November 2007 are gratefully acknowledged for their contributions to this work. The following organisations and agencies are acknowledged for their financial support: Bureau of Meteorology; CSIRO; European Space Agency; French Space Agency (CNES); French Service Hydrographique Océanographique de la Marine (SHOM); Japan Meteorological Agency; the Japanese Ministry of Education, Culture, Sports, Science and Technology; Office of Naval Research; and the Royal Australian Navy. Satellite altimetry is provided by NASA, CNES and NOAA and SST observations are provided by NOAA. Argo data are provided by the Coriolis and USGODAE data centres.

References


Table 1: RMS difference (in cm) between tide gauge sea-level and mapped altimetry for the old delayed-time, the new delayed-time and the near-real-time system; adapted from Pascual et al. (2008). Comparisons are for the period October 2002 - August 2003.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Old Delayed-Time</th>
<th>New Delayed-Time</th>
<th>Near-Real-Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 missions</td>
<td>4.72</td>
<td>4.26</td>
<td>4.82</td>
</tr>
<tr>
<td>4 missions</td>
<td>4.27</td>
<td>3.94</td>
<td>4.42</td>
</tr>
</tbody>
</table>

Table 2: RMS of the difference between Jason-1 observation and 7-day forecast, Nowcast (real-time analysis) and hindcast (best analysis) for several OSEs where altimeter data from 0, 1, 2, 3 and 4 satellites are assimilated in addition to in situ T/S profiles and SST; adapted from Benkiran et al. (2008).

<table>
<thead>
<tr>
<th>SLA RMS difference</th>
<th>No altimetry</th>
<th>Jason-1 only + Envisat</th>
<th>Jason-1 + Envisat + GFO</th>
<th>Jason-1 + Envisat + GFO + T/P</th>
</tr>
</thead>
<tbody>
<tr>
<td>7-day forecast (cm)</td>
<td>10.27</td>
<td>9.67</td>
<td>8.95</td>
<td>8.62</td>
</tr>
<tr>
<td>Nowcast (cm)</td>
<td>9.15</td>
<td>8.36</td>
<td>7.50</td>
<td>7.08</td>
</tr>
<tr>
<td>Hindcast (cm)</td>
<td>12.94</td>
<td>8.38</td>
<td>7.07</td>
<td>6.18</td>
</tr>
</tbody>
</table>
Figure 1: Normalised measure of 7-day forecast error in the North Atlantic, when no altimeter data are assimilated and when data from 1, 2 and 4 altimeters are assimilated. Forecast skill is measured against the forecast error when Jason+Envisat+GFO data are assimilated (REF; see equation (1)). A positive % implies a degradation of the forecast skill, 0 is the baseline and negative means an improvement; adapted from Benkiran et al. (2008).

Figure 2: Anomaly correlation between forecast (left) SLA and along-track altimetric SLA from all satellites and (right) forecast near-surface velocity and near-surface velocity derived from drifting buoys; based on a series of OSEs that assimilate SLA data from 0-3 satellites, using the 1/9° North Atlantic FOAM configuration for the first 3 months of 2006.
Figure 3: RMS (solid lines) and mean (dotted lines) error in predicting sub-surface temperature (left) and S (right) anomalies using Levitus monthly mean climatology (red), synthetic fields (blue), combined fields (green); adapted from Larnicol et al. (2006).

Figure 4: Preliminary estimates of the Information Content (IC; %), degrees of freedom of signal (DFS) and the number of assimilated super-observations (# Obs) for the Bluelink reanalysis system in the region 90-180°E, 60°S-equator, computed for 1 January 2006. The scale for the IC is to the left and the scale for the DFS and # Obs is to the right.