Submitted to *Ocean Modelling*, 1 September 2010.

GODAE Inter-comparisons in the Tasman and Coral Seas

Peter R. Oke[1], Gary B. Brassington[2], James Cummings[3], Matthew Martin[4], Fabrice Hernandez[5]*

[1]CAWCR, CSIRO, Hobart, Australia

[2]CAWCR, BoM, Melbourne, Australia

[3]NRL, Monterey, USA

[4]Met Office, Exeter, UK

[5]Mercator Ocean, Ramonville, France

** Corresponding Author: CSIRO Marine and Atmospheric Research, GPO 1538, Hobart, TAS, 7001, Australia; phone: +61362325387; fax: +613623265123; peter.oke@csiro.au*

Keywords: GODAE, Operational Oceanography, Data Assimilation, Model validation, 140°E -180°E, 50°S-EQ

Abstract

The purpose of this paper is to compare the performance of operational short-range ocean forecast systems developed under the Global Ocean Data Assimilation Experiment (GODAE) – an international effort to demonstrate the feasibility of operational ocean forecasting. Four different forecast systems are inter-compared for the Tasman and Coral Seas, off eastern Australia. Systems considered include those developed in Australia, France, the United States, and the United Kingdom. Each system is compared to observations of along-track sea-level anomaly, sea-surface temperature, near-surface velocity, and sub-surface temperature and salinity. All systems have their strengths and weaknesses, and each system out-performs all others in one aspect or another. With few exceptions, all systems demonstrate signal to noise ratios greater than one for all variables. The products compared in this study include analyses, hindcasts, and forecasts; some computed in near-real-time, others in delayed-mode. Bearing in mind these differences, of the GODAE products compared, the Australian system generally performs the best for sea-level anomaly; the French

system is best for near-surface velocities; the US system generally performs the best for sea-surface temperature; and the UK system is best for sub-surface temperature and salinity. These findings provide useful indicators of deficiencies in each system; and provide clear metrics by which future developments should be assessed. Based on these results we promote the use of multi-model consensus forecasting, using all available forecasts from all systems, as the most robust approach for the user community. The results demonstrate the success of GODAE in demonstrating the feasibility of operational oceanography. Development of these ocean forecast systems continues under GODAE OceanView, the successor of GODAE.

1. Introduction

The primary goal of the Global Ocean Data Assimilation Experiment (GODAE; Smith, 2000; Le Traon et al., 2001) was to demonstrate the feasibility of operational ocean forecasting. Additional goals of GODAE included the development and application of state-of-the-art ocean models and data assimilation methods for short-range open-ocean forecasts; the provision of global ocean analyses and reanalyses for developing and improving understanding of the oceans; improving assessments of the predictability of ocean systems; and improving the design and effectiveness of the global ocean observing system.

Several data-assimilating ocean modelling systems have been developed under GODAE and are used for operational ocean forecasting and ocean reanalysis on either regional or global scales. These include Bluelink from Australia (Oke et al., 2005; Brassington et al., 2007); FOAM from the United Kingdom (e.g., Bell et al., 2000; Martin et al., 2007); HYCOM from the United States (e.g., Cummings, 2005; Metzger et al. 2010); Mercator from France (Drevillon et al. 2008); TOPAZ from Norway (Bertino and Liseter 2008); MOVE and COMPASS-K from Japan (Kamachi et al., 2004), C-NOOFS from Canada (www.c-noofs.gc.ca/), and the ECCO group (www.ecco-group.org). All of these systems are unique – in most cases using different model codes, different model resolution, different domains, different parameterisations, different observation sources, different surface forcing, and different data assimilation methods.

Prior to the final GODAE Symposium (www.godae.org/Final-GODAE-Symposium.html), in November 2008, an internationally coordinated effort to

undertake model inter-comparisons of the operational ocean forecast systems was undertaken. Hernandez et al. (2009) present results from these inter-comparisons, with a focus on the North Atlantic, tropical Atlantic, and North Pacific basins. They concluded that the ocean dynamics of all systems were consistent; that the wind-driven circulation was satisfactorily represented by all systems considered; all systems reasonably represented the thermohaline circulation; and although there were differences in the representation of the eddy-scale variability, further analysis is required to understand these differences. These activities did not include a detailed inter-comparison with the Australian system, because the Australian system lacks sufficient resolution in the focus regions of that study. We note that the Australian region that is chosen for this study is particularly interesting for assessing GODAE systems. It includes a western boundary current, with intense currents, fronts, and evolving (propagating) features, strong horizontal gradients, and non-linearities. All of these features pose significant challenges for models and assimilation schemes, particularly where observations only marginally resolve the features of interest. The region considered here also includes a tropical area, the Coral Sea, where vertical stratification and fast propagating dynamics are often difficult to reproduce with models. Finally, the coastal areas considered in this study also permit some possible insights into how well each system represents the slope and shelf dynamics, and its interactions with open ocean processes.

The motivation for this paper is threefold. Firstly, we seek to extend the inter-comparisons of Hernandez et al. (2009) to the Australian region, and to include comparisons with the Australian system. Secondly, we seek to quantify the performance of GODAE systems at the end of GODAE, by comparing each system to observations; and thirdly, we seek to understand when, where, and why different systems perform well, or poorly, so that all systems can be improved in terms of their modelling and data assimilation techniques. Notably, developments on GODAE systems continue under GODAE OceanView (www.godae-oceanview.org), the successor to GODAE.

The remainder of this paper includes a brief description of the models and assimilation systems along with the products compared in section 2, the results and a discussion of the inter-comparisons are presented in section 3, and the conclusions are given in section 4.

2. System Descriptions

The important elements of the models and assimilation systems used in the forecast systems are presented in Table 1. Detailed descriptions of each assimilation system are available in the published literature (e.g., Oke at al. 2008; Martin et al. 2007; Cummings 2005; Brasseur et al. 2005), including summary papers that contrast all assimilation and model systems (Dombrowsky et al. 2009; Cummings et al. 2009; Hurlburt et al. 2009) to which the interested reader is referred. The descriptions that follow are focussed on identifying the differences between the systems that may explain differences in the results presented in section 3. These descriptions only include those aspects of the systems directly relevant to the specific versions of each system that was used for the final GODAE inter-comparison exercise. However, we note that development on each system continues.

1.1 Model overview

The important elements of the model components of the four ocean forecast systems considered here are summarised in Table 1. Of these systems, Bluelink and HYCOM are eddy-resolving in the region of interest (1/10 and 1/12 degrees respectively), and FOAM and Mercator are eddy permitting (both 1/4 degrees). Bluelink, FOAM and Mercator all use z-level models, while HYCOM uses a hybrid, adaptive vertical grid (www.hycom.org). Both FOAM and Mercator use the same model code and grid (NEMO; www.nemo-ocean.eu/). Bluelink uses MOM4 (Griffies et al. 2004). All systems use different NWP flux products. The heat and freshwater fluxes for Bluelink and FOAM include diurnal variability, while Mercator and HYCOM use daily averaged heat and freshwater fluxes. Similarly, the wind forcing for Bluelink, FOAM, and HYCOM resolves diurnal variability, but again Mercator uses daily averaged winds.

1.2 Assimilation overview

The important elements of the data assimilation systems of the four ocean forecast systems considered here are included in Table 1. Each system uses a different method for data assimilation. Both Bluelink and Mercator use ensemble-based schemes to represent the background error covariance – Bluelink uses Ensemble Optimal Interpolation (EnOI; Oke et al. 2002; Evensen 2003) and Mercator uses a variant of the Singular Extended Evolutive Kalman (SEEK) filter (Testut et al. 2003). Bluelink

uses a time-invariant ensemble (BODAS; Oke et al. 2008), and Mercator uses a seasonally-varying ensemble (SAM2; Drevillon et al. 2008). Neither the Bluelink nor Mercator ensembles are state-dependent (i.e., not true ensemble Kalman filters). By contrast, both FOAM and HYCOM use an optimal interpolation approach that uses some form of analytical function to approximate the background error correlations derived from either the differences between model background fields and observations, or from model forecast fields of different lengths that are valid at the same time. The FOAM system uses the Analysis Correction (AC) approach, described by Lorenc (1991), together with the so-called NMC approach, described by Kanamitsu (1989). HYCOM uses a system called NCODA, described by Cummings (2005).

The observations assimilated by each system are quite similar (Table 1); and due to continuing GODAE efforts at data assembly centres, the observations available in near-real time (NRT) for assimilation into different operational forecasts systems are increasingly consistent. Bluelink does not assimilate all available observations – AMSRE is the only sea surface temperature (SST) product used by Bluelink, and Bluelink does not assimilate sea-level anomalies (SLA) from GFO. HYCOM and FOAM are the only systems that assimilate temperature data from surface drifting buoys. Bluelink and HYCOM are the only systems that do not use the First-Guess at Appropriate Time (FGAT) method, and Bluelink and Mercator do not use Incremental Analysis Updating (IAU; Bloom et al. 1996). Briefly, FGAT is often employed when observations from some time window, say over 5 days, are assimilated. When FGAT is used the model background field is compared to the observations at the observation time. So a time-varying background field is used. When FGAT is not used, it is common to compare the model background field at the analysis time, to observations over the whole time window. Also IAU relates to model initialisation. When IAU is used, the increment to the model state (the adjustment to all variables computed by the assimilation system) is added to the model field over multiple time steps. For example, if the increment is added over 10 time steps, one tenth of the increment is added at each time step – though the magnitude of the increment often varies over the initialisation period (Bloom et al. 2006).

1.3 Products compared

The inter-comparison period spans 1 February 2008 to 15 May 2008. The fields provided by each group are best described as a "best-estimate", plus one forecast product from Bluelink. The fields provided by the US, UK, and French groups are the class 1 fields where each group has interpolated model fields onto a common horizontal grid, for a predefined set of depths (Hernandez et al. 2009). Specifically, we compare nowcasts from the operational systems from the UK Met Office, hereafter referred to as UKMet; delayed-mode hind-casts from HYCOM (version 74.2), hereafter referred to as HYCOM; 7-14 day behind real-time hindcasts from Mercator using the PSY3V2R2 system, hereafter referred to as Mercator; and three products from the Bluelink system. The Bluelink products used here include fields from version 2p2 of the Bluelink ReANalysis (BRAN2p2; Oke and Griffin 2010), hereafter referred to as BRAN; 6-9 day behind real-time hindcasts from version 1p1 of OceanMAPS (Brassington et al. 2007), hereafter referred to as OMAPS-hc, and 3-4 day real-time forecasts from version 1p1 of OceanMAPS, hereafter referred to as OMAPS-fc. Because OceanMAPS is initialised 5-days behind real-time, the 3-4 day forecasts reported here are the model forecasts 8-9 days after initialisation. BRAN and the OMAPS products are based on the Bluelink system, however they are configured slightly differently. For example, BRAN assimilates delayed-mode quality controlled altimetry and Argo T/S profiles but no XBT data, uses a 7-day update cycle, using 120 ensemble members, and is forced with NRT surface fluxes from ECMWF. By contrast, the OMAPS products assimilate NRT altimeter data and Argo TS profiles plus XBT data from the Global Telecommunications System (GTS), uses a 3-4 day update cycle, using 72 ensemble members, and is forced with NRT surface fluxes from the Bureau of Meteorology's operational NWP system (GASP).

In section 3, we compare results from each GODAE system to observations and observation-based products. Specifically, we compare to along-track Sea Level Anomalies (atSLA) from Jason-1, Envisat and GFO satellite altimeters; L2P AMSRE SST (www.remss.com/), surface velocities from satellite-tracked surface drifting buoys (www.aoml.noaa.gov/phod/dac), and potential temperature (T) and salinity (S) observations from the Argo program (www.argo.net), using delayed-mode Argo data where it is available (accessed from Coriolis and US GODAE in June 2009; duplicates are removed). We also compare results from GODAE systems to observation-based analysis products. These include gridded SLA maps, produced

using optimal interpolation (OI) at CSIRO (www.cmar.csiro.au/remotesensing/oceancurrents/), and a gridded SST product, also produced using OI under the Group for High Resolution SST (GHRSST) project (www.ghrsst.org) – specifically the Regional Australian Multi-Sensor SST Analysis (RAMSSA; Beggs et al. 2010).

3. Results and Discussion

3.1 Regions

Inter-comparisons of each system are presented here for two regions in the south-west Pacific Ocean – the East Australian Current (EAC) region, and the Coral Sea and Papua New Guinea (PNG) region. These regions are denoted in Figure 1. The EAC region is characterised by the strong southward currents of the EAC that flow adjacent to the east Australian coast between about 20^{o}S and 33^{o}S. South of about 33^{o}S the EAC typically separates from the coast and degenerates into a complex field of mesoscale eddies (e.g., Oke and Griffin 2010; and references therein). The most salient feature of the Coral Sea is a quasi-stationary gyre, centered around 148^{o}E and 12^{o}S, called the North Queensland or Hiri Current (e.g., Steinberg 2007). The New Guinea Coastal Current flows west-north-west along the northern coast of PNG, feeding the West Pacific warm pool (e.g., Cresswell 2000). Both the EAC and the Hiri Current are fed from the westward flow of the South Equatorial Current.

3.2 Taylor diagrams

In this section, modelled and observed estimates of oceanic properties are compared using Taylor diagrams (Taylor 2001). Taylor diagrams succinctly represent the unbiased root-mean-squared difference (RMSD; i.e., the RMSD after the mean has been removed) and cross-correlation between observed and modelled estimates of some quantity. Taylor diagrams also show the standard deviation of observed and model estimates. With reference to Figure 2a, showing comparisons of atSLA in the EAC region, we will explain how Taylor diagrams can be interpreted. Taylor diagrams exploit a geometric relationship between RMSD, cross-correlation and standard deviation through the use of two over-lapping polar coordinates. Note that the mean of each estimate is removed prior to generation of the Taylor diagram. So the Taylor diagram does not provide any explicit comparison of the mean, or bias, of each estimate. The radial distance of each dot, for either a set of model estimates or observations, from the origin represents the standard deviation. Figure 2a indicates

that the standard deviation of the observed atSLA is slightly greater than 0.15 m. The estimate of atSLA from BRAN, for example, is slightly less that 0.15 m. The arc distance from the vertical axis corresponds to the cross-correlation between each model estimate and the observed estimate. The vertical axis corresponds to zero correlation; the horizontal axis corresponds to a correlation of 1; with an inverse cosine scaling in between. The observed dot is always along the horizontal axis because it is perfectly correlated with itself. For Figure 2a, the cross-correlation between the observed atSLA and BRAN is approximately 0.8. The distance between the dot representing the observed estimate and each model estimate is the RMSD. So Figure 2a indicates that the RMSD between the observed atSLA and BRAN is slightly less than 0.1 m.

Taylor diagrams can be used to get a sense for the signal to noise ratio of each estimate. If the RMSD exceeds the standard deviation, the signal to noise is less than 1. In such a case, the model field is not quantitatively skilful. If the standard deviation of a model is greater than the standard deviation of the observations, then we might conclude that the model is noisier than the observations. This might be due to unconstrained instabilities in the model, or because there is more energy in the model than in the observations – this might occur if a high resolution model is compared to a coarse resolution observation, like a Real-Time Global (RTG) SST analysis product for example.

3.3 EAC Region

Modelled and observed estimates of atSLA, SST, and near-surface velocities in the EAC region are compared in the Taylor diagrams presented in Figure 2. Similarly, modelled and observed estimates of sub-surface T and S in the EAC region are compared in Figure 3, showing profiles of the RMSD between observed and modelled estimates. Also included in Figure 3, are profiles of the RMSD between observed fields and climatology (we use climatology from the CSIRO Atlas for Regional Seas; CARS; Ridgway and Dunn, 2003).

Using the RMSD as the most basic metric for evaluation, Figure 2 and 3 indicate that the model estimates closest to observations of atSLA, SST, near-surface velocity, and sub-surface T/S in the EAC come from BRAN, HYCOM, Mercator, and UKMet respectively. Both Mercator and UKMet tend to under-represent the variability of

near-surface velocity, with standard deviations that are measurably less than the observations and the HYCOM and Bluelink models. This is expected, since both Mercator and UKMet models are $1/4^{o}$, that is appreciable coarser than $1/10^{o}$ and $1/12^{o}$ resolution used by Bluelink and HYCOM in this region. It is also interesting to note that UKMet tends to produce better T/S profiles compared to Mercator, while Mercator tends to produce better SLA. Noting that these systems use the same model code and grid, it is logical to conclude that these differences are attributable to the assimilation schemes that have presumably been "tuned" differently for different variables – though the differences may also be related to different surface forcing.

Considering only the series of Bluelink results, the estimates from the operational forecast (OMAPS-fc) are always less skilful than the hindcast products (BRAN and OMAPS-hc), as we might reasonably expect. However, the delayed-mode reanalysis (BRAN) is *not* always more skilful than the operational hindcast product (OMAPS-hc). This is somewhat surprising, because the operational hindcast product uses NRT observations. The main difference between BRAN and OMAPS-hc that might explain this is the update cycle. BRAN assimilates observations once every 7 days, while OMAPS-hc assimilates observations once every 3 or 4 days (with twice weekly updates). We expect that the more frequent update cycle in OMAPS-hc translates to an improvement in the system's performance for some variables, since instabilities in the EAC region are known to develop and evolve rapidly over very short periods like 2-3 days. BRAN and OMAPS-hc also use different surface fluxes from operational NWP systems (see Table 1), and the assimilation system is implemented slightly differently – BRAN uses shorter localising length-scales (5 degrees) than OMAPS-hc (8 degrees). These differences may also contribute to the different performance of the systems, but the details are the subject of ongoing consideration.

Considering the sub-surface profiles of the RMSD for T and S (Figure 3), the results show that, with the exception of the forecast product, all estimates have smaller errors than climatology. This is encouraging, particularly in this region that is characterised by transient, mesoscale features. In such a region, the displacement of an eddy can result in very large RMSD statistics, as discussed by Oke et al. (2008).

Close inspection of Figure 3 near the surface seems to give a result that is somewhat inconsistent with the SST comparisons, presented in Figure 2a. Specifically, the RMSD of the near-surface T is smallest for UKMet in Figure 3a, but the RMSD of

SST is smallest for HYCOM in Figure 2b. We expect this is due to different spatial and temporal sampling of Argo (approximately 200 profiles in total for the whole region over the entire time period) and satellite SST (approximately 25 km resolution maps spanning most of the region every day). But this difference highlights one of the limitations of this study, namely that it is short – spanning less than 4 months – and is focussed on only a single region of the ocean.

3.4 Coral Sea and PNG region

Modelled and observed estimates of atSLA, SST, and near-surface velocities in the Coral Sea and PNG region are compared in the Taylor diagrams presented in Figure 4. Similarly, modelled and observed estimates of sub-surface T and S in the Coral Sea and PNG region are compared in Figure 5, showing profiles of the RMSD between observed and modelled estimates.

The RMSD statistics shown in Figure 4 and 5 again indicate that the model estimates closest to observations of atSLA, SST, near-surface velocity, and sub-surface T/S in the EAC come from BRAN, OMAPS-hc, Mercator, and UKMet respectively. For the meridional velocities, both Mercator and UKMet again tend to under-represent the variability (Figure 4d), with smaller standard deviations than the other estimates.

In this region, the RMSD between observed and modelled sub-surface T for all systems are significantly less than climatology for depths shallower that 400 m – the depths that are of particular interest for GODAE systems. However, the RMSD for S is not very much less than climatology. This suggests that S is not particularly well constrained by any of the models in this region – perhaps because there are too few observations. It could also be an indication that the assimilation systems are generally not well tuned for S.

3.5 XBT transect

The frequently repeated XBT transect PX-6, between Auckland and Fiji (see Figure 1), was occupied during the inter-comparison period. The observed T section from this line is shown in Figure 6a. Comparisons of T anomalies (i.e., anomalies from CARS) at several depths are also shown in Figure 6b-e. Note that the metrics retrieved from the various modelling systems for this inter-comparison exercise only included an agreed subset of depths (Crosnier et al. 2006), so a detailed comparison of this transect is not feasible using the class 1 metrics that have been used in this study.

These comparisons demonstrate that the GODAE models generally provide a realistic representation of the sub-surface T field. Anomalies along this section range from -1.2^{o}C to 2.1^{o}C; and the GODAE systems generally represent the sign and locations of the anomalies, and the approximate magnitude of the anomalies along the transect. Consistent with the comparisons in Figures 3 and 5, the results from UKMet, along with OMAPS-hc, appear to out-perform the other systems – though the differences are generally small relative to the magnitude of the anomalies.

3.6 SLA sequence

A sequence of SLA maps are shown in Figure 7, comparing an analysis product, based on optimal interpolation (labelled OI-SLA), and SLA fields from HYCOM, UKMet, Mercator, and BRAN. The purpose of this comparison is simply to demonstrate that all systems produce realistic estimates of the time-varying state of the ocean. During the period displayed in Figure 7, the salient features include two warm-core eddies and three cold-core eddies, all in the vicinity of 31-37^{o}S. All systems include some representation of each of these features – though their evolution over time differs. Specifically, HYCOM shows the two warm-core features being drawn together, with a hint that they will coalesce. The variability in the UKMet fields is notably less than the OI-SLA and the models. Mercator, BRAN, and OMAPS-hc all show the south-eastern-most cold-core eddy splitting into two by 26 February – a feature that is also evident in the OI-SLA, but not in HYCOM or UKmet. The differences between the models occur in the positions and intensity of mesoscale features. Notably, the performance of OMAPS-fc, the only true operational forecast product considered in this study, shows good agreement with the other products, including the OI-SLA.

3.7 SST sequence

A sequence of SST maps (Figure 8) compares a GHRSST analysis product (RAMSSA; Beggs et al. 2010), with daily mean SST fields from HYCOM, UKMet, Mercator, and BRAN. In general, all of the GODAE systems reproduce SST fields that compare well with the GHRSST analysis. The SST fields from UKMet and Mercator are slightly smoother than GHRSST, HYCOM, and the Bluelink products – though one could argue that they are less noisy and therefore possibly more reliable. There are times when several systems appear to have somewhat noisy SST fields. For

example, note the small-scale filaments around 160°E and 35°S on 9 April in UKMet, Mercator, and BRAN. Small-scale features are also evident in GHRSST and HYCOM – though the correspondence between those two products is very good. More generally, all model products appear to show somewhat noisy SST fields around 35°S. At different times, some GODAE systems include features that appear fictitious – possibly due to assimilation of bad data, or due to dynamical instabilities in the models that are known to occur in this region (e.g., O'Kane et al. 2010). Again, we note that Bluelink and HYCOM offer higher resolution than UKMet and Mercator. This might explain some of the differences. Moreover, if the GHRSST product is too smooth, some SST features that appear in the high resolution models might be absent in the GHRSST product.

3.8 Variability

The standard deviation of SLA from OI-SLA and from each GODAE system is shown for the EAC region in Figure 9. The observation-based product (OI-SLA) shows 6 local maxima in SLA variability, with a magnitude that exceeds 0.2 m. OI-SLA shows one local maxima around 37°S, one around 36°S (about 159°E), two around 34°S, and two around 32°S. All GODAE systems reproduce the local maxima around 37°S, all reproduce two local maxima around 34°S, and all reproduce two local maxima around 32°S. This result is very encouraging, because it demonstrates that the variability reproduced by each system is qualitatively realistic. Only HYCOM and BRAN reproduce the local maxima at 36°S; and only Mercator and BRAN realistically reproduce the magnitude of the largest maximum (at 34°S, 155°E). We also note that the reproduction of many of these features by OMAPS-fc is good. This demonstrates that the GODAE systems can realistically forecast (both persist and dynamically evolve) this variability.

The standard deviation of SST from the GHRSST product and from each GODAE system is shown for the south-west Pacific in Figure 10. In general the GHRSST product has lower variability than all of the GODAE systems. The most prominent feature in the observation-based product (GHRSST) is the diagonally oriented maximum centered on about 155°E and 35°S. All of the GODAE systems reproduce a similar maximum. Another prominent feature is the band of high variability around 22°S that extends along the north-east coastal of Australia, to about 15°S.

All GODAE systems reproduce relatively high variability in these regions, though the models tend to have greater variability there, particularly near the coast, over the Great Barrier Reef (GBR). It's not entirely clear whether the model variability is too high, or the GHRSST variability is too low. The GHRSST product may under-predict the true SST variability because it uses the previous analysis as the background field for the next analysis (see Beggs et al. 2010) – so in the absence of new observations, the previous analysis persists. BRAN also significantly over-estimates the SST variability in Bass Strait, around 40^{o}S. The over-estimate of SST variability in these coastal regions (Bass Strait and GBR) is an indication that the models are poorly constrained there and possibly lacking some important physical processes (e.g., tidal mixing – none of the systems considered include tides). Also, the models may be overly sensitive to the quality of the atmospheric fluxes. BRAN, for example, only assimilates AMSR-E SST that is not available within about 50 km of the coast, and has no data in Bass Strait (at least none that are assimilated into BRAN). The modelled SST is often quite variable near the coast due to strong local wind forcing, and in the absence of sufficient observations, the modelled variability can become somewhat unreliable. In those locations, the accuracy of the model is closely related to the accuracy of the local wind forcing. This was shown to be true for BRAN by Oke et al. (2008). In general, the SST standard deviation in the models tend to be on small scales and quite noisy in the southern part of the domain shown in Figure 10. This is an indication of the dynamic nature of the variability there, and may imply that predictability in that region is limited by dynamic instabilities, as suggested by O'Kane et al. (2010). Another region of interest, where there is notable disagreement with observations is on the coastal side of the Hiri Current (around 10^{o}S). All models show relatively high SST variability there (Figure 10), but the observations do not. The reason for this is unclear, but again, it may be an indication that the models are insufficiently constrained there due to a paucity of observations.

4. Conclusions

The purpose of this paper is to document the performance of GODAE systems at the end of the GODAE project and to extend the inter-comparisons of Hernandez et al. (2009) to the Australian region, and to include inter-comparisons with the Australian system. Comparisons between observations and GODAE systems demonstrate that each system has certain strengths and weaknesses. We recognise that several aspects

of the inter-comparisons presented in this study are not ideal. For example, the products we compare include analyses, hindcasts, and forecasts; some computed in NRT, others in delayed-mode. We also note that this study is limited to two regions of the ocean, and results are likely to be different for other regions. The period over which the inter-comparisons are performed is short – less than 4 months. This means that the quantitative comparisons presented here may be statistically indistinguishable. Despite these issues, we regard inter-comparisons as valuable exercises that provide important insights into the ocean forecast systems for both developers and users. We also note that this inter-comparison study, and that of Hernandez et al. (2009), is one of the first attempts at performing GODAE inter-comparisons. Now that the tools have been developed, the metrics have been defined, and the community has engaged in inter-comparisons, the GODAE OceanView community is well placed to undertake more comprehensive inter-comparisons, using more complete datasets, and more compatible datasets, including class 2, 3 and 4 metrics (Hernandez et al. 2009). We expect that future inter-comparisons are likely to focus more and more on operational forecast products. One of the key challenges for those inter-comparisons and related studies will be on the characterisation of ocean predictability, and on the capacity for GODAE systems to deliver forecasts that are skilful enough to be of real benefit to the user community.

Bearing in mind the caveats and limitations outlined above, we conclude that off the east coast of Australia, the Bluelink system produces the most accurate representation of SLA, the HYCOM system produces the most accurate representation of SST, the Mercator system produces the most accurate representation of near-surface velocity, and the UKMet system produces the most accurate representation of sub-surface T and S.

Different variables are important for different applications. So, based on this study a user might benefit from using analyses from the system with strengths in the variable most relevant to their application. For example, during a search and rescue operation, a user might adopt Mercator; to identify the position of an eddy, say for a fisheries application, a user might adopt Bluelink (BRAN); to monitor the chance of a warming event that might impact corals or aquaculture, a user might adopt HYCOM or Bluelink (OMAPS); and to predict underwater acoustics that depend strongly on sub-surface T, a user might adopt UKMet. However, noting that the skill of an individual

analysis or forecast varies with time, we propose that a more robust approach is to combine information from all systems together, in the form of a so-called super ensemble, or consensus forecast. Given their different strengths and weaknesses, such an ensemble could prove most useful for communities that depend on operational oceanography. We also note that user-oriented inter-comparison studies (e.g., Hackett et al. 2009; Davidson et al. 2009) should also be performed to comprehensively assess the suitability of GODAE systems for different applications.

We note that development of the GODAE systems used here continues under GODAE OceanView, the successor to GODAE, and the performance of each system continues to improve (e.g., Oke and Grifin 2010; Storkey et al. 2010; Ferry et al. 2010; see also the GODAE OceanView national reports at www.godae.org/documents and www.godae-oceanview.org/documents). Based on the results of this study however, it is clear that operational oceanography is indeed feasible, and is now in full swing at several agencies. The transition of these systems into operational services is being aided through the WMO/IOC Joint Technical Commission for Oceanography and Marine Meteorology – JCOMM (www.jcomm.info). As the skill of operational systems improve through adoption of new observing systems, improvements to assimilation methods, and more accurate models, the impact of GODAE systems on industry groups and marine users is likely to continue to increase; and as cooperation between international partners continues under GODAE OceanView, the potential to produce accurate, reliable ocean forecasts are being realised.

Acknowledgements

The following organisations and agencies are acknowledged for their financial support: European Space Agency; French Service Hydrographique Océanographique de la Marine (SHOM); Office of Naval Research; and the Royal Australian Navy. Satellite altimetry is provided by NASA, NOAA, ESA and CNES. Drifter data are provided by NOAA-AOML and SST observations are provided by NASA, NOAA and Remote Sensing Systems. Argo data were collected and made freely available by the International Argo Project and the national programmes that contribute to it (www.argo.net). Inter-comparison studies were financed at Mercator Ocean through the funding of the European Union MERSEA Integrated Project (Contract no. SIP3-CT-2003-502885). We thank M. Drévillon and C. Regnier at Mercator Ocean for their

contribution in preparing Mercator simulations and GODAE metrics.

References

Beggs, H., A. Zhong, G. Warren, O. Alves, G. B. Brassington, T. F. Pugh (2010) RAMSSA - An Operational, High-Resolution, Multi-Sensor Sea Surface Temperature Analysis over the Australian Region, submitted.

Bell, M. J., R. M. Forbes, A. Hines, 2000. Assessment of the FOAM global data assimilation system for real-time operational ocean forecasting. *J. Mar. Syst.,* **25**, 1–22.

Bertino, L., K. A. Liseter (2008) The TOPAZ monitoring and prediction system for the Atlantic and Arctic Oceans, *Journal of Operational Oceanography,* **2**, 15-19.

Bloom, S. C., L. L. Takas, A. M. Da Silva, D. Ledvina (1996) Data assimilation using incremental analysis updates. *Monthly Weather Review,* **124**, 1256–1271.

Brasseur P., Bahurel P., Bertino L., Birol F., Brankart J.-M., Ferry N., Losa S., Rémy E., Schröter J., Skachko S., Testut C.-E., Tranchant B., van Leeuwen P.J., Verron J., 2005 : Data Assimilation for marine monitoring and prediction : The Mercator operational assimilation systems and the MERSEA developments, *Q. J. R. Met. Soc*., **131**, 3561-3582.

Brassington, G. B., T. F. Pugh, C. Spillman, E. Schulz, H. Beggs, A. Schiller, P. R. Oke (2007) BLUElink> development of operational oceanography and servicing in Australia. *J. Res. Pract. Inf. Tech.,* **39**, 151–164.

Cresswell, G. R. (2000) Coastal currents of northern Papua New Guinea, and the Sepik River outflow, *Mar. Freshwater Res.*, **51**, 553–64, 10.1071/MF99135 1323-1650/00/060553.

Crosnier, L., C. Le Provost, and MERSEA Strand Team. 2006. Internal metrics definition for operational forecast systems inter-comparison: Examples in the North Atlantic and Mediterranean Sea. pp. 455–465 in *GODAE Summer School in "Ocean Weather Forecasting: An Integrated View of Oceanography."* E.P. Chassignet and J. Verron, eds, Springer, Lallonde les Maures, France.

Cummings, J.A. 2005. Operational multivariate ocean data assimilation. *Quart. J. Royal Met. Soc.*, **131**, 3583-3604.

Cummings, J. A., L. Bertino, P. Brasseur, I. Fukumori, M. Kamachi, M. Martin, K. Mogensen, P. R. Oke, C. Emmanuel Testut, J. Verron, A. Weaver, 2009: Ocean data assimilation systems for GODAE. *Oceanography,* **20**, 102-115.

Davidson, F. J. M., A. Allen, G. B. Brassington, O. Brevik, P. Daniel, M. Kamachi, S. Sato, B. King, F. Lefevre, M. Sutton, H. Kaneko (2009) Applications of GODAE Ocean Current Forecasts to Search and Rescue and Ship Routing. *Oceanography,* **22**, 176-181.

Dombrowsky, E., L. Bertino, G.B. Brassington, E.P. Chassignet, F. Davidson, H.E. Hurlburt, M. Kamachi, T. Lee, M.J. Martin, S. Mei, M. Tonani (2009) GODAE systems in operation. *Oceanography*, **20**, 80-95.

Drevillon, M.; Bourdallé-Badie, R.; Derval, C.; Drillet, Y.; Lellouche, J-M.; Rémy, E.; Tranchant, B.; Benkiran, M.; Greiner, E.; Guinehut, S.; Verbrugge, N.; Garric, G.; Testut, C-E.; Laborie, M.; Nouel, L.; Bahurel, P.; Bricaud, C.; Crosnier, L.; Dombrowsky, E.; Durand, E.; Ferry, N.; Hernandez, F.; Galloudec, O Le; Messal, F.; Parent, L. (2008) The GODAE/Mercator-Ocean global ocean forecasting system: results, applications and prospects, *Journal of Operational Oceanography*, 1, 51-57.

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

Ferry, M., L. Parent, G. Garric, B. Barnier, N. C. Jourdin, Mercator Ocean team (2010) Mercator Global Eddy Permitting Ocean Reanalysis GLORYS1V1: Description and Results. *Mercator Ocean Quarterly Newsletter,* **36**, 15-27.

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

Griffies, S. M., R. C. Pacanowski, A. Rosati (2004) A technical guide to MOM4. GFDL Ocean Group Technical Report No. 5 NOAA/Geophysical Fluid Dynamics Laboratory, 371pp.

Hackett, B., E. Comerma, P. Daniel, H. Ichikawa (2009) Marine Oil Pollution Prediction. *Oceanography,* **22**, 168-175.

Hernandez, F., L. Bertino, G. Brassington, E. Chassignet, J. Cummings, F. Davidson, M. Drévillon, G. Garric, M. Kamachi, J.-M. Lellouche, R. Mahdon, M.J. Martin, A. Ratsimandresy, C. Regnier (2009) Inter-comparison studies within GODAE. *Oceanography,* **20**, 128-143.

Hurlburt, H. E., G. B. Brassington, Y. Drillet, M. Kamachi, M. Benkiran, R. Bourdalle-Badie, E. P. Chassignet, G. A. Jacobs, O. Le Galloudec, J.-M. Lellouche, E. J. Metzgar, P. R. Oke, T. F. Pugh, A. Schiller, O. M. Smedstad, B. Tranchant, H. Tsujino, N. Usui, A. J. Wallcraft (2009) High-resolution global and basin-scale ocean analyses and forecasts. *Oceanography,* **22**, 80-97.

Kamachi, M., T. Kurangano, H. Ichikawa, H. Nakamura, A. Nishina, A. Isobe, D. Ambe, M. Arai, N. Gohda (2004) Operational data assimilation system for the Kuroshio South of Japan: reanalysis and validation. *J. Oceanogr.,* **60**, 303–312.

Kanamitsu, M., (1989) Description of the NMC global data assimilation and forecast system. *We. Forecasting,* **4**, 335-342.

Le Traon, P.Y., Reinecker, M., Smith, N.R., Bahurel, P., Bell, M., Hurlburt, H., Dandin, P., 2001. Operational oceanography and prediction: a GODAE perspective. In: Koblinsky, C.J., Smith, N.R. (Eds.), Observing the Oceans in the 21st Century, GODAE Project Office. Bureau of Meteorology, Melbourne, pp. 529–545.

Lorenc A. C., R. S. Bell, B. MacPherson (1991) The Met Office analysis correction data assimilation scheme. *Q. J. R. Meteorol. Soc.,* **117**: 59–89.

Martin, M.J., A. Hines, M. J. Bell (2007) Data assimilation in the FOAM operational short-range ocean forecasting system: a description of the scheme and its impact. *Quarterly Journal of the Royal Meteorological Society,* **133,** 981–995.

Metzger, E.J., O.M. Smedstad, P. Thoppil, H.E. Hurlburt, D.S. Franklin, G. Peggion, J.F. Shriver, T.L. Townsend and A.J. Wallcraft, 2010: Validation Test Report for the Global Ocean Forecast System V3.0 - 1/12º HYCOM/NCODA: Phase II. NRL Memo. Report, NRL/MR/7320--10-9236.

O'Kane, T. J., P. R. Oke, P. A. Sandery, 2010: Predicting the East Australian Current: Errors of the day, Ocean Modelling, submitted.

Oke, P. R., J. S. Allen, R. N. Miller, G. D. Egbert, P. M. Kosro (2002) Assimilation of surface velocity data into a primitive equation coastal ocean model. *Journal of Geophysical Research,* **107**(C9), 3122-3147.

Oke, P. R., G. B. Brassington, D. A. Griffin A. Schiller (2008) The Bluelink Ocean Data Assimilation System (BODAS), Ocean Modelling, 21: 46-70.

Oke, P. R., D. A. Griffin (2010) The cold-core eddy and strong upwelling off the coast of New South Wales in early 2007. Deep Sea Research, doi:10.1016/j.dsr2.2010.06.006.

Ridgway, K. R., J. R. Dunn (2003) Mesoscale structure of the mean East Australian Current System and its relationship with topography. *Progress in Oceanography*, 56, 189-222.

Rio, M.-H., F. Hernandez, 2004: A mean dynamic topography computed over the world ocean from altimetry, in situ measurements, and a geoid model, *J. Geophys. Res.,* **109**, C12032.

Smith, N.R., (2000) The global ocean data assimilation experiment. *Adv. Space Res.,* **25**, 1089–1098.

Steinberg, C. (2007) Impacts of climate change on the physical oceanography of the Great Barrier Reef, in Climate Change and the Great Barrier Reef, edited by J. E. Johnson and P. A. Marshall, pp. 51– 74, Great Barrier Reef Mar. Park Auth., Townsville, Queensl., Australia.

Storkey, D., E.W. Blockley, R. Furner, C. Giuavarc'h, D. Lea, M.J. Martin, R.M. Barciela, A. Hines, P. Hyder, J.R. Siddorn (2010). Forecasting the ocean state using NEMO: The new FOAM system. *J. Operational Oceanography*, **3**, 3-15.

Taylor, K. E., (2001) Summarizing multiple aspects of model performance in a single diagram. *J. Geophys*. Res., 106, 7183-7192.

Testut, C-E., P. Brasseur, J-M Brankart, J. Verron (2003) Assimilation of sea-surface temperature and altimetric observations during 1992–1993 into an eddy

permitting primitive equation model of the North Atlantic Ocean. *J. Mar. Sys.,* **40-41**, 291-316.

Table 1: Summary of the model and assimilation systems for each GODAE system. All of the acronyms listed below are either defined in the text of this paper, or follow here: RTG refers to Real-Time Global SST product (polar.ncep.noaa.gov/sst/); TAO* includes observations from TAO, PIRATA and RAMA;; GC99 refers to Gaspari and Cohn (1999); Rio-v5 refer to the mean sea-level described by Rio and Hernandez (2004).

	BRAN	**OMAPS**	**UKMet**	**HYCOM**	**Mercator**
Model code	MOM4	MOM4	NEMO 1.09	HYCOM	NEMO 1.09
Horizontal grid	1/10°	1/10°	1/4°	1/12°	1/4°
Vertical grid	47 levels	47 levels	50 levels	32 hybrid	50 levels
Heat and freshwater fluxes	ECMWF 6-hr	GASP 3-hr	UKMO 6-hr	NOGAPS 1-d	ECMWF 1-d
Wind forcing	ECMWF 6-hr	GASP 3-hr	UKMO 6-hr	NOGAPS 3-hr	ECMWF 1-d
Assimilation System	BODAS	BODAS	AC/OI	NCODA	SAM2
Background error covariance	Time invariant Ensemble	Time invariant Ensemble	NMC	Flow-dependent Gaussian	Seasonally-varying Ensemble
Scheme	EnOI (120-members)	EnOI (72-members)	FOAM	MvOI	SEEK (200-220 modes)
Localising length-scale	5° GC99	8° GC99	n/a	n/a	200-500 km
SST data	AMSRE	AMSRE	AVHRR + AMSRE + In Situ	AVHRR + AMSRE + In Situ	RTG
SSH data	Jason-1 + Envisat	Jason-1 + Envisat	Jason-1 + Envisat + GFO	Jason-1 + Envisat + GFO	Jason-1 + Envisat + GFO
In Situ	Argo + TAO*	Argo + XBT + TAO*	Argo + XBT + buoys + TAO*	Argo + XBT + buoys + TAO*	Argo + XBT + TAO*
FGAT	No	No	Yes	No	Yes
MSL	15-y model	15-y model	Rio-v5	20-yr model	Rio-v5
Assimilation cycle	7-d	3-4-d	1-d	1-d	7-d
Obs window (SST/SLA/TS)	3/11/7-d	3/11/7-d	5/5/10-d	1/3/12-d	7/7/7-d
Initialisation	Nudging (1-d)	Nudging (1-d)	IAU (1-d)	IAU (6-hr)	Instantaneous update
Version	2p2	1p1	1p0	Exp 74.2	PSY3V2R2

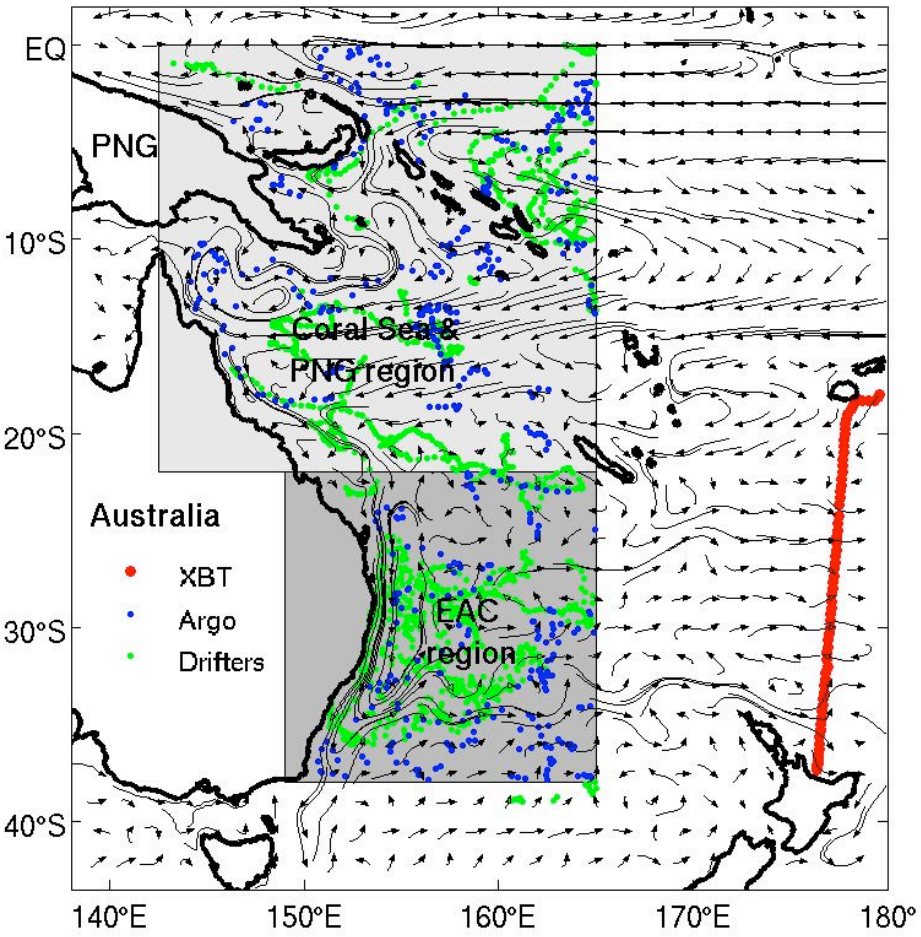

Figure 1: Map of the SW Pacific Ocean showing an estimate of the mean ocean circulation over the top 200 m depth, two geographic regions referred to in this study, the distribution of Argo T/S profiles, and the XBT section.

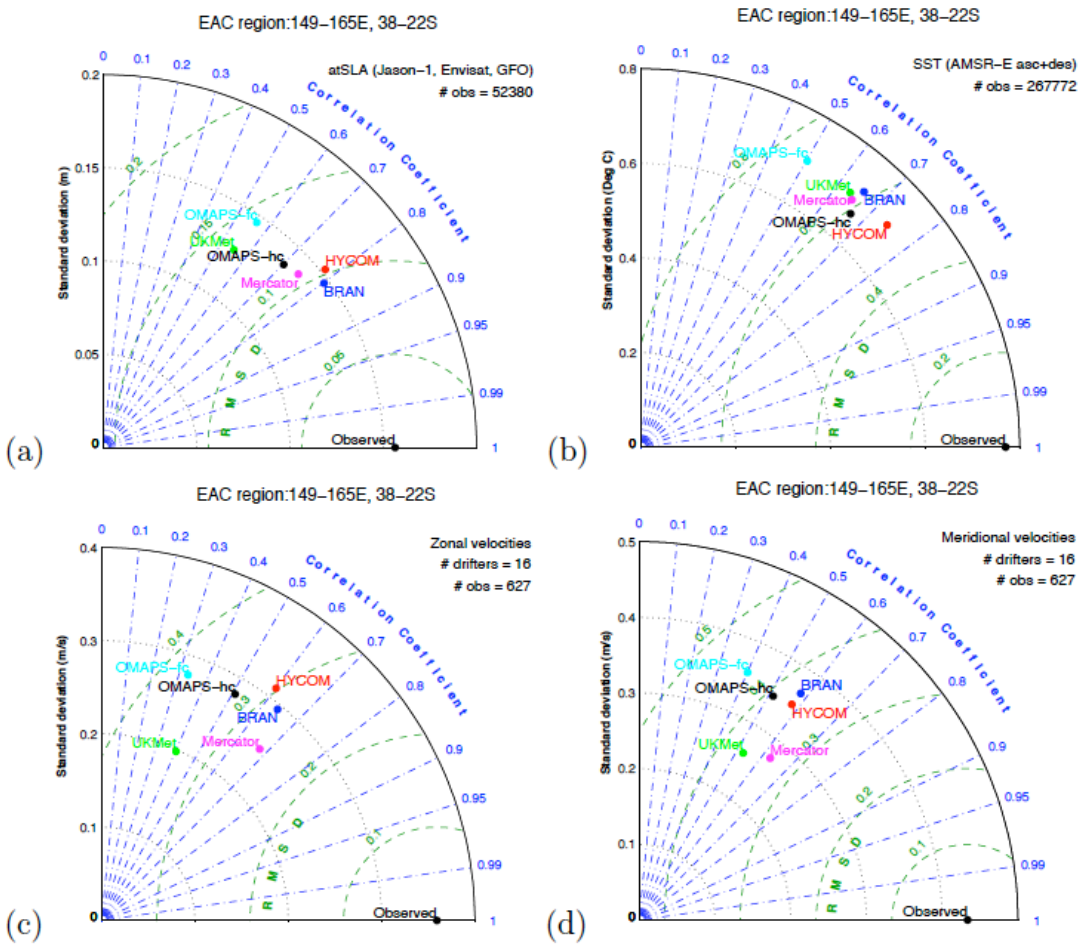

Figure 2: Taylor diagrams for (a) atSLA, (b) AMSRE SST (c) drifter-derived zonal velocities, and (d) drifter-derived meridional velocities in the EAC region. The number of observations used for each panel is listed explicitly (# obs and # drifters).

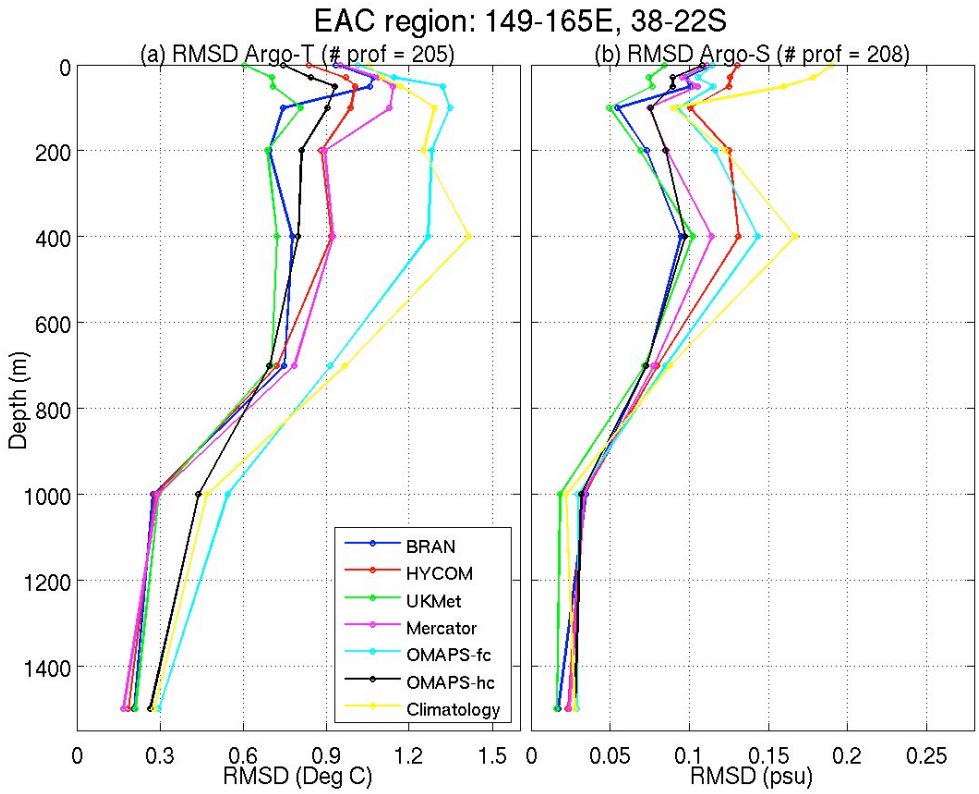

Figure 3: Profiles of the RMSD for T (left) and S(right) for each GODAE system, and for climatology, for the EAC region. The number of Argo profiles used for each panel is listed explicitly (# prof).

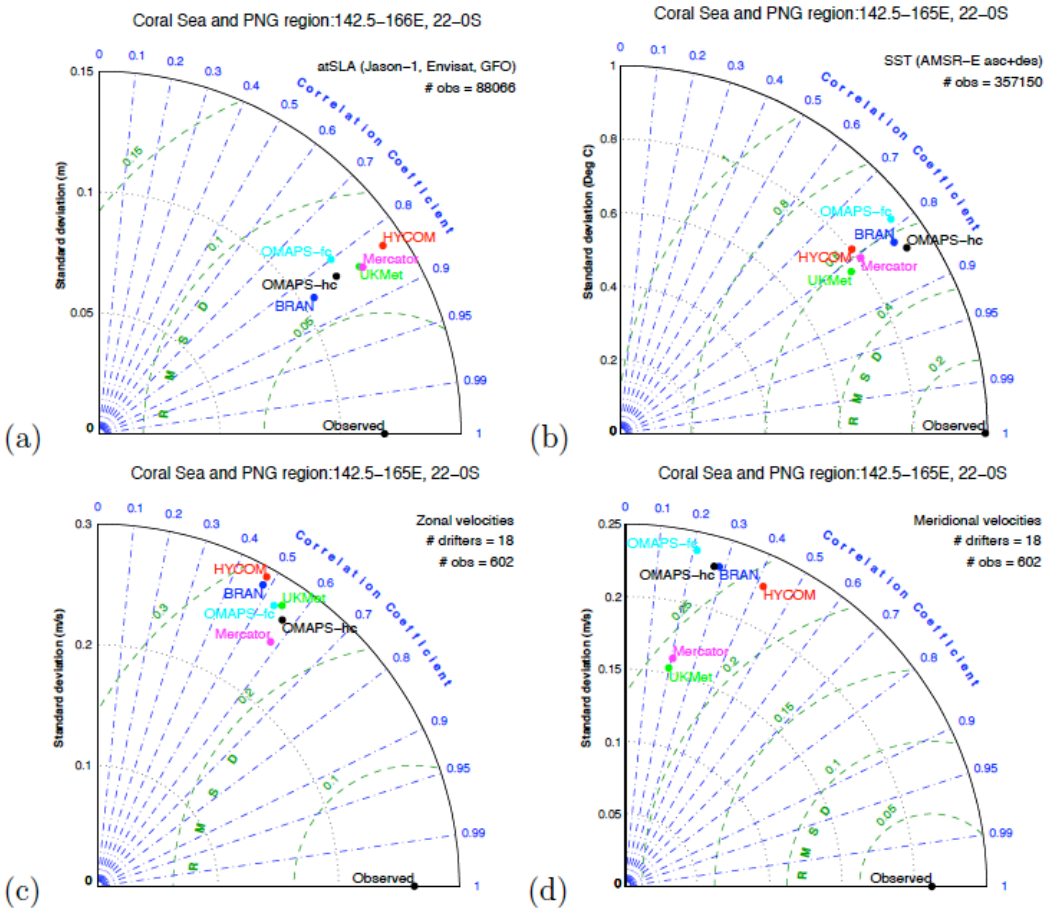

Figure 4: As for Figure 2, except for the Coral Sea and PNG region.

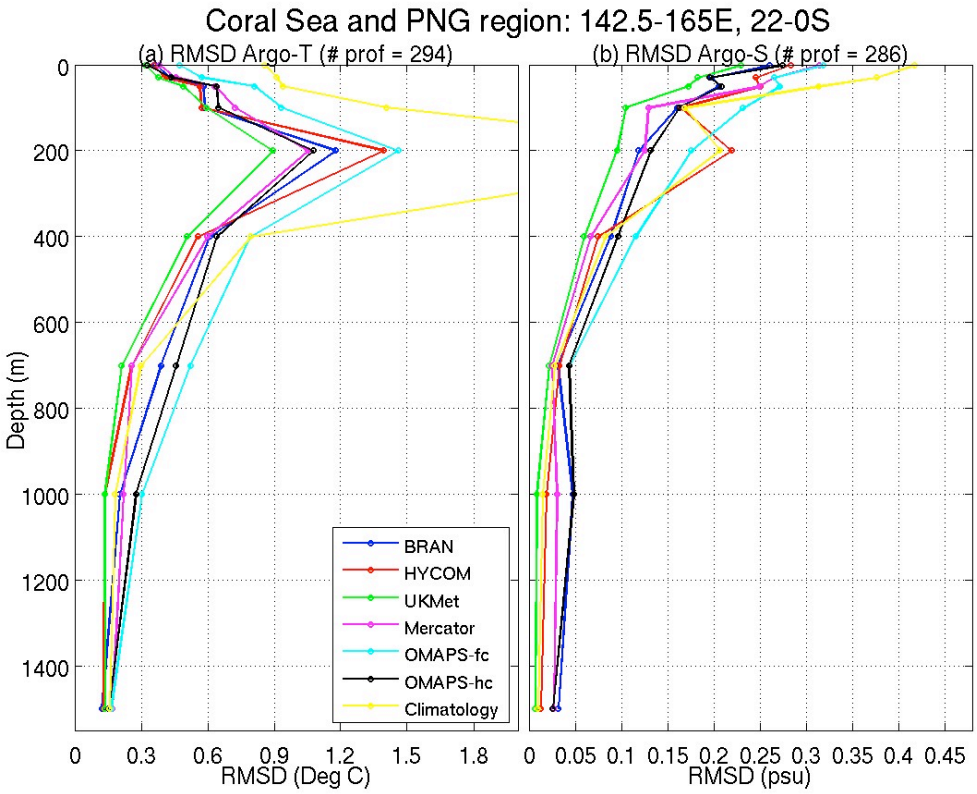

Figure 5: As for Figure 3, except for the Coral Sea and PNG region.

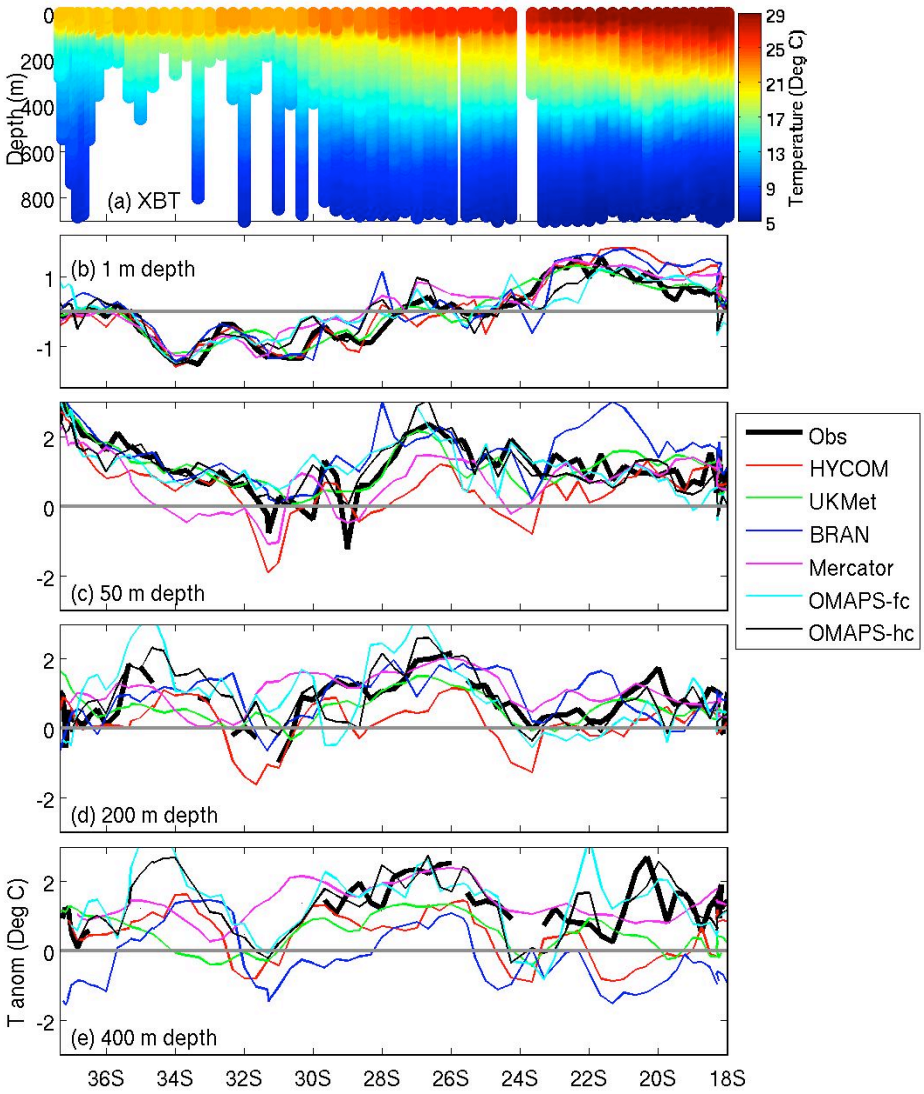

Figure 6: Comparison of T along the PX-6 XBT transect, denoted in Figure 1, that was occupied between 21-25 February 2008, showing (a) the latitude-depth section from observations, and (b-e) T anomalies (relative to climatology) from the XBT observations (bold black) and from each GODAE system (see legend) for 1, 50, 200, and 800 m depth.

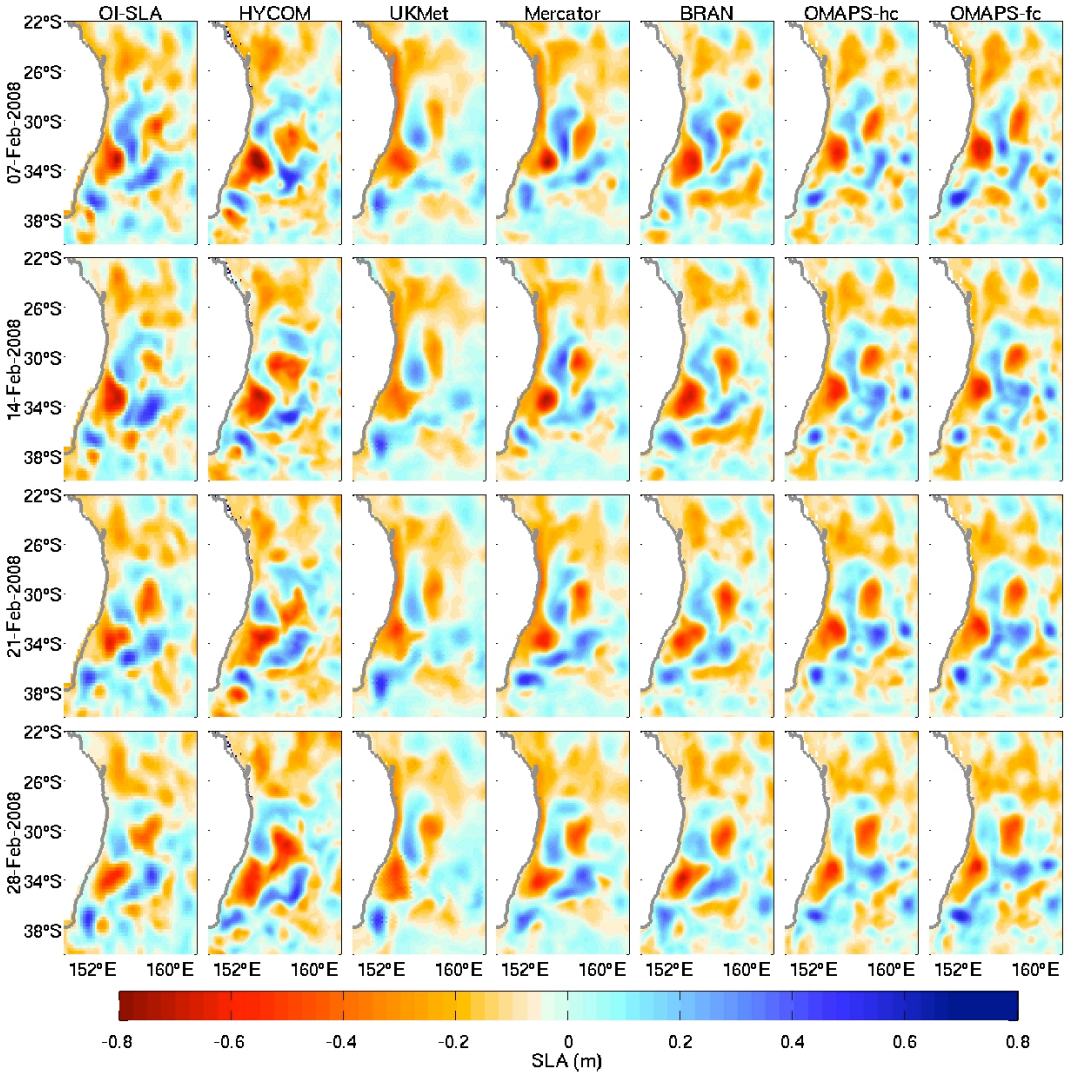

Figure 7: Sequence of daily averaged SLA during February 2008 from optimally interpolated observations (column 1), and each GODAE system (columns 2-7).

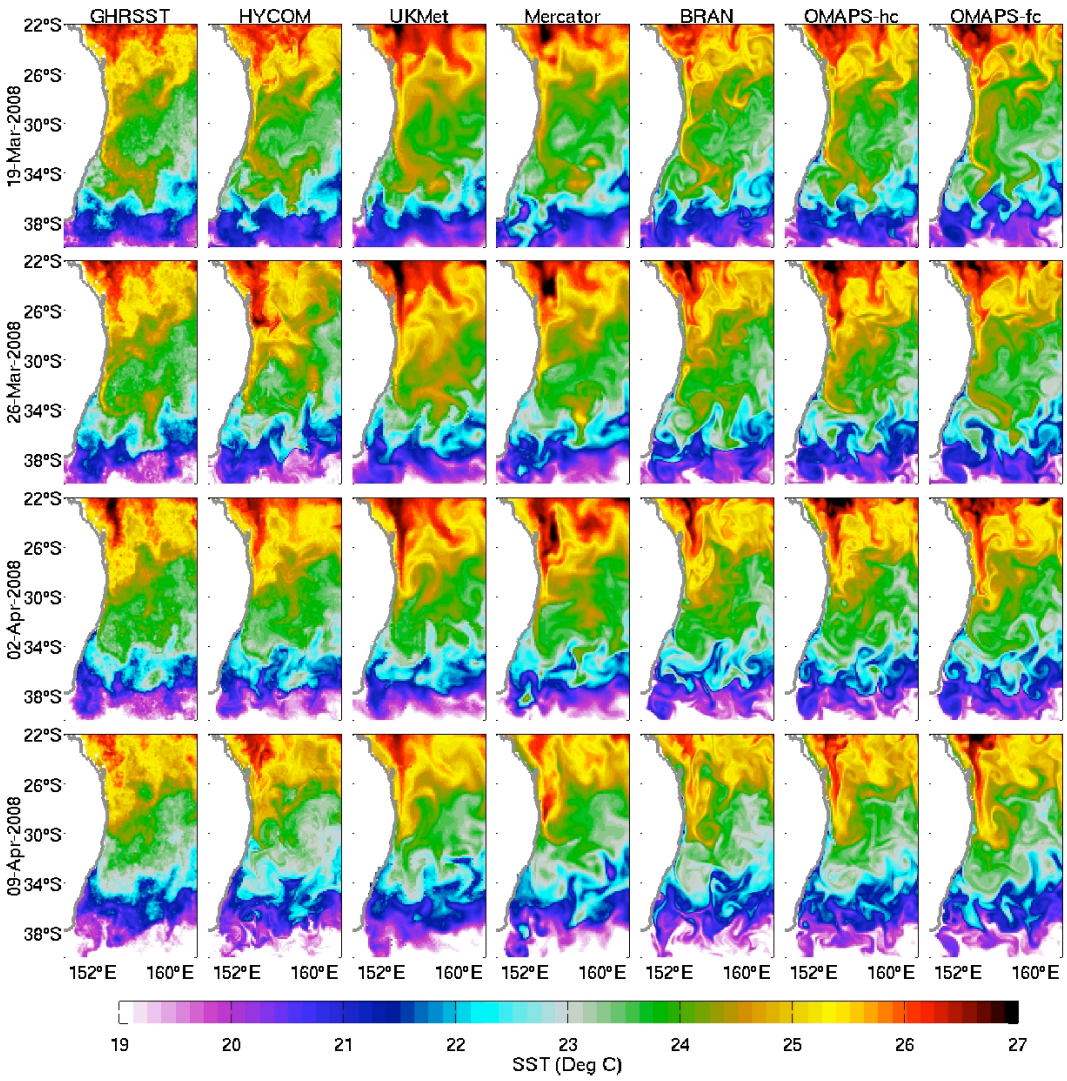

Figure 8: Sequence of daily averaged SST during March and April 2008 from a GHRSST analysis (column 1; RAMSSA, Beggs et al. 2010), and each GODAE system (columns 2-7).

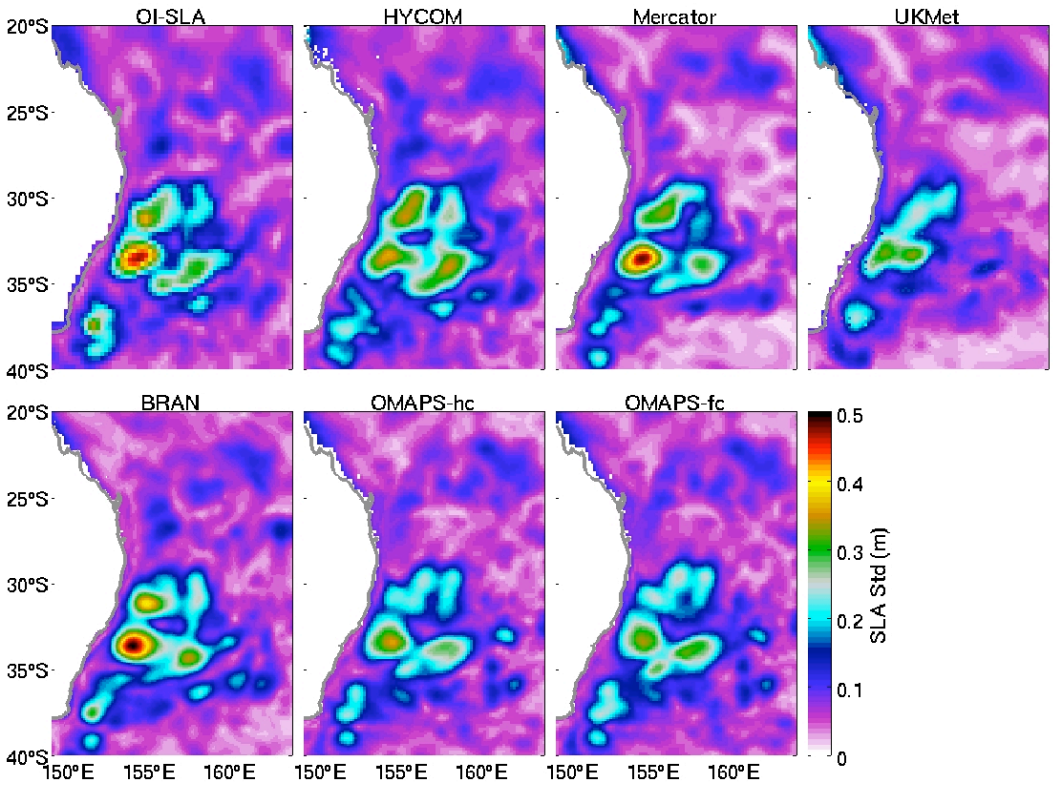

Figure 9: Standard deviation of SLA from optimally interpolated atSLA and each GODAE product.

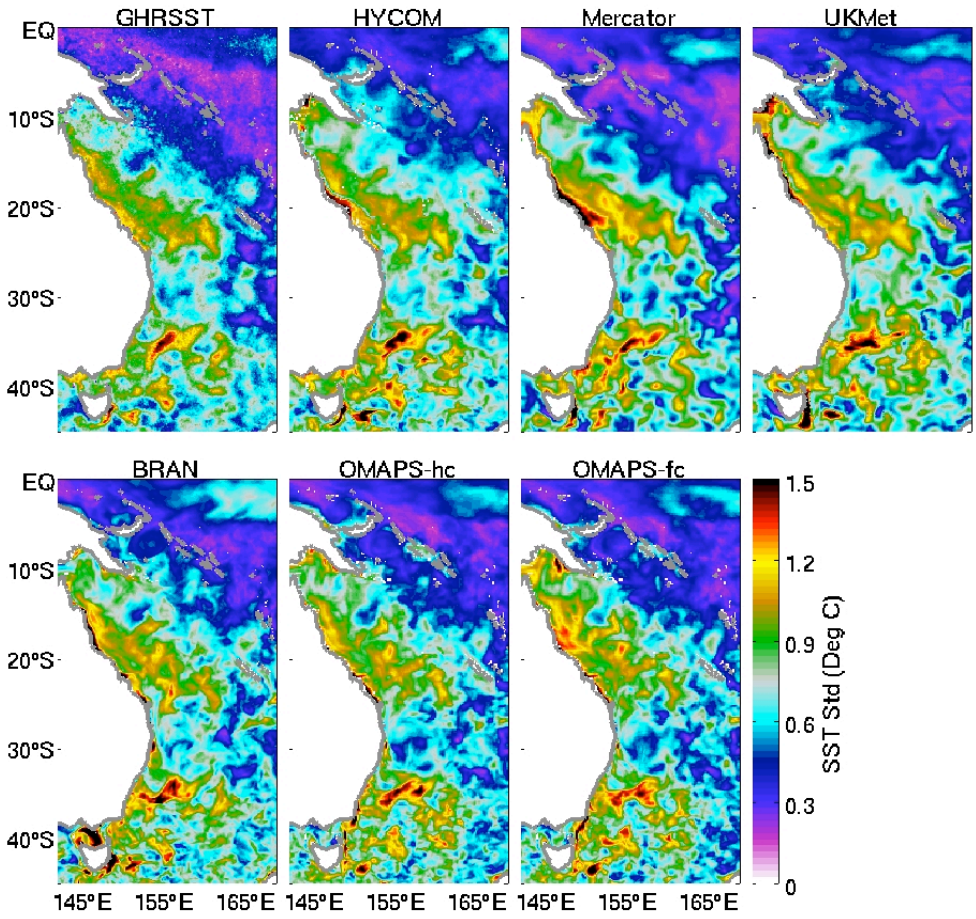

Figure 10: Standard deviation of SST from a GHRSST analysis and each GODAE product.