White Paper #4 – Operational Forecasting Systems

Balmaseda, M.A.[1], Kumar, A.[2], Andersson, E.[1], Takaya, Y.[3], Anderson, D., Janssen, P.[1], Martin, M.[4], and Fujii, Y.[1]

[1] European Centre for Medium-Range Weather Forecasts, United Kingdom
[2] National Centers for Environmental Prediction, United States
[3] Japan Meteorological Agency, Japan
[4] UK Meteorological Office, United Kingdom
[5] Meteorological Research Institute, Japan

1. Introduction

There is clear demand for reliable weather, marine and climate forecasts at different time scales for a variety of societal applications. The improvement of the observing systems, model development, and computer resources have pushed the operational forecasting activities to expand well beyond the traditional short-range (1-3 days) weather forecast. Current operational forecasting capabilities take advantage of better initialization techniques, incorporate probabilistic methods to cope with the chaotic nature of the atmosphere, and rely on coupled ocean -atmosphere models that can predict the slowly evolving sea-surface-temperature and its impact on the atmosphere, to progressively increase the lead time of forecast horizon.

Medium-range (10-15 days) weather and marine forecasts are now produced operationally in the major forecasting centers, as well as forecasts of climate at seasonal (up to 6-12 months lead time) time scales, and more recently, forecasts at subseasonal time scales (1-2 months lead time), bridging the gap between weather and climate. Figure 1.1 shows schematically the time and spatial scales characteristics of the different forecasting systems. Given the coupled nature of the ocean-atmosphere system, it is expected that the ocean will play an active role in the forecasting systems at all lead-times in the future[1].

Although the observational needs of the different forecasting systems vary, all of them revolve around four main activities: initialization of the ocean and atmosphere for subsequent prediction; model and data assimilation development; forecast verification and, in some cases, calibration of model output; and hindcasts for calibration and skill assessment. Both verification and calibration of forecasts require long ocean and atmosphere reanalyses and reforecasts. The reanalyses are also used for monitoring of the Earth System's climate. This paper discusses the requirements of ocean and boundary layer observations in the tropical Pacific from the perspective of forecasting systems spanning different time-scales: medium range (section 2), seasonal (section 3) and monthly (section 4), organized in order of their maturity. The observational needs for reanalyses, with emphasis on their use for climate monitoring, are discussed in more detailed in section 5. Section 6 provides a summary of recent and future developments, including decadal forecasts, coupled forecasting systems, as well as coupled ocean-atmosphere reanalyses and initialization of forecasting systems. Section 7 presents a summary of general considerations that are common to all the different applications, with

---

[1] For example, the ECMWF medium range ensemble prediction system has used a coupled ocean-wave-atmosphere model since November 2013.

emphasis on the interpretation of observing system experiments, model error, and the different applications of the observations. The paper ends with a summary of data requirements and specific recommendations (section 8).

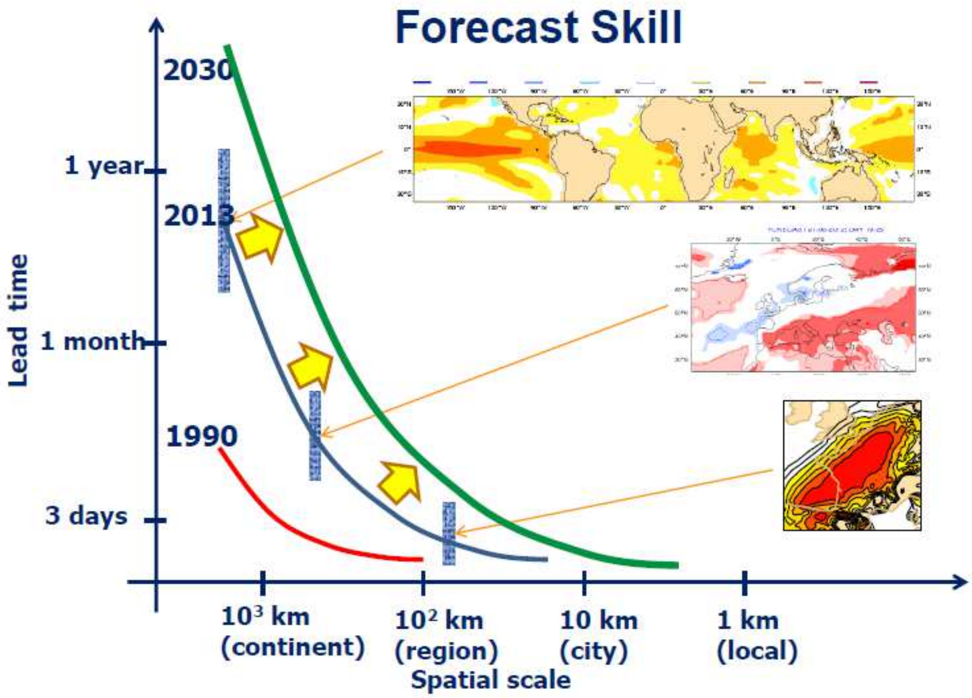


Figure 1.1 - Schematics of the dependence of forecast skill as a function of lead time and spatial resolution. The different size of the vertical bars is indicative of the characteristic averaging time (hours-days, weeks, seasons) for which the forecast is issued.

As such, the scope of the paper outlined above is already too wide, and there are topics not of relevance not explicitly addressed here, but are covered in other white papers. For example, the needs for medium-range high resolution marine forecasting systems are discussed in TPOS-2020 white paper 5 (WP5, Fujii et al., (2014)). The needs for model development regarding parameterization of atmospheric convection and boundary layer processes, a cross-cutting theme for forecasting activities on all time-scales, are discussed in TPOS-2020 WP11 (Cronin et al., 2014). Ocean biochemistry requirements are addressed in a separate paper (TPOS-2020 WP6, Mathis et al., 2014). The needs to calibrate and reprocess satellite data are covered in TPOS-2020 WP9 (Lindstrom et al., 2014).

To facilitate the discussion of the data needs, Appendix A introduces a naming convention regarding the processing level of the observations, the quality control, and the time-delivery properties.

2. Global NWP and Wave forecasts at the medium range

Global Numerical Weather Prediction (NWP) models are used to produce medium range weather forecasts (out to 15 days), with a horizontal resolution of typically 15-50 km and a vertical resolution of 10-30 m near the surface increasing to 500 m-1 km in the stratosphere.

There is a strong interest in using NWP model output to predict the risk for extremes or severe and damaging weather events. Statistical approaches based on forecast ensembles are used to predict the probability for extreme or rare events at longer lead times. Such ensembles require good knowledge of the uncertainty in all input data including the observations. Global NWP models are also used to provide boundary conditions for regional NWP models.

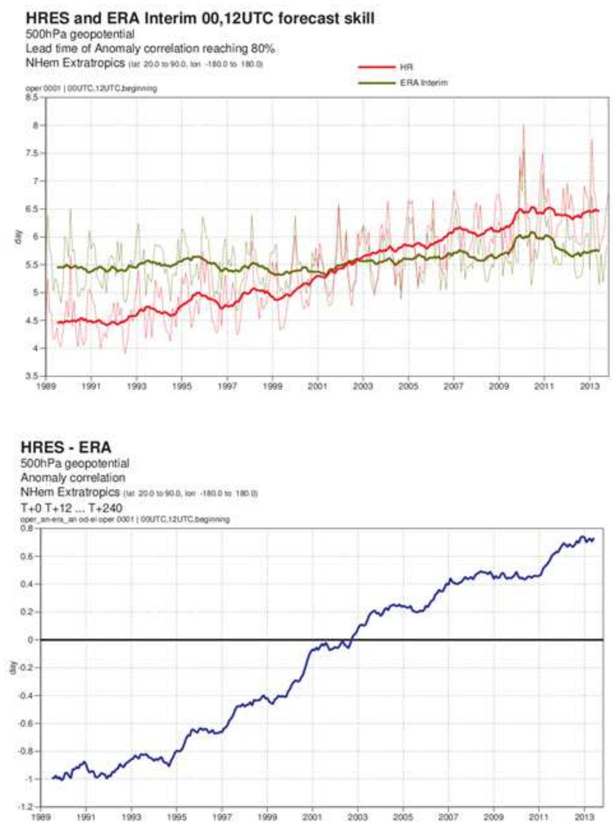


Figure 2.1 - Progress in the ECMWF operational NWP forecast skill since 1989 (OPS, red line) and in ERA-Interim (green line). The skill is measured as the forecast lead time when correlation drops below 80%. The statistics are for the Northern Hemisphere Z500. Progress is about 1 day per decade. The difference between OPS and ERA-Interim (lower panel) filters out changes due to intrinsic predictability of the atmosphere, and highlights progress due to model and data assimilation improvement. The observing system in the 90's is better exploited with the ERA-Interim forecasting system (which was state of the art one decade later) than with the forecasting system used at the time.

Figure 2.1 shows the progress in NWP forecast skill from the ECMWF operational forecasting system (red line, top panel). The metric is the forecast lead time when the anomaly correlation drops below 80% when predicting the geopotential at 500hPa (Z500) in the Northern Hemisphere. The progress is slow but steady (about 1 day per decade). The skill depends very much on the seasonal cycle, and it is convenient to filter the seasonal cycle with a 12-month running mean (red thick line). Another way of filtering out changes in the intrinsic predictability of the atmosphere is to compare with a reference experiment given by forecasts initialized from the ERA-Interim reanalysis system (Dee et al., 2011), which uses a frozen model and data assimilation cycle (the one operational around 2006). The differences between operational and ERA-Interim show the impact of model and data assimilation development efforts. The observing system in the 90's produces better skill when analyzed with a more advanced forecasting system (typical of the state of the art one decade later). Equally, increases in model resolution and continuous development are incorporated in the operational NWP system, which soon leaves ERa-Interim behind.

To initialize NWP models, an accurate estimate of the complete atmospheric state is required. Observations from surface-based, airborne and space-based platforms are all used to help define this initial state. Reliable error estimates of all observations are needed to estimate the accuracy of the initial state. The observational requirements for global NWP are based on the need to provide an accurate analysis of the complete atmospheric state and the Earth's surface at regular intervals (typically every 6 hours). Through a "data assimilation" system, new observations are used to update and improve an initial estimate of the atmospheric and surface states provided by an earlier short-range forecast. The uncertainty in the initial conditions is generally captured by ensembles of data assimilations.

The key atmospheric model variables for which observations are needed are: 3-dimensional fields of wind, temperature and humidity, and the 2-dimensional field of surface pressure. Also important are surface boundary variables, particularly sea surface temperature, soil moisture and vegetation, ice and snow cover. Of increasing importance in NWP systems are observations of cloud and precipitation. In the latter part of the medium-range, the upper layers of the ocean become increasingly important, and therefore, relevant observations of the ocean are also needed.

The highest benefit is derived from observations available in near real-time; NWP centers derive more benefit from observational data, particularly continuously generated asynoptic data (e.g. polar orbiting satellite data), the earlier they are received, with a goal of less than 30 minutes' delay for observations of geophysical quantities that vary rapidly in time. However, most centers can derive some benefit from data that is up to 6 hours old.

In general, conventional observations have limited horizontal resolution and coverage, but high accuracy and vertical resolution. In situ observations over the ocean or from remote land areas can occasionally be of vital importance. Also, a baseline network of in situ observations is currently necessary for calibrating the use of some satellite data. Observations are more important in some areas than in others; it is desirable to make more accurate analyses in areas where forecast errors grow rapidly, e.g. baroclinic zones and in areas of intense convection, such as the warm pool in the tropical Pacific.

2.1 Surface pressure and surface wind

Over the ocean, ships and buoys provide observations with good frequency. Accuracy is good for pressure and acceptable/marginal for wind. Coverage is generally good but marginal or absent over some areas in the tropics and the Arctic. The coverage in the tropical Pacific has degraded in recent years. Scatterometers on polar-orbiting satellites provide information on surface wind - with global coverage and acceptable horizontal and temporal resolution and accuracy. Scatterometers give information on both wind speed and direction, whereas passive microwave imagers provide information on wind speed only.

Surface pressure is not observed by present or planned satellite systems except for: some contribution from radio occultation data and measurements of differential atmospheric optical depth for a gas of known composition such as oxygen.

2.2 Sea surface temperature

Ships and buoys provide observations of sea surface temperature of good temporal frequency and accuracy. Coverage is marginal or absent over some areas of the Earth, but recent improvements in the in situ network have enhanced coverage considerably. Infrared instruments on polar satellites provide information with global coverage, good horizontal resolution and accuracy, except in areas that are persistently cloud-covered. Here data from passive microwave instruments on research satellites has been shown to be complementary. Observation of the diurnal cycle is becoming increasingly important, for which present and planned geostationary satellites offer a capability.

2.3 Ocean sub-surface variables

In the latter part of the medium-range (~7-15 days), the role of the sub-surface layers of the ocean becomes increasingly important, and hence observations of these variables become relevant. In this respect the requirements of global NWP are similar to those of seasonal and sub-seasonal forecasting (see section 3).

2.4 Sea State

Observations of the sea state from ships have become available since the middle of the nineteenth century. These observations are manually made and concern wave height, period and direction of the wind sea and the swell part of the sea state. The coverage of these observations is marginal over large areas of the Earth. Although ship manual observations have a great historical value, the observations are of marginal accuracy. Increases in accuracy can only be expected when observations by humans are replaced by instruments such as a shipborne wave recorder. The introduction of these will also benefit the observation frequency which is presently every 3 to 6 hours.

Moored buoy provide sea state observations of acceptable frequency (hourly) and acceptable quality. The buoys record the time series of the surface elevation which gives the frequency spectrum and, as a consequence, parameters such as wave height, peak period and several versions of mean period. Certain buoy types are able to give the directional wave spectrum, while buoys from the Canadian network produce estimates of maximum wave height, a parameter which is important for extreme sea states.

The highest quality wave height observations are nowadays provided by altimeters on board of polar orbiters. Although the coverage is global the spatial and temporal resolution is marginal as only along track observations are available. Despite this, these observations have played a major role in the improvement of the physics and numerics of ocean wave models. A number of weather centers use the altimeter wave heights in their wave height analysis.

Information on the low-frequency part of the two-dimensional spectrum can be obtained from Synthetic Aperture Radar (SAR) instruments. The accuracy is good, but horizontal and temporal resolution is marginal. These observations have been used in the wave analysis of a number of weather centers.

2.5 Data assimilation in the tropics

The Final Report of the Fifth WMO Workshop on the Impact of Various Observing Systems on Numerical Weather Prediction (Andersson et al., 2012) concluded: “Current global observing systems are heavily skewed towards mass observations over wind measurements, especially for the satellite components. And yet many studies presented at the workshop pointed to a higher than average impact of wind observations, both on a component and on a “per-observation” basis. There is a need to invest in enhanced wind observations in the tropics and over the oceans especially.”

A substantial fraction of the tropical large-scale variability can be explained by equatorially trapped waves; the equatorial waves coupled to convection can explain on average 60–70% of the error variance in the tropical atmosphere (Zagar et al., 2005). The largest part of this explained variance is represented by the equatorial Rossby (ER) modes, and a significant percentage pertains to the equatorial inertio-gravity (EIG) modes. Most likely, deep convection, acting as a generator of equatorial wave motion, is the dominant mechanism.

To ensure that the observational information is assimilated mainly in terms of Rossby modes, initialization procedures and methods for generating geostrophically balanced increments have been developed. In the tropics a dominant relationship similar to geostrophy is lacking; the analysis here has thus traditionally been undertaken in the univariate fashion. Consequently, large-scale divergence fields, such as the Hadley and Brewer–Dobson circulations, are analysed nearly univariately. Since the Global Observing System in the tropics relies on mass-field information, uncertainties in the analysed wind field are significant. Furthermore, large-scale motion in the tropics cannot be considered without taking into account inertio-gravity (IG) waves (Browning et al., 2000). In addition, the change of sign of the Coriolis parameter at the equator gives rise to important types of large-scale non-rotational motion, which are absent in the mid-latitude atmosphere: the Kelvin and mixed Rossby–gravity (MRG) modes. Indeed, equatorially trapped Kelvin, MRG and equatorial IG waves have regularly been detected in observations since the 1960s. Observational studies (Wheeler and Kiladis, 1999) identified equatorially trapped wave structures in long-term satellite observations of outgoing long-wave radiation, a proxy for deep tropical cloudiness. The waves have been denoted the ‘convectively-coupled equatorial waves’, as their presence in areas of moist convection implies an interaction between convection and the dynamics. In order to initialize these structures accurately, mass and wind information is required in the areas surrounding the equator, particularly in the tropical Pacific warm pool and surrounding areas.

2.6 Data Withdrawal Experiment: Impact of Winds from Moored Buoys

Traditionally, in-situ observations of surface pressure and wind field have played an important role in weather analysis and weather forecasting. Using data assimilation techniques, these observations provide information on the initial state of the forecast, while much can be learned from the validation of the forecast by means of these so-called conventional observations. Nowadays, however, we are facing a different situation because over the past 20 years weather forecasting centers have introduced a massive amount of satellite data in their analysis systems. The question now is whether in the presence of overwhelming amount of satellite observations the relatively small number of in-situ observations can still add information to the weather analysis in the Tropics.

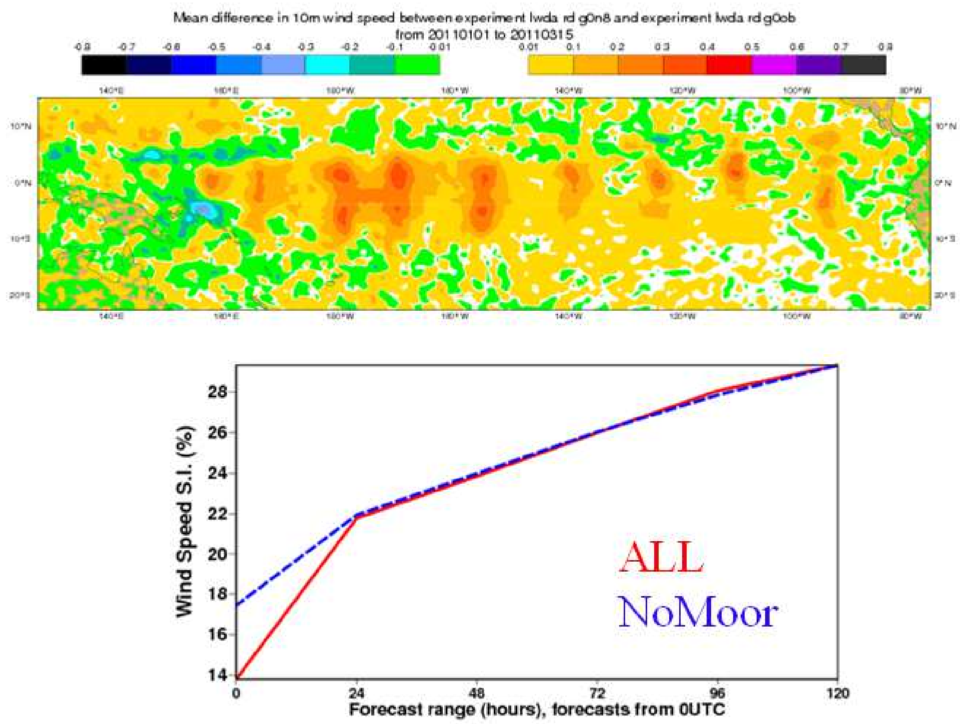


Figure 2.2 - Impact of withdrawing the wind information from the moored buoys in NWP experiments. Experiment ALL is the standard control experiment; experiment NoMoor is equivalent to ALL, but the wind from moored buoys is not used. (Top) Mean differences in analyzed 10m wind speed (ALL - NoMoor), for the period 20110101-20110315, showing that the moorings have a pronounced impact on the analyzed wind speed. (Bottom) Scatter index as a function of forecast lead time, verified against all tropical buoy data. This shows that the information from the moorings is quickly lost in the forecasts.

In order to investigate this, a data denial experiment was performed in which all the pressure and wind vector observations from moored buoys in the Tropics were removed (Bidlot and De Chiara, in preparation). Analysis and forecast results from the data denial experiment were subsequently compared with results from the control experiment, i.e., an experiment which includes all the relevant observations from the Tropics, but has otherwise an identical set-up. Recently weather centers experience a dramatic decrease in the amount of TAO array observations that are received through the GTS. For this reason we had to go back to the year 2011 to have a sufficient number of observations to do a meaningful experiment. Cycle 40R1, which is the latest cycle of the IFS, was chosen for the experimentation and analyses and 10 day forecasts were produced for the period 1 January 2011 until 31 March 2011. The spatial resolution was 40 km (corresponding to a T511 truncation in spectral space) while in the vertical there were 91 levels.

Figure 2.2 (top) shows the systematic difference between analysed surface wind speed from control and data denial experiments, averaged over a 2 1/2 month period. Differences are quite considerable, of the order of 0.4 m/s, and the use of buoy data leads to larger wind speeds in the analysis. It should also be clear from difference patterns where a number of TAO/TRITON moorings are located. This is particularly evident in the west Tropical Pacific. It may come as a surprise that removal of the buoy data in the Tropics has such a big impact on the wind speed analysis. It can be understood by realizing that the TAO array produces good quality wind vector observations on a frequent, hourly basis. They apparently can compete with the scatterometer observations on board polar orbiters as these visit the area where the buoys are located relatively infrequently.

The impact on the wind speed and wave height scores in the Tropical area is, however, fairly limited. This is illustrated by Figure 2.2 (bottom), which shows the scatter index (normalised standard deviation of error) in forecast wind speed, obtained from a comparison with buoy wind speeds. The area is the Tropics. It is evident from this plot that the impact of the Tropical buoy data on the forecast scores for surface winds is already fairly small after one day in the forecast. This also follows from a verification of the forecast wind speed against the control analysis, although in this case there is impact until Day 2 of the forecast. The impression is that the Tropical analysis of wind is not very well-balanced at the initial time, and ingested information is lost rapidly due to initial shocks. As already mentioned in the previous section, unlike the extra-tropics, there is no dominant balance relationship similar to geostrophy. This imbalance is supported by the experience that in the first 12 hours of the Tropical forecast there are signs of spin-down in, for example, the average Tropical wind speed.

Despite the fact that the weather analysis in the Tropics is relatively poor (hence more effort is needed to alleviate the problem with imbalance in the analysis) the in-situ observations of surface wind and pressure are considered to be of value for weather forecasting. In a coupled ocean-atmosphere context the observations are potentially of greater value, as parameters such as SST are quite sensitive to errors in the forcing wind field.

3. Seasonal Forecasts

3.1 Description of a Seasonal Forecasting System

Good-quality seasonal forecasts with reliable uncertainty estimates are of great value to society, allowing institutions and governments to plan actions to minimize risks, manage resources, and increase prosperity and security. Human and economic losses that may be caused by adverse climate events can be mitigated with early warning systems (e.g. famine, epidemics) and disaster preparedness. Equally, adequate planning can aid the exploitation of favorable climate conditions.

Seasonal forecasts predict variations in the atmospheric circulation in response to anomalous boundary forcing, changing significantly the probability of occurrence of weather patterns (Palmer and Anderson, 1994). Seasonal forecasting systems are based on coupled ocean-atmosphere general circulation models that predict both the surface boundary forcing and its impact on the atmospheric circulation. The chaotic nature of the atmosphere is taken into account by issuing probabilistic forecasts based on an ensemble of coupled integrations.

An added requirement for seasonal, and in general, extended-range forecast systems is correction for model biases. This step is required as forecast anomalies can easily be of the same magnitude as the model bias, and thus, can be overwhelmed by model errors. The bias correction of real-time forecasts is done by conducting a series of past seasonal hindcasts (also referred to asa reforecasts), which in turn requires initial conditions for a historical period (typically 15-25 years), usually obtained from reanalyses. The reforecasts are also needed for skill assessment for the seasonal forecast system that needs to be conveyed to the user community.

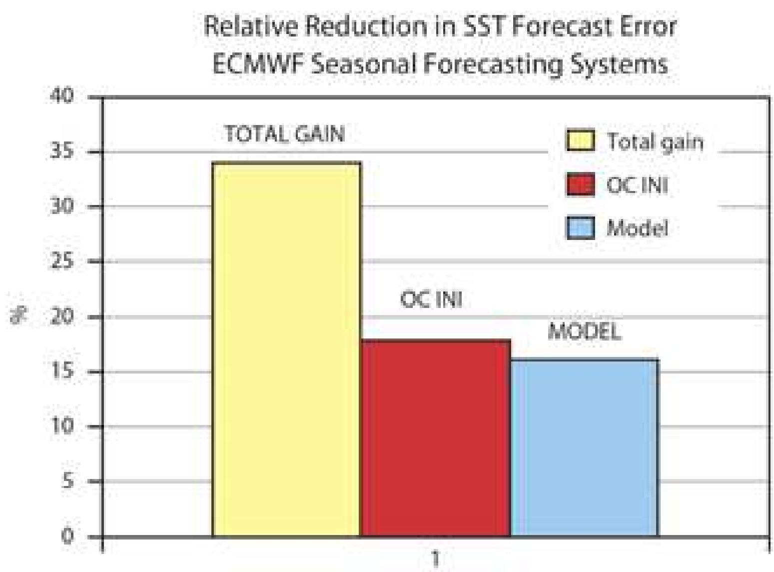


Figure 3.1.Progress in the seasonal forecast skill of the ECMWF operational system since it became operational around 1996. The yellow bar shows the relative reduction in mean absolute error of forecast of SST in the eastern Pacific (NINO3) integrated over the 1-6 months lead time. Contribution from model development (blue bar) and ocean initialization (red bar) are equally important. Developments in ocean and atmosphere models also contribute to the ocean initialization.

Of special importance for seasonal predictions are the variations of the tropical SST in the Pacific sector associated with El Niño Southern Oscillation (ENSO), and is the underpinning of operational seasonal prediction efforts. SST variations associated with ENSO alter the tropical convection and associated changes in heat sources lead to changes in atmospheric circulation. The importance of ENSO in seasonal forecasts is further enhanced by its relatively high potential predictability (Zebiak and Cane, 1987), which is largely inherent equatorial wave dynamics. Thus, the predictability of climate variability on seasonal time-scales depends critically on the adequacy of initial conditions of the ocean. See supplementary figure S1 for an illustration of the equatorial wave dynamics and impact on SST anomalies. However, linear wave dynamics is insufficient to predict the SST outcome, as can be seen in the same figure: not every eastern propagating Kelvin wave leads to an SST anomaly of the expected sign. Figure S1 also shows the wealth of temporal and spatial scales of the SST; especially noticeable is the tropical instability waves (TIW) activity in the Eastern Pacific.

Since seasonal forecasts became operational, their skill has been slowly but steadily increasing. The improvement in skill is equally attributed to better initialization of the ocean and improved coupled models, as shown in Figure 3.1 from Balmaseda et al. (2010). Improved initialization reflects not only the contribution of the ocean observing system, but also improved atmospheric surfaces fluxes, and better exploitation of the observations by more advanced data assimilation methods and models.

Seasonal forecasts use lower resolution models that those in NWP, mainly because the length of the integration, the number of ensemble members and the need for bias correction and calibration adds to the computational cost. The atmospheric model has a typical resolution of 0.5-1 degree in the horizontal, with 60 to 90 vertical levels. The ocean resolution is typically 1 degree (with equatorial refinement), although in the latest Met Office seasonal forecasting system the ocean resolution is 0.25 degrees (at expense of reducing the reforecast data set). The forecast lead time is typically 6-7 months, sometimes extended up to 12 months. The real-time forecasts require about 40-50 ensemble members. The reforecasts span a period of approximately 30 years, with hindcasts initialized every month using a reduced ensemble (~11-15 members). In total, about 200 years-worth of coupled model integration years are needed for a seasonal forecast at 7 months lead time initialized from a single calendar month. Or in other words, 2400 years-worth of coupled integrations are needed for seasonal forecasts initialized each month.

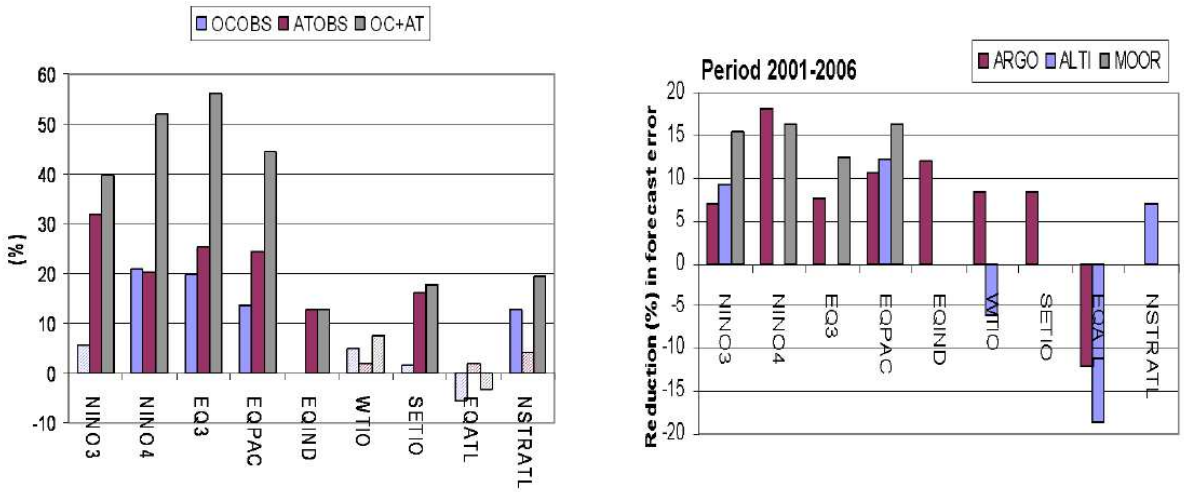
Seasonal forecasts use both the Near Real Time (NRT) data stream for initialization of real time, and the Behind Real Time (BRT) data stream in the reanalyses needed for the calibration data set. BRT data are also used for verification.

3.1 Ocean Initialization

The simplest way of providing initial conditions is to run an ocean model forced with observed winds and fresh-water fluxes from atmospheric reanalyses and with a strong constraint to observations of SST. Such a 0^{th} order ocean data assimilation system has been shown to generate realistic subsurface ocean structures (Luo et al., 2005; Kumar et al., 2013) in the equatorial Pacific. Although the information about wind forcing (wind stress or surface wind) and SST is essential to initialize seasonal forecasts, it is often not enough. The quality of the models and of surface forcing is not sufficient to provide an accurate estimation of the subsurface ocean state. By assimilating subsurface ocean observations it is possible to reduce the uncertainty in the ocean estimate and improve seasonal forecasts. See supplementary material for a summary of the current ocean observing system used in operational seasonal forecasts.

Figure 3.2 from Balmaseda and Anderson (2009) shows the contribution of ocean and atmospheric observations to the skill of seasonal forecasts, as well as the individual impact of different ocean observing systems. Their results highlight the importance of the surface wind information, and they also show than in the equatorial Pacific all the ocean observing systems contributed to the skill of seasonal forecast. These experiments were conducted with the previous ECMWF seasonal forecasting system (S3), and have been revisited with the new S4 (see below), as well as with a variety of other operational systems (see WP5). See also section 7.4 for a discussion on limitation of this methodology.

Aside from winds and SST, subsurface temperature observations are the next most important variable for the initialization of seasonal forecasts in the tropical Pacific. Salinity observations are also important (Yin et al 2011), especially in the mixed layer, and because they contribute to the assimilation of temperature data (by providing better constraints for density field. Altimeter-derived sea-level observations can also be helpful to constrain the upper thermal structure by projection onto the baroclinic ocean density structure. In order to obtain an accurate projection, the model vertical density structure needs to be reasonably realistic, i.e., in-situ observations of temperature and salinity are needed. The importance of altimeter-derived data is increasing with increased ocean model resolution. The assimilation of altimeter sea level needs additional information about the geoid which can be derived from gravity missions.



NINO3	5ºS 5ºN, 90-150ºW	EQIND	5ºS-5ºN, 40º-120ºE
NINO34	5ºS-5ºN, 170-120ºW	WTIO	10ºS-10ºN, 50º-70ºW
NINO4	5ºS-5ºN, 160ºE-150ºW	SETIO	10ºS-0ºN, 90º-110ºE
EQ3	5ºS-5ºN, 150ºE-170ºW	EQATL	5ºS-5ºN, 70ºW-30ºE
EQPAC	5ºS-5ºN, 130ºE-80ºW	NSTRATL	5ºN-28ºN, 80ºW-20ºE

Figure 3.2 - Impact of observations in forecast skill for different regions in table above, as measured by the reduction in mean absolute error for the forecast range. (Left) ocean observations (OCOBS), atmospheric observations (ATOBS) and both, for the forecast range 1-3 months, period 1987-2008. (Right) Impact of Argo, altimeter and moorings for the period 2001-2006. Results illustrate the importance of wind information, and also show that in the Equatorial Pacific all observing systems contribute to the skill.

Although the emphasis for initialization of seasonal forecasts is in the upper thermal structure (the upper 300m are more likely to influence the atmosphere on seasonal time scales), it does not mean that only observations of temperature in the upper ocean are needed. A full profile rather than a truncated one also makes a difference in the resulting stability of the water column. Equally, the upper thermal structure is better initialized with T/S profiles rather than only T (Troccoli et al., 2002; Ricci et al., 2005). However the assimilation of temperature and salinity separately (as it is common in variational assimilation methods) may induce problems, especially when the first guess salinity is lower than the observations, and the water column is not very strongly stratified.

There is large uncertainty in the fresh water flux (precipitation, evaporation and river runoff), affecting the surface salinity and mixed layer properties. It is probably the largest source of uncertainty in the estimation of salinity in the upper 100m. Information about Sea Surface Salinity (SSS) from either in-situ measurements close to the surface or from satellite (Aquarius or SMOS) can be useful (TPOS 2020 WP9 Lindstrom).

Aside from the biochemistry applications, time and spatially varying ocean color can be used as forcing fields of the ocean models to specify the depth for solar penetration. Ocean model simulations exhibit high sensitivity to ocean color. So far most of the ocean color products consist of climatologies and are not available in real time. A Level 4 (L4, see Appendix) time dependent ocean color maps, delivered in NRT or BRT will be useful.

Assimilation of altimeter-derived sea level (as opposed to sea level anomalies) needs information from the geoid. This is obtained from gravity missions. In addition, gravity missions can provide bottom pressure information, which can be used globally to constrain the non-steric part of global sea level variations. Gravity-derived variations of the global mass field are also useful for verification of ocean reanalyses (Balmaseda et al 2013a), BRT L4 data is desired. Bottom pressure also has the potential to constrain the barotropic mode; however, more experience is needed.

3.3 Weakness of ocean data assimilation

The assimilation of ocean observations in the equatorial wave-guide remains challenging, in spite of progress on data assimilation methods. Preliminary results from the Ocean ReAnalysis Intercomparsion Project (ORA-IP) show large spread in meridional mass and heat fluxes at the Equator (Valdivieso et al., in preparation). Observing system experiments (OSEs) show very little impact of the different observing systems right at the Equator (see WP5), which is indicative of either redundant information or poor assimilation methods. The relative impact of TAO and Argo at the equator is comparable, although varies among the data assimilation systems.

Fig 3.3 shows results from OSEs conducted with the ECMWF ORAS4 system, where Mooring, Altimeter and Argo are withdrawn from the ocean analyses, once at a time. The figure shows the fit (rms error) of the first-guess to mooring observations during the 10-day assimilation cycle (also called departures) in the Eastern Pacific (EQ1) and in the Western Pacific (EQ3). Since the observations have not yet been assimilated (the comparison is done just before the analysis), they can be considered fairly independent. In both Eastern and Western Pacific withdrawing the moorings increases the rms error. The assimilation of Argo does not seem to improve the fit to the moorings. In fact, in the western Pacific, the fit to the moorings is degraded when Argo is assimilated. These results suggest either that moorings and/or Argo provide different information (for instance, if they are at different locations and there is a lot of spatial structure) or/and that there are problems with the assimilation system. Curiously the assimilation of moorings does not degrade the fit to Argo (not shown).

There are several reasons for the small impact of observing systems at the Equator: i) it is difficult to constrain large scale biases with short spatial decorrelation scales (most of the assimilation methods only use one decorrelation scale) and ii) the equatorial dynamical balance needs longer time scales than the typical ocean assimilation cycle (1-to-10 days). Similar problems to those in the equatorial region are seen in coastal boundary current regions, where there is no dedicated observing system for the time-being. In the future, both equatorial and boundary current regions may become adequately sampled by Argo (see design plans in TPOS 2020 WP10). It can be argued that the observational needs are larger in the areas where model and data assimilation are poor, although it can also be argued otherwise.

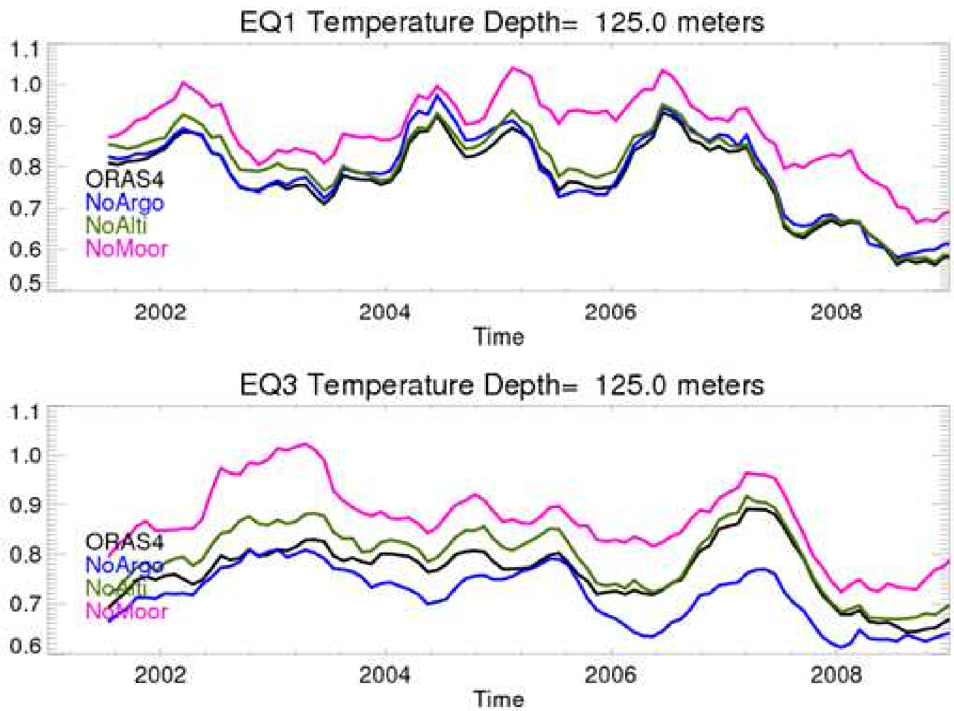


Figure 3.3 - Fit to mooring observations (root mean square error) in the Eastern Pacific EQ1 (top) and Western Pacific EQ3 (bottom) from the 10 days forecasts using the ECMWF ORAS4 data assimilation system (black), and in equivalent data withdrawal experiments NoArgo (blue), NoMoor(pink) and NoAlti (green), where Argo, Moorings and altimeter have been respectively removed. EQ1 (150W-90W, ±5^{o}), EQ3 (150E-170W, ±5^{o}). The verifying mooring observations have not yet being assimilated in any of the experiments. Withdrawing the moorings from the analyses degrades the fit of the 10-day forecast, in both Eastern and Western Pacific. Argo does not improve the fit to the moorings. On the contrary, in the Western Pacific withdrawing Argo improves the fit to the moorings. This can be interpreted as i) Argo and moorings providing different information or ii) sub-optimal data assimilation systems.

In summary, although the information from ocean observations is essential for initialization of seasonal forecasts, its extraction is not always straight forward. There is evidence that the current data assimilation systems are not exploiting the full potential of the observations (see Figure 3.3). Challenging areas are the Western Boundary Currents (WBC) and the Equatorial regions, where the information is quickly lost._ Constraining the density field by separate assimilation of temperature and salinity remains difficult, and so is the assimilation of altimeter sea level (both because the methods for projection into subsurface density and because the need of external information in form of geoid or MDT). It is expected that most of these problems will be solved by future developments in data assimilation methods and reduction of model error.

3.4 Model and Data Assimilation Development

Continued deficiencies in ocean and atmospheric models, and their corresponding data assimilation systems have led to ongoing developmental efforts. Model and data assimilation development efforts hinge on observational data sets for (a) validation of model simulations to document biases, and (b) testing and implementing new parameterization and data assimilation schemes. Although model parameterization often falls under the purview of focused field programs that target a specific process, sustained observations are helpful to evaluate model performance under various climate regimes. A good example is parameterization of stratus cloud decks over the western coast of continental areas (such as the western coast of equatorial South America), that are often associated with warm biases in the ocean models.

Another example of a familiar model error is the equatorial cold-tongue in coupled models. The attribution of the error still remains unresolved. From the ocean perspective, the cold tongue is usually associated with too strong zonal wind stress and/or too much poleward heat transport by the TIW. This could be due to errors in the atmospheric model (deficient resolution or others). In ocean-only simulations, the strong zonal wind stress can be mitigated by taking into account the ocean currents when deriving wind stress from atmospheric analysis winds. But this itself can lead to overcompensation and masking of the errors (there is some inconsistency in this formulation, since the atmospheric analysed winds -produced in uncoupled mode- have not seen the ocean currents). Having measurements of the TIW activity level and heat transports, as well as in-situ measurements of surface stress and winds is essential to solve this persistent problem.

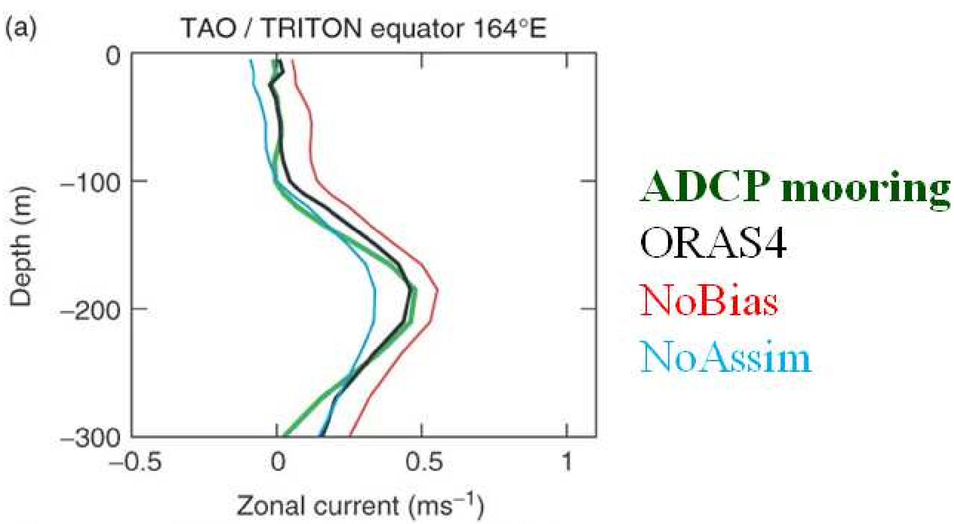


Figure 3.4 – Sub-surface currents for moorings are very valuable to verify ocean reanalyses and develop the data assimilation system.

Ocean currents from moorings have proved very useful in the development of data assimilation (see Figure 3.4). Often the assimilation of density information can lead to spurious circulations, especially at the equator. By looking at the impact on the equatorial undercurrent it is possible to assess if the data assimilation is adequately balanced. This has led to the development of balance constraints for equatorial velocity (second order geostrophic balance, Burgers et al., 2002), and the pressure bias correction suggested by Bell et al. (2004). This latter appears essential to obtain good velocity fields (Balmaseda et al., 2007). The scheme is quite sensitive to the choice of some parameters, which are tuned by using the currents from moorings.

Another method for validating the ocean data assimilation system is by monitoring error growth. It is easy to overfit the data if the only criterion is the fit of the analysis to the observations. More important is how the information is retained (or how the error grows during the very short forecasts, before the results get contaminated by model error). To this end, and in order to obtain reliable statistics, the sampling of the verifying observations should be homogeneous in space and time. The moorings provide an excellent data set for verification, since they guarantee similar number of observations at the same locations.

Sea level from tide gauges also provide valuable independent information for validation of ocean reanalyses, with the added benefit that some of them span long time-records. These are particularly important for evaluating the quality of the ocean re-analyses prior to the satellite period (Chepurin et al., 2013). Surface currents derived from the combination of altimeter and drifting buoys, such as the OSCAR product (Bonjean and Largelof, 2002) are also used for reanalyses verification.

It can be argued that for model and data assimilation development it is not necessary to have a continuous and permanent observing system in place. However, the errors of model and data assimilation are very flow dependent. Having a reduced data set for model/data assimilation development can lead to over-tuning algorithms for some case studies, which may then not be suitable for other cases. Numerical models, forcing fluxes and observations are also changing, and there is always the need to test new components with the most recent data (for example, it would be difficult to test the assimilation methods for altimeter without a reasonable coverage of in-situ observations).

3.5 Bias correction, verification and skill assessment

Seasonal forecasts require bias correction as the anomalies one seeks to predict can often be as large as model biases and can easily overwhelm the signal one strives to predict. The first order calibration is the a-posteriori removal of the mean bias, which depends on the lead time and on the seasonal cycle (Stockdale, 1997). This strategy assumes that the model bias is stationary, but this is not always the case. Figure 3.5 from Kumar et al. 2012 shows that the bias depends on the lead time and the seasonal cycle, a dependency that is accounted for in the a-posteriori removal of the bias. This figure also shows a non-stationary behavior in the bias, with a tendency towards colder (warmer) bias before (after) 1999, contrary to the assumptions in the a-posteriori bias correction. This can lead to complications in bias correction procedures and degradation of forecast skill. Kumar et al. (2012) discuss how changes in the observing system possibly led to non-stationarity in the forecast bias, and sub-optimal forecast skill.

The dominant change in the forecast bias in the CFSv2 system analyzed by Kumar et al. (2012) occurred around 1999, coinciding with the assimilation of AMSU data. Although some of the shifts in the reanalyses time-series can be attributed to the assimilation of AMSU, there are others that may be related to real shifts in nature (Zhang et al., 2012). Long reference time-series of good quality observations representative of the large scale circulation are therefore needed to be able to distinguish between spurious and real signals. See the section 5 on reanalysis for further discussion on this topic.

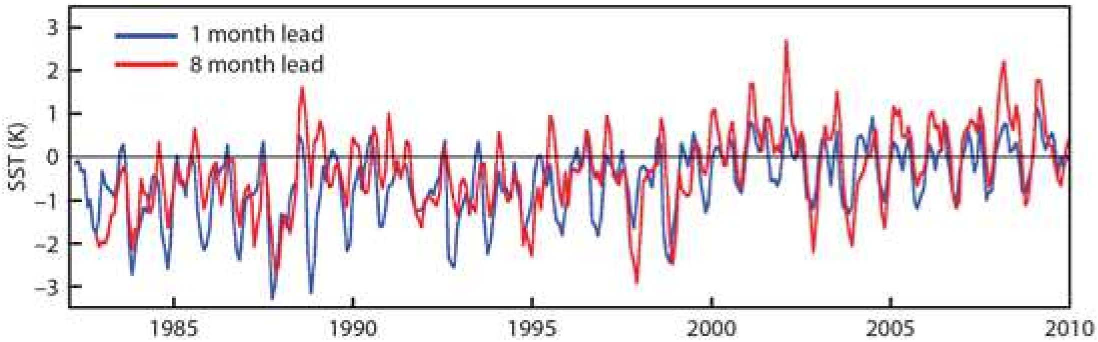


Figure 3.5 - Time evolution of the SST forecast bias in the NCEP CFS version 2. The figure shows the bias at 1-month and 8-months lead time, and it illustrates the non-stationarity of the bias (from Kumar et al., 2012).

For bias-correction and verification of seasonal forecasts, gridded maps (usually monthly means, 1x1 degree) of relevant variables (winds, precipitation, SST, OLR, surface fluxes etc.) are needed. These gridded fields are usually the results of post-processing, either via model-based reanalyses or other gridding algorithms. The quality of the seasonal forecasts will be influenced by the length and the quality of these products. The accuracy of these products is not considered the limiting factor for the forecast quality (with some exceptions such as precipitation, T2m over land, surface fluxes). The length of the temporal record and the stability of the error can be a reason for concern, even for variables like SST. Ideally, one would like records spanning a minimum of 30 years with stable errors and free of spurious variability and trends. These requirements are even stronger when it comes to the seasonal forecasts of extreme events.

3.6 Summary of seasonal forecasts data needs

(See appendix for an explanation of the acronyms used in this section).

<u>Initialization</u>

- Surface winds/wind stress (L4), SST(L2-L3-L4); subsurface temperature/salinity (L2, L2-QC) and sea-level altimeter (L2, L2-QC) are essential variables for initialization;
- Equatorial wave-guide needs intense sampling, which currently is only provided by moorings, but could be better sampled with the new Iridium Argo floats which avoid drift by staying at the surface for only a few minutes;
- Increased horizontal resolution model initialization needs high spatial resolution altimetry;
- SSS, ocean color, heat, freshwater, and turbulent kinetic energy surface fluxes (L3-L4);
- Gravity derived geoids and bottom pressure complementary to altimeter (L3-L4);
- Delivery in two streams: NRT (no more than 24 hours delay) and QC BRT (with delays ranging from a few days to update the current reanalyses, to years or decades, to be used in future reanalyses and verification).

Model and Data Assimilation Development

- Independent data for validation of ocean data assimilation and models: current profiles at the Equator (provided by moorings, L2-QC, L3); sea surface currents (SSC, L4); sea level from tide gauges. Long records. Time series of L2-QC, L3;
- Quality controlled flux data from reference sites (wind, wind stress, long and shortwave radiation, relative humidity, surface temperature, rain gauges). Long time series of L2-QC, L3;
- Controlled profiles of in-situ surface and subsurface data for validation of ocean-atmospheric reanalyses and models. Long time-series of L2-QC, L3;
- Processed gridded products of surface fluxes. Long records of L3-L4;
- Indonesian Throughflow transports (heat, salt, volume) time-series;
- Equatorial transports (heat, salt, volume). Time series.

Bias correction, Verification and Skill Assessment

- Long (>30 years) stable ocean-atmospheric and SST reanalyses for initialization of hindcasts;
- Long (>30 years) stable records of end-user related variables (such as surface winds, precipitation, surface temperature, sea-level pressure). Other indirect meteorological variables that can help the calibration and interpretation of forecasts are useful (Z500, OLR). (L3-L4);
- Long records of L2-QC, L3 variables in reference sites;
- Continuous delivery BRT, preferably with delays no longer than 1-3 months, for prompt verification.

4. Sub-seasonal forecasts

Sub-seasonal forecasts are currently produced operationally at various major forecasting centers. Configurations of models range from an uncoupled atmospheric model to coupled ocean-atmosphere models (Table 1). Oceanic observations required for this application may be similar to those for medium-range and/or seasonal forecasts, but there are some differences. Regardless of whether a model is coupled or uncoupled, the sub-seasonal forecast requires ocean analysis (SST or sub-surface analysis) and observations of influence to be consistent in quality over a long period, since the sub-seasonal forecasts also need bias correction based on re-forecasts similar to that for their seasonal counterparts (Section 3). The poor time-consistency of the ocean observations and analysis may hamper a proper calibration and fail to provide opportunity for gains on forecast skill. Furthermore, the sub-seasonal forecasts may be performed with models at a higher resolution (up to ~30 km for the ECMWF monthly forecast system) than seasonal forecasts and the higher resolution models ideally need with higher SST conditions. Hence the incentive to use higher-resolution SST boundary/initial conditions requiring higher-resolution observations by a combination of in-situ and satellite measurements.

There is some phenomenological rationale for the requirement of oceanic observation for the sub-seasonal forecast, the primary one may be the Madden-Julian Oscillation (MJO). The MJO is the pronounced variability in an intraseasonal time-scale (30-90 days), accompanying coherent deep convection and large-scale atmospheric circulations in the tropics. The MJO has a strong influence on tropical weather as well as extra-tropical weather through so-called teleconnections (Cassou, 2008; Mori and Watanabe, 2008; Vitart, 2013). Better representation of the MJO and the teleconnections should lead to better skill of sub-seasonal forecasts (Vitart 2013). Many modeling studies have been conducted to evaluate impacts of ocean coupling in predicting MJO, and indicate that the ocean coupling contributes to improve a representation and forecast skill of MJO (e.g., Klingaman and Woolnough, 2013; Woolnough et al., 2007). Meanwhile in-situ and satellite observations have revealed that the MJO is related to the ocean temperature and ocean salinity variations (Anderson et al., 1998; Grunseich et al., 2013; Matthew et al., 2010). Given that the MJO is a coupled atmosphere-ocean phenomenon as the many sensitivity experiments have suggested, better analysis of ocean states should bring better MJO forecasts and sub-seasonal forecasts, at least in principle.

In the traditional ocean data assimilation for seasonal forecasts, the main focus is on relatively large spatial scale and low-frequency variability (i.e. oceanic equatorial Kelvin and Rossby waves and ENSO, Indian Ocean Dipole (IOD)), while the future sub-seasonal forecasting would shed light on the small-scale atmosphere-ocean interaction over tropical instability waves or ocean fronts (Small et al., 2008; Kelly et al., 2010). In the foreseeable future, oceanic observations with finer spatial and time resolutions would be required in order for these phenomena to be analyzed and initialized. Although an oceanic contribution to improving sub-seasonal forecasts has been recognized as mentioned above, an evaluation of the observing system from the sub-seasonal forecast perspective has never been conducted and a research effort should be made to explore the benefit of oceanic observations for sub-seasonal forecasts in the near future.

Table 1 - Operational sub-seasonal prediction systems and utilized ocean observations (as of April 2013).

Institutions	Resolution	Ocean coupling	Ocean observations
ECMWF	TL639L62 (day0-10) TL319L62 (day10-32)	Yes	1,2,3,4,5,6
JMA	TL159L60	No	1,2,3,4
UKMO	N216 (~50km)L85	Yes	1,2,3,4,5,6
NCEP	T126L64	Yes	1,2,3,4,5,6
EC	T63L31/T63L35	Yes	1,2,3,4,5,6
CAWCR	T47L17	Yes	1,2,3,4,5,6
KMA	T106L21	No	
CMA	T63L16	No	
CPTEC	T126L28	Yes	1,2,3,4,5,6
HMCR	1.1x1.4 deg., L28	No	

1: Mooring buoy (in-situ, TAO/TRITON/PIRATA) 2: Argo float (in-situ)
3: Drifter buoy (in-situ) 4: Ship (in-situ)
5: Altimeter (satellite) 6: Infrared/Microwave (satellite)

5. Reanalyses

As discussed before, extended-range predictions require calibration of real-time forecast anomalies to reduce the impact of model errors. This is achieved by running the forecast system back in time and developing a database that can be used for quantifying the statistics of model biases, against which real-time forecasts are bias corrected. Running the extended-range forecast system back in time, generally referred to as the reforecast, requires initialized ocean forecasts, and availability of ocean analyses for the initial conditions. This is one of the reasons for conducting the ocean reanalysis going back in time, and which are updated on a periodic basis either after sufficient advancements in models and assimilation systems have been made or when new sets of observations are added in the historical data bases through data mining efforts or improved quality control methodology. Reanalyses (both for the ocean and atmosphere), therefore, are integral components of real-time forecast systems, while at the same time, are extensively used for climate monitoring efforts to place the evolution of the current climate system into a historical context.

Reanalysis efforts, although do not explicitly pertain to real-time ocean and atmospheric observing systems, do provide guidance on (a) the influence of the changing observing system on the analyses and forecasts, and (b) a means of testing the influence of various observing platforms on extended-range forecasts through Observing System Experiments (OSEs), and are briefly discussed next.

Observing systems are in a state of continual evolution either due to development of new technologies, for example, Argo, or due to the phase-out of older observing systems, such as a decline in XBTs (see Figure S3). Such changes in observing systems, even if intended to improve the quality of ocean analysis, can also lead to discontinuities when changes in the ocean observing system interact with the data assimilation. An example could be that as the density of observation increases, the ocean analysis is drawn more towards the observations away from the assimilation system's initial guess state (which is based on the model forecast). If the initial guess state (that is obtained by a forward integration of the assimilation model has biases), then a sudden appearance of a new observing system in a data void region can create a spurious jump in the analyses from a model state towards in situ observations. Similar discontinuities could occur due to changes in the QC systems or changes in the correction to raw observations, such as XBT fall rate correction (Wijffels et al., 2006), or the reported pressure sensor biases in the Argo floats (Lymann et al., 2006). Influence of such changes in the reanalysis can subsequently affect the forecast biases, so invalidating one of the fundamental assumptions of forecast calibration (Figure 3.5). Such discontinuities in the historical analysis, and their influence on the reforecasts, also provide valuable lessons for the real-time analysis of the ocean state that can occur due to ongoing changes in the observing system, and care needs to be taken in the design and evaluation any of tropical Pacific observing system.

Reanalyses also provide an opportunity for conducting OSEs, whereby influence of a particular observational platform on ocean analysis and subsequent forecast skill can be assessed. The importance of OSEs in the context of TPOS 2020 is of obvious importance for the design of observing system to provide adequate initial states for operational forecasting systems for different time-scales. Reanalysis based on OSEs are also a means of quantifying the accuracy requirements for the observing system in the context of forecasts. Although OSEs represent a computationally expensive exercise, primarily due to requirements of doing an extensive set of reforecasts to obtain results that are statistically significant (for more discussion see white paper 5), it is recommended that a robust OSE activity involving multiple operational forecasting centers should be maintained to inform the design and assessment of the current and future configuration of the tropical Pacific observing system. In summary, although the reanalysis efforts may not be of direct relevance to the real-time observing systems, they encompass a set of tools that are of value for assessing and informing the design of the tropical Pacific observing system.

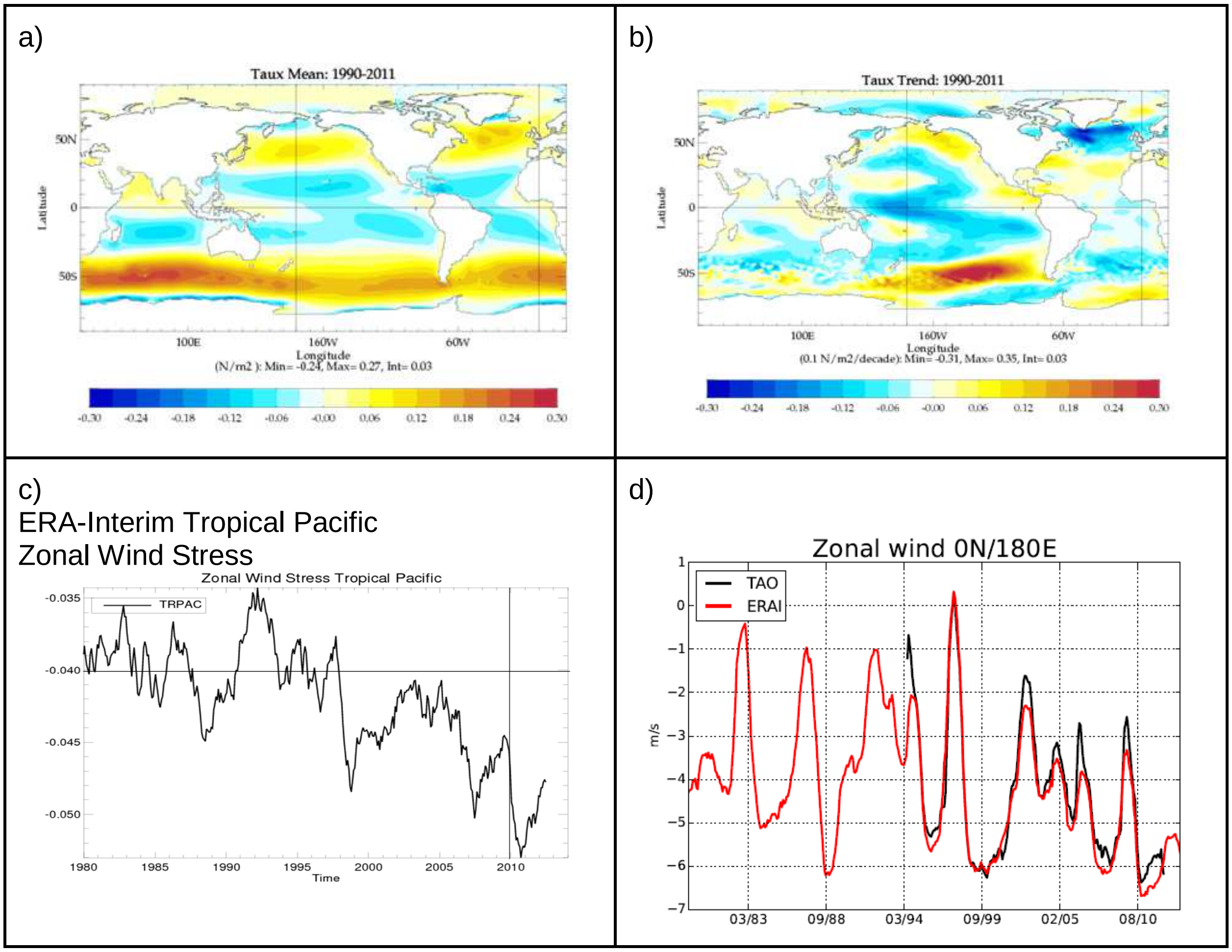


Figure 5.1 - a) Mean zonal mean stress from Era-Interim (EI) (1990-2011). b) Linear trend of EI zonal mean stress for the same period. c) Time series of EI zonal wind stress in the Tropical Pacific and d) EI and TAO zonal wind at 0N-180E.

Atmospheric reanalyses are undoubtedly a great asset for climate research and forecasts applications. They are so widely used that often are taken as truth. Atmospheric reanalyses, as well as the oceanic reanalyses, can present spurious signals due to changes in the observing system (Zhang et al., 2012; for instance, see section 3 above). When changes in the observing system coincide with real changes in climate variability, the true and spurious signals are difficult to disentangle. This is the case for the interesting changes in tropical Pacific winds after the 1998-1999 La Niña event.

Figure 5.1 shows the maps of zonal wind stress from ERA-Interim (mean and linear trend, top), and corresponding time-series averaged over the tropical Pacific. Pronounced changes can be seen in the zonal wind, consistent with changes in the gyre circulation. The time-series shows that the trend is punctuated by intense easterly episodes (1999, 2006-2007, 2009-2010). After 2009 (with the end of Quickscat) no scatterometer winds are assimilated in ERA-Interim. Establishing if the changes in the reanalysis wind stress are robust is vital for the understanding and prediction of climate. For instance, Balmaseda et al. (2013) and England et al. (2014) argue that wind variability is instrumental for the ocean heat uptake. Backscatter from altimeter also detected some extreme signals in the surface winds around 2010, but doubts were cast about the calibration (Abdalla, personal communication). The record from scatterometer winds is often too discontinuous to provide a reliable assessment. The TAO moorings offer a more consistent picture, although of limited spatial extent (De Boisseson et al., in preparation). The implications for the observing system are obvious: there is a clear need for good quality reference long time series, stable in time and representative in space. Spatial sampling to obtain this representativeness needs to be evaluated.

6. Future Developments

6.1 Decadal predictions

Decadal forecasting is a rapidly evolving field. External forcing influences the predictions throughout, but the initial state influenced by natural variability also plays an important role in the evolution of coupled system on shorter lead time, e.g., in the first six-to-nine years. Hawkins and Sutton (2009), Meehl et al. (2013) discuss the importance of initialization to produce time-evolving predictions of regional climate. The analysis of multi-model ensemble results suggesting that some aspects of decadal variability like the mid-1970s shift in the Pacific, the mid-1990s shift in the western Pacific, and the early-2000s hiatus, are better represented in initialized hindcasts compared to uninitialized simulations. The difference between initialized and non-initialized forecasts becomes more evident when using the multi-model ensemble than in any individual forecasting system. The tropical Pacific appears as one of the regions with real predictive skill at 1-2 year lead times arising from the initialization of the ocean (Pohlmann et al., 2013).

Some of the decadal forecasting systems rely on information from existing atmospheric and ocean reanalyses (the later produced for seasonal forecasts) for initialization. This information is used directly, by nudging the coupled model (anomaly or full field) or by forcing an ocean model with atmospheric fluxes. Some systems use a specific data assimilation system designed for decadal forecasts, such as the Met Office system (Smith and Murphy, 2007). A key difference of initialized decadal predictions from initialized predictions on shorter time-scales is the need for ocean observations into the deeper ocean.

From the Tropical Pacific perspective, the observational needs for decadal forecasting are similar to those of seasonal, except the need for even longer and stable observational records is stronger, both for initialization and verification. At decadal timescales the deeper ocean (below

500m) also plays a role and observations up to 2000m are considered important. Results from synthetic observing system experiments suggest that even observations below 2000m are likely to play a role, especially in the prediction of the Atlantic Meridional Overturning Circulation. Initializing the large scale modes of decadal variability such as the PDO may be important.

6.2 Coupled forecasting Systems: thermal and dynamic coupling

Whenever the term coupled ocean-atmosphere system is used, the thermodynamic coupling in the tropics springs to mind, where the atmosphere responds to the SST values and its gradients. And indeed, the benefits of the ocean-atmosphere thermal coupling in the tropics for MJO and NWP prediction have been demonstrated (De Boisseson et al. 2012, Janssen et al. 2013, for instance). However, there are other processes in the air-sea interaction related to the momentum transfer that have potential impact on the resulting SST.

Evidence is growing that ocean wave dynamics plays an important role in the evolution of currents and temperatures of the upper ocean, suggesting that it makes sense to develop a tightly coupled ocean-wave, ocean circulation model. The first indication came from the theoretical work of Craig and Banner (1994) and the experimental work of Terray et al. (1996) that highlighted the prominent role of breaking waves in the upper-ocean mixing of momentum and heat. Monin-Obukhov similarity is based on the balance between production and dissipation of turbulent kinetic energy and breaking waves generate so much additional turbulence that large deviations from similarity occur resulting in enhanced mixing. Additional sea state effects that are relevant for the upper ocean are the generation of turbulence by Langmuir circulation (McWilliams and Restrepo, 1999; Grant and Belcher, 2009) and the Stokes-Coriolis force (Hasselmann, 1970; McWilliams and Restrepo, 1999). Furthermore, although it is well-known that ocean waves experience refraction in the presence of spatially uniform currents, it is perhaps less well-known that by Newtons third law this implies that there will be a sea-state dependent force exerted on the ocean (Garrett, 1976). In return, it should be emphasized that ocean circulation also affects the sea state, through for example current refraction and through Doppler shifting. In an inhomogeneous current system the refraction may lead to focusing of wave energy, triggering the generation of freak waves.

The development and validation of a coupled ocean, ocean-wave, and atmosphere system is a very interesting development, and opens the door to better exploitation of the observations. Data on the ocean and atmosphere boundary layer, as well as on sea-state, are of vital importance in order to validate and further develop these coupled systems. An example of the use of in-situ data in the development of a mixed layer model, including sea state effects, is given in Janssen (2012).

6.3 Coupled Data Assimilation

A recent focus for development is the implementation of coupled ocean-atmosphere-land-sea ice data assimilation (CDA) systems which are expected to improve forecasts at various time ranges (from medium range out to seasonal and decadal). These developments, initiated with the weakly coupled reanalyses at NCEP (Saha et al., 2010), are gaining momentum at different operational and research centers (Met Office, ECMWF, BMRC, JMA-MRI, NASA). Observations of the air-sea interface are crucial to better understand the important coupled processes which should be represented in the CDA systems. Observations in the boundary layers of both the atmosphere and ocean can help to constrain these coupled processes in this important area.

CDA is an important approach to improving the forecast skills for the phenomena in which the air-sea interaction plays an essential role. Fujii et al. (2009) demonstrated that distribution and variability of precipitation in the tropics are improved in their weakly CDA run, in which ocean observation data alone is assimilated into a coupled model, compared to an AMIP run (i.e., a free simulation of the atmospheric model using the observed SST temperature). They found that the negative feedback between the change of SST and atmospheric convective activity is not properly represented in the AMIP run due to the prescribed SST, but it is recovered in the CDA run (see the negative correlation between SST and precipitation around the Philippine Sea in the CDA run (see Figure 6.1). They also showed that the negative feedback improves the precipitation fields and atmospheric circulations.

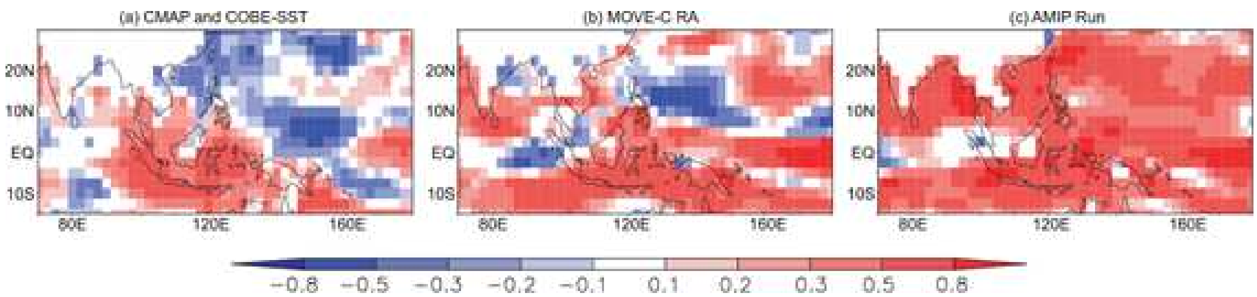


Figure 6.1 - Maps of the Correlations between SST and precipitation in summer (June-August) calculated from (a) CMAP and the gridded SST data in JMA, (b) the weekly CDA run, (c) the AMIP run (adjusted from Fujii et al., 2009).

A few international meetings have taken place to discuss progress in this area, one of which was organized jointly by WGNE and GODAE OceanView in March 2013. A series of white papers are being developed following that meeting and the current status of those is described at www.godae-oceanview.org/outreach/meetings-workshops/task-team-meetings/coupled-prediction-workshop-gov-wgne-2013/white-papers.

It is envisaged that in the near future the SST analysis and initialization will be carried out with such CDA systems. L2 SST bias-corrected data will become necessary (as opposed to using a gridded L4 SST product), frequently enough to be able to represent the diurnal cycle.

One of the most challenging aspects of CDA is the formulation of the error covariance between the variables involved in the air sea interaction. Balance relationships between variations in the ocean mixed layer and in the atmospheric boundary layer are needed. These can be obtained from model integrations, but verification data are needed. From this perspective, ocean and atmosphere observations collocated in a single column are considered very valuable. No systematic studies have been carried out as yet to assess directly the impact of tropical moored buoy data on these systems. However, the tropical moored buoys are expected to provide a very useful input to develop and improve these coupled systems.

7. General considerations

7.1 Different lives of an observation

It is perhaps not always realized how often and in how many ways an observation or set of observations is used. The most immediate way is in weather and climate forecasting via the analyses used to initiate short range high resolution (~4km) weather forecasts up to 2 days, moderate resolution (20km) forecasts up to 10 days, ensembles of medium resolution forecasts up to 15 days, ensembles of intraseasonal forecasts (40km) up to 30 days, seasonal forecasts (80km) up to a year, and in the future, as decadal forecasts become more mature, to multi-annual time scales.

A completely different stream would use the same observations, not in near-real-time analyses, but in delayed reanalyses. These are used for validation and model bias correction and calibration, whereby some aspects of model deficiencies can be corrected. Calibration is applied in some way to all forecasts from days to years.

An interesting feature of reuse of data is that it influences not just real-time forecasts of today but will influence forecasts for many years into the future. Additionally much greater use can be made of observations today than was possible when they were taken 10, 20, 50, etc. years ago. This results from the improvements in models and in analysis procedures.

A third stream of use of an observation is in scientific research, ranging from improving parameterisation of physical processes to greater understanding of major phenomena such as the Intraseasonal Oscillation and ENSO.

Finally, an aspect of the sustained observing system often overlooked, is its contribution to the forecast verification. There is a lot of experience gained during the routine monitoring and verification of forecasts (or learning from errors). Verification is essential to identify different sort of errors, and it often sets the directions for model development.

7.2 Panorama for TPOS 2020

7.2.1 Ready-Get Set-Go

The IRI and the Red Cross have been advocating an approach to the use of forecasts of many timescales, in order to improve their use for society. The idea is a simple one. Use the seasonal forecasts to provide an outlook of potential major climatic anomalies, several months in advance, so that steps can be taken to start getting ready. The next stage (the GET SET stage) is to use intraseasonal forecasts which will start to refine the nature of the anomaly and refine the probability of what happening where. This is refined further through the use of medium range forecasts in the GO stage.

7.2.2 Interpretation of climate and decadal signals

As well as expanding the forecast capabilities from the seasonal to daily timescales, as illustrated with the Ready-Get Set-Go example, TPOS 2020 should be ready to serve the needs for climate monitoring and decadal forecasting. There is increasing need to detect and predict decadal variability. Indeed, the Pacific basin has received a lot of attention in relation with the recent hiatus of the surface warming. Quoting the feature news of Nature (16th January 2014), "And here, the spotlight falls on the Equatorial Pacific".

7.3 Continuous but steady progress

There is a perception that progress in seasonal forecasting has plateaued and that there has been little progress in the last few years. We think that this not the case. While it is generally accepted that seasonal forecasting is a more difficult problem than was initially conceived in the early days of TOGA (1985-1994) progress is being made, albeit more steady than revolutionary, but if experience in weather forecasting is anything to go by, this is hardly surprising.

Figures 2.1 and 3.1 show progress over the years from the ECMWF medium range and seasonal systems respectively. Progress is being made on two fronts. Not only is the RMS error of the forecasts being reduced (the usual measure of skill), but the forecasts are becoming more reliable in that the growth in the ensemble spread now matches the growth of error; the forecasts are no longer far too confident with the observed values lying frequently outside the ensemble, when retrospectively verified.

What may be less reliably known is if there is increased skill in predicting the evolution of large events. There are only two in the period with better observations (82/3 and 96/7). With a sample of only two events it is difficult to validate the models reliably. Attempts are being made to use more extended atmospheric reanalyses, going back a hundred years buoyed by the encouraging results from simpler models (Chen et al., 2003). Our best approach is to understand processes and improve the various parameterisations used in the models. Here too, there appears to be steady progress.

7.4 On the Observing System Experiments

Observing system experiments (OSEs), or Observing system simulation experiments (OSSEs) provide a mechanism for obtaining guidance on the potential importance of an observing system (or even on specific measurements or measurement sites). The strategy consists of performing analyses with and without the observing system. In the case of OSSEs the observing system does not exist and observations must be simulated. This could be because the system is still in planning or has been withdrawn or reduced as in the case of TAO/TRITON (TT). We will not consider OSSEs further.

The first check is to compare analyses with and without the observation system, to gauge the impact on the analyses. Examples of this were given in Figure 3.3 and in TPOS 2020 WP5. In Figure 3.3, it was shown that the analyses without TT were noticeably less accurate at mooring sites when the TT data are not used, especially in the west Pacific.

A second check is to compare the forecasts from analyses made with and without the observing system. Because the atmosphere is chaotic and the models used to make forecasts are imperfect, any forecast system must consist of an extensive set of ensembles. It is thus not possible to verify the impact on a single forecast. In fact to obtain reliable statistics it is necessary to run the forecasts over many realizations spanning many years. This is very costly and so only limited sets of forecasts are feasible. Examples of OSEs are given in Figure 3.2 and also in TPOS 2020 WP5.

One could argue that to identify the effect of TAO/TRITON on say the prediction of Nino3.4 SST, one might need an ensemble of at least 3 members, ideally run every month for 10 years (the bigger the effect, the smaller the set needed to establish statistical reliability). More chaotic variables such as 2m temperature (T2m) in the extratropics, or precipitation, need considerably more ensemble members, since the spread of the PDF (Probability Distribution Function) of the target variable is larger and the shift in the mean of the PDF likely to be small compared with its width (Kumar and Hoerling, 1995). A greater effect might be expected in the tails of the distribution but the models are likely to be less reliable here.

Evaluating the significance of OSEs is difficult as the results are likely to be model dependent. For example fig 3.2 , obtained using ECMWF S3 shows an important role for TT based on the 2001-6 period (this period was chosen as the study aimed to compare the relative importance of TT and ARGO and 2001-6 was the then period of ARGO coverage). Later results using ECMWF S4 do not confirm this result. The reason has not yet been investigated but could be because the analysis system involving atmospheric analyses and reanalyses as well as ocean reanalyses has improved (or deteriorated) or because the longer period suggests the earlier results were not as statistically reliable as thought.

There is a further complication; the impact on the analysis is complex and never investigated in its entirety. For example, TT provides not just ocean observations but also surface meteorology (vector winds and humidity) which can influence the atmospheric analyses (see section 2.6). In the OSEs discussed here and in TPOS 2020 WP5, the effect of TT winds on reanalyses is not considered. Further, because of flaws in the atmospheric and oceanic models, systematic errors are present. Attempts to deal with these are usually included in the analysis system through a bias correction. However, this bias is based on an operational system in which the bias will have been evaluated over a substantial period of time, maybe 20 years or more, and so the bias estimate will have knowledge, albeit implicit, of the observing system whose importance is being evaluated. Without the observing system, the bias estimate would likely be less accurate. Since bias is perceived to be important (see Balmaseda et al., 2007) in improving the quality of the analyses and forecasts, and the bias would be less accurate if one withdrew the observing system while evaluating the bias, it is likely that the importance of the observing system being tested is underestimated. On the other hand the bias may be caused by the observing system, resulting from say a mismatch between the wind forcing and the observed thermocline depth (for example, see Bell et al., 2004; and Balmaseda et al., 2007). For these various reasons, blind interpretation of OSEs is not encouraged. They should be carried out using several models to see if a consistent picture emerges. But this may not be sufficient. Analysis systems are flawed (analysis is a very complex process). One example of this is shown in Figure 3.3 lower panel, when the fit to the verifying data actually improves when a data stream, in this case ARGO, is removed. It could be that ARGO and TT are inconsistent, but a more likely interpretation is that the data assimilation system is not using the data in an optimum way. One should note, however, that correcting errors like this is unlikely to be straightforward. Progress will undoubtedly be made, but on a scale of decades rather than years.

In summary, OSEs are a useful and essential tool for assessing the importance of an observation system and their use, particularly multi-model approach, should be encouraged, but they must be interpreted with caution.

8. Data Requirements and Recommendations

There are different aspects of a forecast system (initialization, system development, verification and calibration) that require observational information. Methodology for objective evaluation of the observing system only exists for the initialization, in terms of so-called “forecast sensitivity” evaluation, or from more ad-hoc Observing System Experiments (OSES). Even then, it is not always easy to extract useful information. In NWP there is routine forecast evaluation, but the metrics are not targeted to the boundary layer or surface variables. At seasonal and monthly forecasts range, the size of sample to be evaluated is smaller, and it is harder to establish statistical significance. Some ad-hoc OSES are conducted to evaluate in the initialization and forecast for seasonal time scales. But results are generally not very conclusive; vary a lot from different systems; and appear to be influenced by model error, rendering the evaluation quite difficult. To our knowledge, there has not been any evaluation of the observing system with forecasting systems at monthly time scales. In any case, this sort of evaluation targets only the impact of observations on the initialization of forecasts, but it can not measure the impact of the observations on the other components of a forecasting system (such as verification or model development).

TPOS 2020 WP5 offers a thorough list of observing system requirements from the ocean initialization perspective. Specific requirements for NWP are also captured by the relevant WMO expert groups. Rather than duplicating these documents, in what follows we provide a discussion of the relevant attributes of any observing system, including data delivery, accuracy, stability, resilience, temporal sampling and span, spatial sampling and coverage, including focus areas, and variables for the ocean, for waves, and for atmosphere boundary layer.

8.1. Data Streams

Regarding the delivery of data, operational centers would need at least 2 streams of data delivery: real time stream via GTS, to be used for the initialization of forecasts, and a BRT quality-controlled data stream. The latter is used for a variety of applications, which allow for longer delay times. Close-to-real-time reanalyses used for monitoring and the backbone of a seasonal forecasting system can accept delays of around 10-15 days to receive better processed data (such as better orbit specification for altimeter sea level or bias-corrected Argo salinity data). Verification of current forecasts need data (either station or L2) arriving with a delay or about 1-3 months. Delays of years or decades are accepted for re-processed data, or observations recovered from data mining.

8.2 Accuracy

It is difficult to provide hard set values for accuracy and precision of the observations. Indeed, these will depend very much on the application, and it is likely to vary in time as the models get better. Given that model error is still large, the requirements for verification will be more relaxed than for assimilation. Requirements for model development may vary, and in some cases very precise observations may be needed. From the assimilation perspective, it is important that the observations are unbiased. It is better to have larger random errors in the observations than smaller systematic biases. It is also important that the errors are stable in time, and homogeneous in space. If this is not possible, models to parameterize these errors should be provided. Currently, the instrument error of most in-situ observations is small compared with the representativeness error (which is determined by the model unresolved spatial and temporal scale). One possibility is to require that the observation random error should not exceed a percentage of the variability that the forecast system is trying to predict (say 1-5%).

8.3 Stability

An important requirement for the observing system is continuity and stability in time. Although changes leading to improvement are of course welcome, exploring new ways of obtaining efficient measurements should not compromise more traditional but critical observations. Well established observing systems should only be abandoned when there is clear evidence that they are redundant, or that the needs they were serving can be catered by alternative means. One should bear in mind that for calibration of seasonal forecasts requires stable data records of about 30 years. The longer the record, the better the calibration of extreme events is performed. For decadal forecasting and climate monitoring, records of 50-100 years are required.

8.4 Resilience

The observing system needs to be robust and resilient to guarantee that the needs of operational services are met. This implies some degree of redundancy, to cover for failure of individual components. It may also be advisable to formulate a priority list for sites and observations that need to be maintained in critical situations.

8.5 Spatial sampling, coverage and focus areas

The equatorial region appears challenging for both atmospheric and ocean models. The question is whether regions where forecasting systems do not make good use of observations should be more, or on the contrary less, sampled. The Equator is a key region for the earth's climate and for monthly to seasonal forecasts. From that point of view it should be sufficiently sampled. Ocean models also have problems with western boundary currents, which are also key for the ocean circulation and air-sea interaction.

The robust broad scale sampling of in-situ temperature and salinity achieved by Argo should be maintained, and if possible enhanced. The sampling of the deep ocean below 2000m is needed for advancing decadal forecasts, reanalyses of the global ocean, and ocean model development.

Enhanced sampling of the ocean and atmosphere boundary layer should be considered, both for model development and for coupled data assimilation. The enhancement includes higher vertical and temporal resolution, as well as the existence of reference flux sites sampling different regimes. More specific studies are needed to establish the needs for a sustained observing system of the ocean-atmosphere boundary layer.

The high resolution spatial sampling of surface variables achieved by satellite instruments (SST, SSH, SSS, ocean color, surface winds) is needed, and it is likely to have higher impact as the model resolution increases.

8.5 Temporal sampling

The temporal sampling is not independent of the spatial sampling. Currently global coverage of the ocean at 3 x 3 degrees is achieved by Argo every 10 days, but this should be increased if possible. Equally, daily global coverage at, say, 2 x 2 degree of SSH by satellite altimeter would be desirable. Satellite altimetry does not yet resolve the diurnal cycle. The temporal sampling by the moored array (daily and sub-daily) is very valuable.

General guidelines for temporal sampling would be: 1-3 hours sampling for ocean-atmosphere boundary layer process. Daily sampling for in-situ subsurface (upper 1000m) at a spatial resolution of 200km-1000km (depending on the region). Below 2000m, monthly samplings may be adequate, but more specific studies are needed.

8.6 Essential variables

The following variables are considered necessary (see also WP5)

- Ocean subsurface [T(z),S(z),U(z)]
- Ocean surface [SST, SSS, Surface Currents], as well as SSH, Geoid, and Bottom pressure
- Atmosphere Surface and BL: surface winds, T2m, humidity, SLP, and precipitation
- Waves: wave height, wave period, wave spectrum
- Multivariate (collocated at same location), direct or derived: heat flux, humidity flux, wind stress, TKE profile, ocean-atmosphere boundary layer high resolution soundings.

References

Anderson et al. (1998): Surface buoyancy forcing and the mixed layer of the Western Pacific warm pool: Observations and 1D model results, J. Clim., 9, pp. 3056–3085.

Andersson, E., and Yoshiaki, S. (Editors) (2012): Final Report of the Fifth WMO Workshop on the Impact of Various Observing Systems on Numerical Weather Prediction. WIGOS WMO Integrated Global Observing System Technical Report 2012-1.

Balmaseda, M.A., Dee, D., Vidard, A, and Anderson, D.L.T. (2007a): A multivariate treatment of bias for sequential data assimilation: application to the tropical oceans. Q. J. R. Meteorol. Soc. 133: pp. 167–179.

Balmaseda, M., and Anderson, D. (2009): Impact of initialization strategies and observations on seasonal forecast skill, Geophys. Res. Lett., 36, L01701, (doi:10.1029/2008GL035561).

Balmaseda, M. & Co-Authors (2010): "Role of the Ocean Observing System in an End-to-End Seasonal Forecasting System" in Proceedings of OceanObs'09: Sustained Ocean Observations and Information for Society (Vol. 1), Venice, Italy, 21-25 September 2009, Hall, J., Harrison, D.E. & Stammer, D., Eds., ESA Publication WPP-306, doi:10.5270/OceanObs09.pp.03.

Balmaseda, M. A., Trenberth, K.E., and Källén, E. (2013): Distinctive climate signals in reanalysis of global ocean heat content, Geophys. Res. Lett., 40, pp. 1754–1759, (doi:10.1002/grl.50382).

Bell, M. J., Martin, M. J., and Nichols, N. K. (2004): Assimilation of data into an ocean model with systematic errors near the equator. Q. J. R. Meteorol. Soc., 130, pp. 873-893.

Bonjean, F., and Lagerloef, G.S.E. (2002): Diagnostic Model and Analysis of the Surface Currents in the Tropical Pacific Ocean. J. Phys. Oceanogr., 32, pp. 2938–2954.

Browning, G. L., Kreiss, H.O., and van de Kamp, D.W. (2000): "Comments on "Observations of a Mesoscale Ducted Gravity Wave"." *Journal of the atmospheric sciences* 57.4 (2000): pp. 595-598.

Burgers, G., Balmaseda, M.A., Vossepoel, F.C., van Oldenborgh, G.J., and Leeuwen, P.J. (2002): Balanced ocean data assimilation near the equator. J Phys Ocean, 32, pp. 2509-2519.

Cassou (2008): Intraseasonal interaction between the Madden–Julian Oscillation and the North Atlantic Oscillation, Nature, 455, pp. 523-527.

Chepurin, G. A., Carton, J.A., and Leuliette, E. (2014): Sea level in ocean reanalyses and tide gauges, J. Geophys. Res. Oceans, 119, (doi:10.1002/2013JC009365).

Craig, P.D. and Banner, M.L. (1994): Modeling wave-enhanced turbulence in the ocean surface layer. J. Phys. Oceanogr. 24, pp. 2546-2559.

De Boisséson, E., Balmaseda, M. A., Vitart, F., and Mogensen, K. (2012): Impact of the sea surface temperature forcing on hindcasts of Madden-Julian Oscillation events using the ECMWF model, Ocean Sci. Discuss., 9, pp. 2535-2559, (doi:10.5194/osd-9-2535-2012).

Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, MA, Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A.C.M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A.J., Haimberger, L., Healy, S.B., Hersbach, H., Hólm, E.V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A.P., Monge-Sanz, B.M., Morcrette, J.J., Park, B.K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.N., and Vitart, F. (2011): The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. Q. J. R. Meteorol. Soc. 137: pp. 553–597.

England, M.H. (2014): "Recent intensification of wind-driven circulation in the Pacific and the ongoing warming hiatus." Nature Climate Change 4.3 (2014): pp. 222-227.

Fujii, Y., Nakaegawa, T., Matsumoto, S., Yasuda, T., Yamanaka, G., and M. Kamachi (2009): Coupled climate simulation by constraining ocean fields in a coupled model with ocean data. J. Climate, 22(20), pp. 5541-5557, (doi:10.1002/jgrc.20094).

Garrett, C. (1976): The generation of Langmuir circulations by surface waves- A feedback mechanism. J. Mar. Res., 34, pp. 117-130.

Grant, A.L.M. and Belcher, S.E. (2009): Characteristics of Langmuir Turbulence in the Ocean Mixed layer, J. Phys. Oceanogr. 39, pp. 1871-1887.

Grunseich et al. 2013: The Madden-Julian oscillation detected in Aquarius salinity observations, Geophys. Res. Lett., 40, 20, 5461–5466

Hasselmann, K., 1970. Wave-driven inertial oscillations. Geophys. Fluid Dyn. 1, pp. 463-502.

Hawkins E, and Sutton R. (2009): The potential to narrow uncertainty in regional climate predictions. Bull Am Meteor Soc 90: pp. 1095–1107.

Janssen, P.A.E.M. (2012): Ocean wave effects on the daily cycle in SST. J. Geophys. Res, 117, C00J32, 24 pp., (doi:10.1029/2012JC007943).

Janssen, P.A.E.M., Breivik, O., Mogensen, K., Vitart, F., Balmaseda, M.A., Bidlot, J.R., Keeley, S., Leutbecher, M., Magnusson, L., and Molteni, F. (2013): Air-sea interaction and surface waves. ECMWF Technical Memorandum 712. 34 pages. Kelly et al. 2010: Western Boundary Currents and Frontal Air–Sea Interaction: Gulf Stream and Kuroshio Extension. J. Clim., 23, pp. 5644–5667.

Klingaman, and Woolnough (2013): The role of air–sea coupling in the simulation of the Madden–Julian oscillation in the Hadley Centre model, Quart. J. Roy. Meteorol. Soc. accepted.

Kumar, A., and Hoerling, P.M. (1995): Prospects and Limitations of Seasonal Atmospheric GCM Predictions. Bull. Ame. Meteor. Soc., 76, pp. 335-345.

Kumar, A., Chen, M., Zhang, L., Wang, W., Xue, Y., Wen, C., Marx, L., and Huang, B. (2012): An analysis of the non-stationarity in the bias of sea surface temperature forecasts for the NCEP climate forecast system (CFS), version 2. Mon Wea Rev 140:3003–3016, (doi:10.1175/MWR-D-11-00335.1).

Kumar, A., Wang, H., Xue, Y., and Wang, W. (2014): How much of monthly subsurface temperature variability in equatorial pacific can be recovered by the specification of sea surface temperatures? J. Climate, 27, pp. 1559-1557.

Luo, J.-J., et al. (2005): Seasonal Climate Predictability in a Coupled OAGCM Using a Different Approach for Ensemble Forecasts. Journal of climate 18.

Lyman, J. M., Willis, J.K., and Johnson, G.C. (2006): Recent cooling of the upper ocean, Geophys. Res. Lett., 33, L18604, (doi: 10.1029/2006GL027033).

Magnusson, L., Balmaseda, M.A., and Molteni, F. (2012b): On the dependence of ENSO simulation on the coupled model mean state. Clim Dyn., (doi: 10.1007/s00382-012-1574-y).

Magnusson, L., Alonso-Balmaseda, M., Corti, S., Molteni, F., and Stockdale, T. (2012): Evaluation of forecast strategies for seasonal and decadal forecasts in presence of systematic model errors. *Climate Dynamics*, pp. 1-17.

Matthew et al. (2010): Ocean temperature and salinity components of the Madden–Julian oscillation observed by Argo floats, Clim. Dyn., 35, 7-8, pp. 1149-1168.

McWilliams, J.C., and Restrepo, J.M. (1999): The wave-driven ocean circulation. J. Phys. Oceanogr. 29, pp. 2523-2540.

Molteni, F., Stockdale, T., Balmaseda, M.A., Balsamo, G., Buizza, R., Ferranti, L., Magnusson, L., Mogensen, K., Palmer, T.N., and Vitart, F. (2011): 'The new ECMWF seasonal forecast system (System 4)'. Tech. Memo. 656. ECMWF: Reading, UK.

Mori, and Watanabe (2008): The growth and triggering mechanisms of the PNA: a MJO-PNA coherence. J. Meteorol. Soc. Jpn. 86, pp. 213–236.

Palmer, T. N., and Anderson, D. L. T. (1994): The prospects for seasonal forecasting-A review paper. *Quartly. J. Roy. Met. Soc. 120,* pp. 755-793.

Saha, S., Moorthi, S., Pan, H.L., Wu, X., Wang, J., Nadiga, S., Tripp, P., Kistler, R., Woollen, J., Behringer, D., Liu, H., Stokes, D., Grumbine, R., Gayno, G., Wang, J., Hou, Y.T., Chuang, H.Y., Juang, H.M.H., Sela, J., Iredell, M., Treadon, R., Kleist, D., Van Delst, P., Keyser, D., Derber, J., Ek, M., Meng, J., Wei, H., Yang, R., Lord, S., Van Den Dool, H., Kumar, A., Wang, W., Long, C., Chelliah, M., Xue, Y., Huang, B., Schemm, J.K., Ebisuzaki, W., Lin, R., Xie, P., Chen, M., Zhou, S., Higgins, W., Zou, C.Z., Liu, Q., Chen, Y., Han, Y., Cucurull, L., Reynolds, R.W,. Rutledge, G., Goldberg, M. (2010): The NCEP climate forecast system reanalysis. Bull. Am. Meteorol. Soc. 91: pp. 1015–1057.

Ricci, S., Weaver, A.T., Vialard, J., and Rogel, P. (2005): Incorporating temperature-salinity constraints in the background-error covariance of variational ocean data assimilation. Mon. Wea. Rev. 133: pp. 317–338.

Small et al. (2008): Air-sea interaction over ocean fronts and eddies. Dyn. Atmos.-Ocean, 45, pp. 274-319.

Smith, D. M., and Murphy, J.M. (2007): An objective ocean temperature and salinity analysis using covariances from a global climate model, J. Geophys. Res. 112, C02022.

Stockdale T. (1997): Coupled ocean-atmosphere forecast in the presence of climate drift. Mon Wea Rev 125: pp. 809–818.

Terray, E.A., Donelan, M.A., Agrawal, Y.C., Drennan, W.M., Kahma, K.K., Williams, A.J., P.A. Hwang, P.A., Kitaigorodskii, S.A. (1996): Estimates of Kinetic Energy Dissipation under Breaking Waves. J. Phys. Oceanogr. 26, pp. 792-807.

Troccoli, A., and Haines K. (1999): Use of Temperature–Salinity relation in a data assimilation context. J. Atmos. Oceanic Technol. 16: pp. 2011–2025.

Vitart (2013): Evolution of ECMWF sub-seasonal forecast skill scores. Quart. J. Roy. Meteorol. Soc., accepted.

Wheeler, M., and Kiladis, G.N. (1999): Convectively Coupled Equatorial Waves: Analysis of Clouds and Temperature in the Wavenumber–Frequency Domain. *J. Atmos. Sci.*, **56**, 374–399.

Wijffels, S., Willis, J., Domingues, C.M., Barker, P., White, N.J., Gronell, A., Ridgway, K., Church, J.A. (2009): Changing expendable bathythermograph fall rates and their impact on estimates of thermosteric sea level rise. J. Climate 21: pp. 5657–5672.

Woolnough et al. (2007): The role of the ocean in the Madden-Julian Oscillation: Implications for MJO prediction. Quart. J. Roy. Meteorol. Soc., 133 (622), pp. 117-128.

Yin, Y., Alves, O., and Oke, P.R. (2011): An ensemble ocean data assimilation system for seasonal prediction. Mon. Wea. Rev.139, pp. 786-808.

Žagar, N., Andersson, E., and Fisher, M. (2005): Balanced tropical data assimilation based on study of equatorial waves in ECMWF short-range forecast errors. Q. J. R. Meteorol. Soc. 131, pp. 987-1011.

Zhang, L., Kumar, A., and Wang, W. (2012): Influence of changes in observations on precipitation: A case study for the Climate Forecast System Reanalysis (CFSR), J. Geophys. Res.117, D08105.

Zebiak, S.E., and Cane, M.A. (1987): "A Model El Niño–Southern Oscillation". Mon. Wea. Rev. 115 (10): pp. 2262–2278.

Appendix A

A.1 Naming convention used in this paper

To facilitate the discussion of the data needs it is helpful to introduce a naming convention regarding the processing level of the observations, the quality control, and the time-delivery properties.

For the processing level we follow the remote sensing convention, which classifies the data in different levels (L0 to L4) as described in the Appendix A.2. This will also be used for in-situ data. L0 data is not usually required in the applications mentioned in this paper. For operational forecasting we will be talking about level 1 (L1) to level 4 (L4). Most of the in-situ data used in operational forecasting are L2 (intended geophysical variables (IGV) at measurement location and time. When the data have been gridded (such as some Argo gridded products), binned in time, and sometimes combined with other data sources, it becomes L3. These can have regions or periods of time with missing data. If a further degree of processing takes place, the data can become L4. This is the case of analyses conducted with models via data assimilation, or when derived geophysical variables (DGV) are produced by some sort of geophysical constraint (such as geostrophic transports, or Argo derived velocities, or surface currents from altimeter). Some derived variables such as section transports or surface fluxes derived from bulk formulae are available only at precise locations, and no gridding algorithm has taken place. We will refer to these products as L3-D (D for Derived), to distinguish them from the more direct L2-QC and gridded L4 products.

Table A1 - Naming convention for data streams according to the delivery time and processing level. IGV is for Intended Geophysical Variable and DGV is for Derived Geophysical Variable. See text for other acronyms.

Product Level	Processing	Use	Examples
L2	IGV No Gridded	Real Time Initialization (Analyses)	Data acquired via GTS
L2-QC	IGV-QC No Gridded Time series possible	Reanalyses. Verification. Model development	WOA/EN3 data set.
L3	IGV Binned. Areas with missing data permitted. Time series possible	Analyses and Reanalyses Verification. Model development	Gridded sea-level maps from altimeter. Ocean Color climatology.
L3-D	DGV Binned. Areas with missing data permitted. Time series possible	Analyses and Reanalyses. Verification. Model development	Transports. Fluxes from bulk-formulae at observation points
L4	IGV/DGV Gridded.	Analyses and Reanalyses.	Surface fluxes from Reanalyses.

	Time series possible	Verification. Model Development	Surface currents from reanalyses

For many operational applications L2 data is expected to be delivered in near real-time (traditionally within 3 hours in meteorology and say within 24 hours for ocean applications) following a protocol that allows its automatic acquisition and usage, often under the auspices of the WMO. We will refer to this data stream as NRT. The L2 data can have some degree of quality controlled (QC), such as bias correction, black-listing, and others. We will refer to this as L2-QC. The QC often takes place in specialized data processing centers, and this may result in a delay in the delivery, even when the QC is done on a routine/operational basis. We can refer to this data stream as BRT, since it arrives behind real-time, but it can still be used in operational delayed products such as reanalyses. The delay time application of observed data depends very much of the application, and will be specified whenever is needed. When L2-QC data from a fixed location are further treated to provide time-series with the data binned at adequate time-intervals (daily, monthly), the product becomes L3 or L3-timeseries. Note that the L3 data can still have regions or period where the data is missing. Finally, there can be very sophisticated QC and processing procedures (especially some of L4 data), without specific delivery time requirements.

A.2 Levels of processing of remote sensing data

In remote sensing the data is classified in different levels (Level 0 to Level 4) according to the degree of processing:.

- L0: Reconstructed, unprocessed instrument and payload data at full resolution, with any and all communications artifacts (e. g., synchronization frames, communications headers, and duplicate data) removed.

- L1a: Reconstructed, unprocessed instrument data at full resolution, time-referenced, and annotated with ancillary information, including radiometric and geometric calibration coefficients and georeferencing parameters (e.g., platform ephemeris) computed and appended but not applied to the L0 data (or if applied, in a manner that L0 is fully recoverable from L1a data).
- L1b: These are L1a data that have been processed to sensor units (e. g., radar backscatter cross section, brightness temperature, etc.); not all instruments have L1b data; L0 data is not recoverable from L1b data.

- L2: Derived geophysical variables (e. g., ocean wave height, soil moisture, ice concentration) at the same resolution and location as L1 source data.

- L3: Variables mapped on uniform space-time grid scales, usually with some completeness and consistency (e. g., missing points interpolated, complete regions mosaicked together from multiple orbits, etc.).

Supplementary Material

Illustration of Equatorial Dynamics and Time Scales

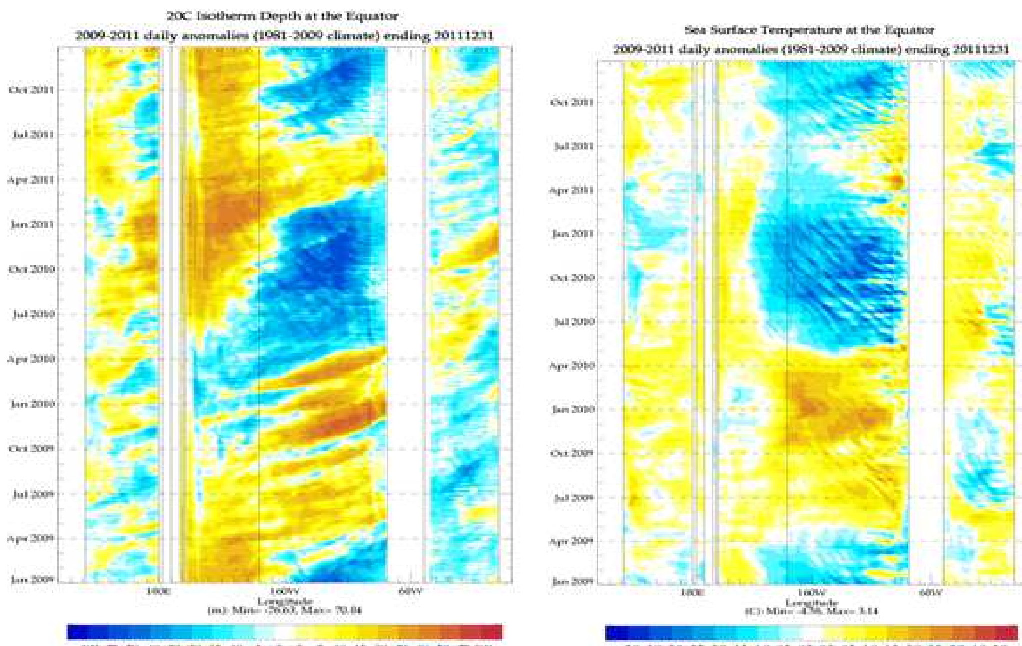


Figure S1 - Longitude-time diagrams of equatorial thermocline depth (left) and SST (right) anomalies. The thermocline depth is represented by the depth of the 20 degree Isotherm (D20). The anomalies are computed respect the 1989-2008 climatology. The eastward propagation of equatorial Kelvin waves is visible in D20 usually preceding the appearance of SST anomalies in the Eastern Pacific (from the ORAS4 ocean reanalysis (Balmaseda et al., 2013)). The figure also shows that a thermocline anomaly associated with an individual Kelvin wave does not always translate into a large scale SST anomaly, as it happened in the "failed" El Niño of 2011, when in spite of a substantial propagation of the thermocline anomalies the warm SST anomaly was very short lived, and the El Niño did not materialized. The figure also shows the variability associate with the westward propagating Tropical Instability Waves in Eastern Pacific.

Ocean Observing System

Supplementary Figure S2 shows schematically the different components of the ocean observing system and their availability in time. SST observations are essential for seasonal forecasts. Most of the initialization systems also use subsurface temperature from XBT's (Expendable bathythermograph), CTDs (Conductivity, Temperature and Depth) usually from scientific cruises, moored buoys (TAO/TRITON in the Pacific, PIRATA in the Atlantic, RAMA in the Indian Ocean) and Argo floats. Salinity (mainly from Argo and CTDs), and altimeter-derived sea-level anomalies (SLAs, since approximately 1993) are also assimilated. The latter usually need a prescribed external Mean Dynamic Topography (MDT), which can be derived indirectly from gravity missions such as GRACE (Gravity Recovery and Climate Experiment) and, in the near future, GOCE (Gravity field and steady-state Ocean Circulation Explorer). Supplementary figure S3 shows the time evolution of the number of temperature and salinity observations, as well as the typical spatial distribution. The figure shows the large increase in observations associated to the advent of Argo. The properties of spatial and temporal sampling varies substantially between instruments: the XBTs usually follow commercial ship routes, CTDs are associated with intense scientific missions, the moored array samples the equatorial oceans at few selected fix positions; Argo, is only observing system that sample uniformly the subsurface of the ocean, measuring temperature and salinity up to depth of 2000m. Altimeter sea-level (not shown) also samples the surface of the ocean quite uniformly, but a good relation between sea level variations and subsurface structure is only possible in regions of strong stratification (the tropics).

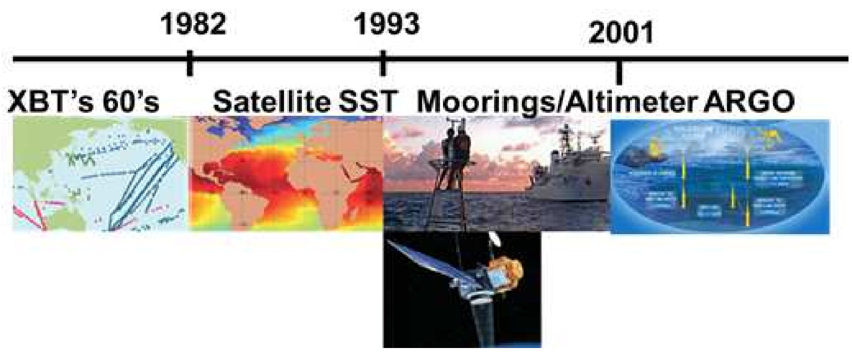


Figure S2 - Time evolution of the ocean observing system by instrument.

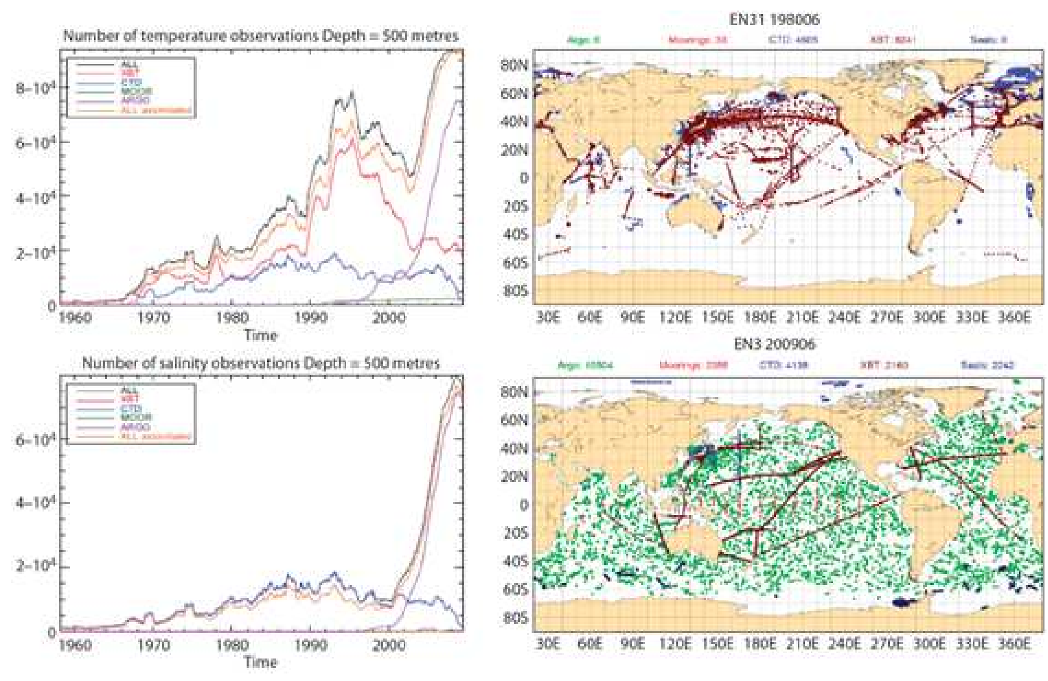


Figure S3 - (left) Number of temperature (top) and salinity (bottom) observations within the depth range 400m-600m as a function of time per instrument type. The black curve is the total number of observations. The orange curve shows the number of assimilated observations. (right) Typical observation coverage in June 1980 (top) and in June 2005 (bottom). Note that the color coding for the instruments is not the same in the left and right panels.