Origin and impact of initialisation shocks in coupled atmosphere-ocean forecasts

Article

Accepted Version

Origin and impact of initialisation shocks in coupled atmosphere-ocean forecasts

David P. Mulholland*

Department of Meteorology, University of Reading, Reading, UK

Patrick Laloyaux

European Centre for Medium-range Weather Forecasts, Shinfield Park, Reading, UK

Keith Haines

Department of Meteorology and National Centre for Earth Observation, University of Reading, Reading, UK

Magdalena Alonso Balmaseda

European Centre for Medium-range Weather Forecasts, Shinfield Park, Reading, UK

* *Corresponding author address:* David Mulholland, Department of Meteorology, University of Reading, Earley Gate, PO Box 243, Reading RG6 6BB, UK

E-mail: d.p.mulholland@reading.ac.uk

ABSTRACT

Current methods for initialising coupled atmosphere-ocean forecasts often rely on the use of separate atmosphere and ocean analyses, the combination of which can leave the coupled system imbalanced at the beginning of the forecast, potentially accelerating the development of errors. Using a series of experiments with the European Centre for Medium-range Weather Forecasts coupled system, the magnitude and extent of these so-called initialisation shocks is quantified, and their impact on forecast skill measured. It is found that forecasts initialised by separate ocean and atmospheric analyses do exhibit initialisation shocks in lower atmospheric temperature, when compared to forecasts initialised using a coupled data assimilation method. These shocks result in as much as a doubling of root-mean-square error on the first day of the forecast in some regions, and in increases that are sustained for the duration of the 10-day forecasts performed here. However, the impacts of this choice of initialisation on forecast skill, assessed using independent datasets, were found to be negligible, at least over the limited period studied. Larger initialisation shocks are found to follow a change in either the atmospheric or ocean model component between the analysis and forecast phases: changes in the ocean component can lead to sea surface temperature shocks of more than 0.5 K in some equatorial regions during the first day of the forecast. Implications for the development of coupled forecast systems, particularly with respect to coupled data assimilation methods, are discussed.

1. Introduction

The use of a coupled atmosphere-ocean model, in preference to an atmosphere-only modelling approach, is essential in order to achieve skillful forecasts of climate on the seasonal timescale and beyond, and is increasingly being recognised to provide benefits at shorter forecast lead times too (e.g., Fu et al. 2007; Klingaman et al. 2008; Vitart et al. 2008; Janssen et al. 2013; Shelly et al. 2014). A major challenge of the coupled forecasting approach lies in the initialisation, the goal of which is to incorporate information from the observational network in both atmosphere and ocean into the corresponding model components in an optimal manner. This is commonly achieved through data assimilation (DA), performed using one of a number of established methods for each model component (e.g., Daley 1991; Anderson et al. 1996).

The data assimilation strategy used by operational centres in recent years to initialise coupled forecasts (e.g., Saha et al. 2006; Molteni et al. 2011; Arribas et al. 2011; MacLachlan et al. 2014) is to perform separate analyses of the atmosphere and ocean. A sea surface temperature (SST) product is used to prescribe the boundary condition of the atmospheric model, and the ocean model is constrained by either near-surface atmospheric fields or explicitly specified surface heat, momentum and freshwater fluxes, typically obtained from an atmospheric analysis or from a gridded observational product. One-directional coupling during the initialisation may be achieved with this approach, by using the result of the atmospheric analysis to provide the boundary condition for the ocean model (e.g., Balmaseda et al. 2013). However, the use of different models for the analysis and forecast phases can further complicate matters, particulary when producing historical hindcasts (re-forecasts) for calibration purposes using past initial conditions computed with previous model code versions. In this context, obtaining truly balanced initial conditions requires allowing for some degree of atmosphere-ocean coupling to occur during the analyses themselves, as well as the use of the same coupled model in the analysis and forecast phases.

Various possible coupled data assimilation systems exist, exhibiting varying strengths of coupling between the atmosphere and ocean. Several operational centres are pursuing such methods (Saha et al. 2010; Lea et al. 2014; Alves et al. 2014), including the European Centre for Medium-range Weather Forecasts (ECMWF) which has developed a prototype for a coupled assimilation system that ingests simultaneously atmospheric and ocean observations (Laloyaux et al. 2015). In this system, information is allowed to cross the interface through the multiple integrations of the coupled model performed during the assimilation process, ensuring a consistent atmosphere-ocean analysis is produced (in the sense that each of the two model components have knowledge about the boundary fluxes of the other component, and have been able to establish a balance with one another in this context). Forecasts can be initialised from the output of this coupled analysis. ECMWF operational coupled forecasts currently, however, continue to use the uncoupled analysis method for initialisation.

In choosing an initialisation method, particularly for relatively short-range coupled forecasts, it is important to ensure that the two model components are consistent with one another at the commencement of the forecast, in order to avoid the generation of 'initialisation shocks' (alternatively, coupling shocks, or spin-up effects) (Rahmstorf 1995; Zhang et al. 2007; Balmaseda et al. 2009; Zhang 2011). The likely existence of initialisation shocks in the coupled model context has been acknowledged, particularly in a seasonal forecasting context (Balmaseda and Anderson 2009; Marshall et al. 2011), but neither their formation nor impact in short-range forecasts using a full atmosphere-ocean global climate model has been explored in detail, to our knowledge. A particular problem lies in separating out signals of initialisation shock — that is, those that result purely from an imperfect initialisation method — and those of model drift, which occurs regard-less of the initialisation method used, due to the existence of biases, physical or dynamical, in the model (e.g., Magnusson et al. 2013; Wang et al. 2014). Measuring the magnitude of initialisation shock and investigating its causes are important steps in maximising the effectiveness of coupled forecasts and in pointing the way towards possible improvements to conventional methods.

Here, we define initialisation shock relatively broadly, to encompass several possible causes, each of which we are able to isolate using the experiments that follow:

1. An imbalance, in the vertical fluxes of any of heat, momentum or freshwater, between the atmosphere and ocean initial states, formed due to insufficient communication between the two model components during the calculation of the initial conditions. This situation can arise if model components are coupled to forcing fields other than those of the coupled system during initialisation, such that the near-surface regions of each component are compatible with the relevant forcing fields but will not, in general, be compatible with each other. As a result, when the two components are combined at the beginning of the forecast, rapid changes in surface fluxes are expected, as the two components exchange heat, momentum and/or freshwater in order to establish a new thermodynamical balance. This rapid adjustment could have an undesirable impact on the forecast.

2. The use of different models, or different versions or configurations of the same model, to provide the initial state (for either component) and to compute the forecast. A common example of this is the use of a popular reanalysis such as ERA-Interim (Dee et al. 2011) to directly initialise an atmospheric model different to the one used to generate the reanalysis (the reanalysis may then be described as ‘non-native’ with respect to the forecast model). The result could be an initial state that is incompatible with new model’s attractor, resulting in an adjustment at the beginning of the forecast.

3. The instantaneous removal of bias correction terms in one of the model components, resulting in an abrupt change in the dynamics of the component at the beginning of the forecast, even in the absence of any model drift (this effect is explained in more detail in Section 3d).

This initialisation shock definition is not intended to be a complete list of the contributors to spin-up effects in a model forecast: development of forecast errors due to model biases, in what would be considered ‘standard’ model drift, is *not* included, since this process is unavoidable even with a balanced initialisation using the same models as the forecast itself. Further, model adjustments occurring as a result of the more general problem of assimilating observational information in the initial conditions but not in the forecast itself, are not explicitly considered, as these are also present in all of the forecast systems used in this work. The shocks that are discussed here are those deviations of the forecast from the truth that can demonstrably be reduced or eliminated through changes to the initialisation procedure. Also, we note that a similar initialisation problem exists for the coupling of atmosphere and land surface model components, but do not consider this here: we focus solely on atmosphere-ocean coupling.

In this paper, we use the ECMWF analysis and forecast system, in various configurations, to detect the occurrence of initialisation shocks in coupled forecasts; to quantify the contributions to these shocks of each of the mechanisms listed above; and to evaluate the impact of shock on coupled forecasts. By using forecasts initialised using coupled DA as a control, it is possible to isolate those deviations from a reference state that may be described as initialisation shocks, as a subset of the total model drift, which occurs also via the development of systematic model biases. We attempt to establish if effects can be reduced through changes to the initialisation method, and investigate the extent to which the presence of initialisation shocks might affect forecast skill.

The structure of the paper is as follows. The models and initialisation techniques used in the paper are introduced, and the experiments performed are defined, in Section 2. The results of these experiments, including identification of initialisation shocks and evaluation of forecast skill, are presented in Section 3. Implications for operational coupled forecasting are discussed in Section 4, and the key findings of the paper are summarised in Section 5.

2. Methods

a. Models and experiments

The coupled DA system recently developed at ECMWF, called the Coupled ECMWF ReAnalysis system (CERA), is presented and described in detail in Laloyaux et al. (2015). The CERA system is based on an incremental variational approach in which the misfits with ocean and atmospheric observations are computed by the ECMWF coupled model. Both atmospheric and subsurface ocean observations are assimilated within a common 24-hour assimilation window, leading to the computation of a coupled atmosphere-ocean analysis. The CERA system uses recent versions of the Integrated Forecast System (IFS), at a spectral resolution of T159 with 137 vertical levels, for the atmosphere, and the Nucleus for European Modelling of the Ocean (NEMO) model, in the ORCA1 configuration (corresponding to a horizontal resolution of around 1° in midlatitudes and 1/3° at the equator, with 42 vertical levels) for the ocean (see Table 1 for details of CERA and the other analyses used in this paper).

For the purposes of understanding this paper, additional important points to note regarding the CERA system are that SST is nudged towards a gridded observational product during the coupled model integrations, rather than being explicitly assimilated, and that bias correction (see Section 3d) is not used in the ocean. The initialisation method used in CERA is presented diagrammatically in Fig. 1, along with the other approaches to ocean-atmosphere data assimilation that are relevant to this paper. It is intended that the degree of coupling present in the CERA method is sufficient to ensure a consistent initial ocean-atmosphere state, and thus (along with a consistency of models between analysis and forecast) avoid initialisation shocks of the types listed in the previous section.

Using CERA, coupled reanalyses were performed covering three two-month test periods (to provide some coverage of the seasonal cycle): Apr–May 2008, Dec–Jan 2008/9 and Aug–Sep 2010. 10-day forecasts were initiated at 5-day intervals during these periods, at 00:00 UTC, using the CERA analysis to provide the initial conditions in both the atmosphere and the ocean. This set of 30 forecasts is named C1 (for 'Coupled'; see Table 2). These forecasts were run with the same model configuration (versions and resolutions) as used in CERA. While the three periods used cover a somewhat limited range (less than 3 years) of the possible background states of the climate system, the consistency of results (shown in the next section) across the three periods gives confidence that our forecast sets are adequate for determining the relative importance of each of the sources of shock.

Uncoupled analyses were also performed during these periods. The atmospheric analysis (which is referred to as U_atmos) used the observed SST products as the lower boundary condition, and this analysis was then used as the upper boundary condition during the ocean analysis (referred to as U_ocean), with heat, freshwater and momentum fluxes from U_atmos applied as daily averages (in the same manner as described in Balmaseda et al. (2013)). The same subsurface observations were assimilated, and the same SST nudging scheme was used, as in CERA. A set of forecasts, U1 (for 'Uncoupled'), with the same resolution as C1, was run using initial conditions obtained from these analyses. We refer to this set as 'uncoupled', though in fact a degree of one-directional coupling does exist in the initialisation, through the use of the completed atmospheric analysis during the ocean analysis. Note, also, that the name U1 refers to the uncoupled nature of the analyses only: all forecasts performed here use a coupled system. Comparison of U1 to C1 will reveal the impact on forecasts of the use of coupled DA in creating the initial conditions. With respect to the other experiments detailed subsequently, the key feature of U1 is the use of the same operational ocean and atmosphere models in analyses and forecasts.

A third set of forecasts, M1 (for ‘Model change’), was performed, using the same coupled forecast model versions as used by C1 and U1. In this set, atmosphere and ocean components were initialised using uncoupled reanalyses, namely ERA-Interim (Dee et al. 2011) for the atmosphere, and ORAS4 (Balmaseda et al. 2013) for the ocean. These reanalyses were performed with the atmospheric and ocean components of the ECMWF coupled forecasting system model, respectively (again using a gridded SST product as atmospheric boundary conditions and for ocean SST nudging), but in both cases older, deprecated model versions were used (see Table 1), creating an inconsistency between the analyses and forecasts. In the case of the atmosphere, the resolution between analysis and forecast also differed: ERA-Interim used a resolution of T255 L60, whereas the M1 forecasts were run at T159 L91. In the ocean, analysis and forecast resolutions were the same (ORCA1, 42 vertical levels, as previously). In M1, as in U1, there is some degree of coupling in the initialisation, as ORAS4 was forced by ERA-Interim fluxes during the assimilation.

This method, involving older model versions (and possibly lower resolutions) in the creation of initial conditions, is commonly used for the production of historical hindcasts that are needed for the calibration of operational seasonal forecasts (e.g., Arribas et al. 2011), and changes in model version from analysis to forecast may also be a feature of the operational seasonal forecasts themselves (Molteni et al. 2011).

Details of all the forecast types are summarised in Table 2. Note that in each case, the initial SST values used are taken from the ocean component of the analysis, rather than the atmospheric component (Fig. 1). In short, the comparison between U1 and C1 is designed to reveal the shock that occurs (in U1) due to atmosphere-ocean imbalance in the initial conditions, while the comparison between M1 and U1 is aimed at investigating the sensitivity of forecasts to the choice of uncoupled (re)analysis products used for initialisation, i.e. how this choice of initialisation product can generate shocks of the second and third 'types', as listed in the previous section. It is expected that any shocks will be detectable within the 10-day range of the forecasts.

Two further sets of forecasts are added later (see Section 3d, and Table 2), to distinguish between the second and third sources of shock. Additionally, several 7-month forecasts are performed (see Section 4), to briefly examine the potential for initialisation shocks to impact the forecast on monthly timescales.

b. Forecast evaluation methods

In the results that follow, two common metrics, root-mean-square error (RMSE) and anomaly correlation coefficient (ACC), are used to measure forecast bias and skill respectively. RMSE is sensitive to mean drift so is used to detect shocks and identify absolute-value differences between forecast types. The centred version of ACC, as used here, is insensitive to mean drift (forecast and reference anomalies are calculated with respect to their individual climatologies) so is used to measure forecast skill. For each forecast type, RMSE is calculated with respect to the analysis that was used to initialise that forecast (specifically, CERA for C1, U_atmos and U_ocean for U1, and ERA-Interim and ORAS4 for M1). ACC is calculated for daily mean precipitation, and all forecasts are evaluated against an independent observational dataset (i.e. one not assimilated during any of the analyses), from the Global Precipitation Climatology Project (GPCP; a daily-mean dataset at 1° spatial resolution) (Huffman et al. 2012), so as to avoid biasing the calculation towards one of forecast types, as would be the case if a particular analysis were used. In the calculation of ACC, forecast and observation ensemble means (averaged over the 30 start dates, at consistent lead times) are used as the climatologies (with respect to which anomalies are computed), since no longer record is available for the forecasts.

In several of the figures shown, confidence intervals, with respect to forecast biases or skill being significantly different from the corresponding values in C1, which is taken as a baseline case, are used. These are calculated using a non-parametric bootstrapping approach to account for the finite sample size, (following Goddard et al. 2013; Smith et al. 2013) (details of the procedure are given in the Supplementary Information).

3. Results

a. Shock in the lower atmosphere

In U1 and M1, the one-way coupling during the assimilation phase is such that continuity from analysis to forecast is provided in the ocean — by virtue of its forcing by the same atmospheric analysis used to provide the initial atmospheric state — but not in the atmosphere. The change in SST forcing experienced by the atmosphere at the beginning of the forecast is the switch from a gridded, observed product to the ocean analysis field, which itself was produced using nudging of SST towards the same observed product (Fig. 1). Therefore, the shock in the near-surface atmosphere can be expected to be a function of the accuracy with which the ocean analysis U_ocean reproduces the SST field towards which it has been nudged.

Fig. 2(a) shows the root-mean-square difference (RMSD) between the SST seen by the atmosphere during analysis (i.e. the gridded observed products) and the SST produced by the ocean analysis U_ocean as initial conditions for the U1 forecasts. Discrepancies are largest in regions of large SST temporal variability, near the northern hemisphere western boundary currents, in the eastern tropical Pacific (particularly during Aug–Sep 2010, when tropical instability waves are most active) and in the Antarctic Circumpolar Current. These are also areas in which model biases, which the assimilation attempts to correct, are large. It is these areas in which shocks due to component imbalance may be expected.

Fig. 3 shows the RMSE, after 12 hours, of forecast air temperature at 1000 hPa, for C1 (compared to CERA), U1 (compared to U_atmos) and M1 (compared to ERA-Interim), averaged over all forecast start dates. Widespread errors are present in C1 (Fig. 3(a)), forming due to the presence of biases in the models and to any imperfections in the coupled analysis initialisation method. These errors do not constitute the initialisation shock that is being investigated here, according to our earlier definition. Therefore, C1 is taken as a baseline case, such that any further deviation of a forecast from its reference analysis should represent a shock imparted by an initialisation procedure that differs from that of C1.

Relative to C1, U1 (Fig. 3(b)) shows, over the ocean, small but significant increases in RMSE in several areas, which are generally those areas in which the RMSE between the two SST fields, as shown in Fig. 2(a), is largest. This air temperature shock signal in U1 therefore appears to develop primarily due to the change in SST forcing felt by the atmosphere after the transition from the analysis to the forecast phase. Correlations between the initial SST discrepancy and the 12 h air temperature error in U1 minus that in C1, calculated across the 30-date forecast set, are significant in the same areas of strong SST variability (Fig. S1(a)), confirming that the development of air temperature biases in excess of those found in C1, can be attributed to the imbalance between atmosphere and ocean at the beginning of the U1 forecasts. These air temperature shocks are generally of magnitude 0.2 K or less, but compared to the small baseline RMSE seen in most areas in C1 (Fig. 3(a)), they represent substantial error amplifications: RMSE is increased by 50% or more in the eastern equatorial Pacific, eastern tropical Atlantic, northern Pacific and across most of the Southern Ocean, and it is more than doubled in the Gulf Stream and Arctic regions.

The difference between ORAS4 SST and the gridded products used by ERA-Interim (Fig. 2(b)) shows a similar spatial pattern to the differences between the operational analyses, but with slightly larger values (by an average of $\sim$15%) in most areas, indicating a greater imbalance and larger discontinuity felt by the atmosphere at the beginning of a forecast. These increases in RMSD are partly the result of small differences between the SST products used by ERA-Interim and ORAS4 during two of the three periods covered by these experiments. However, the 1000 hPa air temperature shock in M1 (Fig. 3(c)) is rather different to that in U1: RMSE is increased relative to C1 over most of the ocean, in contrast to the limited areas of amplification seen in U1. Correlations between initial SST discrepancy and 12 h air temperature shock are again significant in some regions (Fig. S1(b)), but are uniformly weaker than those of U1, suggesting the existence of another source of air temperature shock in M1. Also, there is little significant correlation to explain the shocks in parts of the North Pacific, the Southern Ocean near Antarctica and in the Arctic, in which regions (along with most of the globe) the bias is increased several times over its baseline (C1) values.

The additional source of atmospheric initialisation shock in M1 is the change in both atmosphere and ocean model versions that occurs between analysis and forecast, combined with the change in atmospheric vertical resolution. The change in atmospheric model is likely to be the more important with respect to shock in the atmosphere, though the change of ocean model could also contribute (as explored further in the next subsection). Model differences lead to a shock that increases errors above those of C1, over most of the planet, by the end of the first day.

Fig. 4 compares the RMSE in air temperature throughout the atmospheric column after 24 hours in the forecast types C1, U1 and M1, each evaluated against the analysis used for their initialisation, averaged over the Niño3 region (150–90°W, 5°N–5°S). In agreement with the interpretation of the U1 near-surface temperature shock as arising from the initial atmosphere-ocean imbalance, statistically significant differences in RMSE between U1 and C1 are limited to the lower atmosphere (at and below $\sim 850\,\mathrm{hPa}$). In M1, however, RMSE is amplified compared to C1 at all pressure levels, implying the occurrence of a shock that is spread throughout the atmosphere. This effect might very well arise from the difference in vertical resolution that exists between analysis and forecast (60 and 91 vertical levels respectively), together with differences in physics between the two model versions. Note that the errors at this point in the forecast are generally at least as large as differences between the three analyses.

So, although atmospheric initialisation shocks do occur as a result of imbalanced initial conditions (i.e. shocks of the first ‘type’ as listed in Section 1), the evidence here suggests that these are smaller than the adjustments that occur following a change in the atmospheric model (shocks of the second type). In the present case the change is merely from an older to a newer version of the same model, and a larger effect can be anticipated if initial conditions are obtained from a structurally different model altogether.

Fig. 5 shows the evolution of the air temperature forecast error at $1000\,\mathrm{hPa}$ for C1, U1 and M1 against their own analyses, averaged over the Niño3 region. The larger error growth in U1 compared to C1 results from the SST discrepancies shown in Fig. 2(a) during the first day, and the effects of the shock are felt out to at least 10 days’ lead time, through a $\sim$ 5–10% increase in RMSE, showing that initialisation shocks have the potential to impact medium-range (as well as short-range) forecasts. In M1, the effect of the difference in vertical resolution between the forecast and the reference analysis can be seen at lead time $t = 0$, and RMSE rises sharply on day one of the forecast, indicating a strong shock following the change in model version/resolution. Part of the difference between M1 and U1 may be attributable to the lower vertical resolution of

M1 (the number of vertical levels in the lowest ~ 1 km is reduced by around a third compared to U1).

b. Shock in the upper ocean

In the upper ocean, markedly different bias development is seen in M1 compared to the other two forecast types, particularly near the equator. Fig. 6 plots the time series in SST averaged in the Niño3 region, for the three forecast types and their corresponding analyses, in the period Dec–Jan 2008/9 only. In M1, a large shock occurs at the beginning of the forecast, and a cold bias of around 0.5 K has formed after 6 hours, the first output point in the forecast series. A shock of around this size forms consistently ($\pm$20%) in each of the 10 forecasts in this period, and the identification of this error is clearly not sensitive to the reference SST used. The other two periods, shown in Fig. S2, feature similar cold shocks, but with different magnitudes. The shock is therefore a robust effect, but shows some seasonal variation, due to seasonal variation in the difference between the climatological states of the analysis and forecast versions of the ocean model. After the initial shock, a correction is seen to occur; nevertheless, by day 10, the M1 error is still significantly larger than errors in the other forecasts. In this case, the initialisation shock has increased the forecast error, though in general the shock need not be of the same sign as the forecast drift (see e.g. Fig. S2(a)). A similar shock, though with smaller magnitude, is seen in the eastern equatorial Atlantic (see Fig. S3).

The source of this drift is dynamical differences between the two ocean model versions (as used in ORAS4 and M1 respectively; see Tables 1 and 2), combined with differences in ocean analysis methodology. Upper ocean vertical profiles in the Niño3 region, plotted in Fig. 7, show that the ORAS4 analysis (run with NEMO v3.0) features stronger (by up to 50%) upwelling velocities than CERA and U_ocean at 50 m depth and below. All three analyses are nudged to the same (or a very similar) SST field (analysed Niño3 SSTs show a spread of $\sim 0.2^\circ$ K), and the zonal wind forcings supplied to the ocean analyses (from CERA, U_atmos and ERA-Interim) are very similar (not shown), so differences in upwelling must be due to ocean model differences between the two versions used to perform the analyses, and differences in the treatment of model bias during the analysis (examined further in Section 3d). The shock that occurs in Niño3 in the M1 forecasts does so as a result of the use of the ORAS4 equatorial ocean state as initial conditions in the newer version of NEMO, which normally (in U_ocean, with no bias correction) produces realistic near-surface temperatures with much weaker upwelling. The stronger vertical velocities, as well as colder waters at 50–150 m, while not necessarily less realistic than U_ocean, cause the rapid surface cooling due to their incompatibility with the forecast model. The partial recovery of Niño3 SST in Fig. 6 can be interpreted as the equatorial ocean circulation adjusting (weakening) through the use of the newer model version. Differences between the analyses vary seasonally, correlated with the size of the SST shock in M1 in the three forecast periods. A similar explanation can be found for the (weaker) shock that occurs in the eastern equatorial Atlantic.

Returning to Fig. 6, it is seen that the drift in C1, which can again be taken as a baseline case, is small in Niño3 in this season, though more substantive drifts do occur in other seasons (Fig. S2). In U1, a cold bias can be seen to form at the beginning of the forecast. However, the source of this bias is not the same as that of the M1 shock. The source is the weak diurnal variation present in SST in the U_ocean analysis, as a result of the use of daily-mean fluxes in its production. Since forecasts are initialised at 00:00 UTC, a longitude-dependent bias forms once the coupled forecast model generates a larger diurnal SST signal. In the eastern Pacific, the initial SST value, which is essentially a daily-mean value, is too cold given the local time of day (seen by comparing the C1 and U1 lines at $t = 0$), so, as the region cools in the evening, a bias develops relative to U_ocean. The opposite effect occurs in the Indian Ocean (Fig. S3(b)). C1, on the other hand, does not show this drift, as the CERA ocean analysis includes some diurnal SST variation by virtue of its frequent coupling to the atmosphere during the analysis. The time-of-day effect might be considered to be a legitimate form of shock (in line with the definition given in Section 1), stemming from a lack of coupling during the ocean analysis. However, in principle it is possible to obtain a stronger SST diurnal cycle from an uncoupled ocean analysis by forcing using a higher-frequency atmospheric flux product.

Errors introduced due to this effect are of order 0.1 K, and appear to account for most of the U1 drift in this region, which is otherwise not much different to that of C1, implying a limited impact of imbalance-driven shock on SST. Nevertheless, correlations between the SST and air temperature shocks do suggest that part of the U1 SST drift in the eastern Pacific arises due to a compensatory ocean cooling in response to the overlying atmospheric cold shock (Fig. 3(b)).

c. Impact on forecast skill

Having established that initialisation shocks do occur in the upper ocean and in the atmosphere in the forecasts initialised using uncoupled data assimilation, we now investigate whether or not these shocks have any detrimental impact on the forecast skill, using daily average precipitation rates evaluated against GPCP observed rates. The use of an independent reference dataset such as this is the only way to meaningfully compare forecast skill among the different experiments, since each was initialised using a different analysis.

Fig. 8 shows that, in both the tropics and extratropics, differences in forecast skill between C1 and U1, which should form solely due to the effects of shock due to initial imbalance, are very small and generally not significant, implying that the impact on forecast skill of this type of initialisation shock is, using this broad measure, slight. Although, where differences in these wide regional averages do briefly reach 90% significance (on two occasions in the northern extratropics) they do so with larger skill scores in C1 than in U1. A similar evaluation of skill in 1000 hPa temperature, measured against an independent reanalysis, also resulted in negligible differences between C1 and U1 (not shown). A much larger forecast set may be necessary to assess confidently the penalty in skill arising from imbalance-driven shock, since it appears to be a very small one, as far as can be determined from this set.

The precipitation forecast skill of the M1 forecasts (not shown) is consistently lower, by ~ 0.03, than C1 and U1. While this could suggest a sustained impact following the initialisation shock due to the change in model version, it is perhaps more likely to be a symptom of the slightly lower vertical resolution used in M1, and of the less accurate initial atmospheric state provided by ERA-Interim compared to the initial states used in C1 and U1.

d. Sensitivity to ocean initial conditions

Although dynamical differences between the two ocean model versions were seen earlier to explain at least partly the SST shock in M1, there is another difference between the ocean initialisation methods of M1 and U1 — the use of bias correction during the analysis in M1, and not in U1. Bias correction during the assimilation attempts to prevent the rapid destruction of increments by a biased model, and has an impact on ocean velocities, particularly close to the equator, where model biases tend to be large due to uncertain wind stress forcing of the upper ocean (Bell et al. 2004; Balmaseda et al. 2007). The use of bias correction leads ultimately to a different ocean initial state, in the same way as does the use of a different model during analysis. To clarify the reasons for upper ocean shock in M1, a further two sets of forecasts, M2 and M3, were run. Both used ERA-Interim as the atmospheric initial conditions, like M1, and both used the same resolutions as M1, but with different initialisations for the ocean.

Forecasts M2 used as initial conditions a different ocean analysis, one identical to ORAS4 but run without bias correction (ORAS4_nobiascrtn; see Balmaseda et al. (2013)). Due to a limited number of available restart files for this analysis, a smaller set of 6 forecast start dates were run in Apr–May 2008 and Dec–Jan 2008/9, and no forecasts were possible in Aug–Sep 2010, so Aug–Sep 2008 was used instead. For all start dates used for M2, corresponding M1 forecasts were also run, enabling an accurate comparison between these two forecast types, to isolate the roles of changing model version and the use of bias correction, in initialisation shocks originating in the ocean. Then, to complete the attribution of shocks to the three sources identified in the introduction, a set of forecasts M3 was run (for the same 30 start dates as in M1) using the new uncoupled ocean analysis (U_ocean) as the ocean initial conditions. The results of M2 should isolate the contribution to the shock in the ocean of the removal of bias correction at the beginning of the forecast, as distinct from the contribution from a change in model version, while M3 should confirm that ocean shocks are predominantly caused by changes in the ocean component between analysis and forecast (and not by changes in the atmospheric component).

In M2, the shock at Niño3 (Fig. 9(a)) is only slightly weakened relative to M1 — there is an average reduction of $\sim 25\%$, with little variation across the three seasons — and is virtually unchanged in the eastern Atlantic (Fig. S4(a)). This confirms that the change in ocean model version, rather than the use of bias correction during the analysis, is the dominant cause of these equatorial cold shocks. Subsurface profiles (not shown) show that ORAS4_nobiascrtn upwelling velocities in the Niño3 region are up to 25% weaker than those in ORAS4, explaining the reduced surface cold shock.

In other areas, the shock in SST and/or air temperature is slightly increased in M2 relative to M1 (see Fig. S4(a) and (b)). Thus, the inclusion of bias correction in the initialising ocean analysis (and its removal during the forecast) imparts small shocks to the upper ocean and to the lower atmosphere (possibly through an increased component imbalance), which can either amplify or reduce the existing shocks following the change in model. In the tropics, the use of bias correction generally has a negative impact on the forecast, as it shifts the ocean analysis circulation into a state that cannot be maintained for any significant length of time from the beginning of the forecast, therefore resulting in an adjustment.

In M3, errors in the ocean develop in a similar manner to those of U1, as the two share the same ocean initial conditions. The large M1 shocks at Niño3 (Fig. 9(b)) and in the eastern equatorial Atlantic (see Fig. S4(c)) are entirely absent, confirming that the ocean initialisation is the source of the M1 shocks. The air temperature shock in the eastern Pacific (Fig. S4(d)) is also reduced, relative to M1 — the lack of cold shock in the underlying SST is likely the main reason for this, since the two biases (in SST and 1000 hPa temperature) are strongly correlated in this area in M1. A reduction in atmospheric shock here may arise also due to the slightly better initial balance present in this area in M3 (which is very similar to the balance in U1, shown in Fig. 2(a)) compared to M1. Elsewhere, air temperature RMSE is very similar to that of M1, confirming that it is the change in atmospheric model version that produces a large component of these widespread biases on the first day. The influence of the atmospheric initialisation on SST can be seen in the slightly increased SST drift in M3 compared to U1 (Fig. 9(b)).

4. Discussion

The results presented above suggest a definite impact on short-range forecasts of changes in ocean or atmosphere model between analysis and forecast, but show only a small (though significant) effect due to an imbalance in the initial conditions. An important factor in the performance of C1 and U1 forecasts is the use of nudging towards a complete gridded SST product, rather than assimilation of individual SST observations, in the ocean analyses. This ensures that U_ocean

SSTs remain, almost everywhere, very close to the observational product, the field that is seen by the atmosphere during U_atmos (see Fig. 2(a)). While this is beneficial with regard to minimising initialisation shock in U1, direct assimilation of satellite SST observations may be preferred to nudging in ocean analyses, since the latter is currently done by modifying air-sea fluxes rather than the ocean model itself (Balmaseda et al. 2013). If assimilation is used, any gaps in observational coverage will lead to periods without observational increments during which uncorrected model SSTs could diverge substantially from the field seen by the atmosphere. This would result in imbalances that are more temporally variable, and at times larger, than those shown in Fig. 2.

Therefore, the differences in C1 and U1 forecast RMSE and skill shown here should perhaps be taken as a lower limit. That is, the benefits of coupled DA to forecasting may be felt more strongly if assimilation of SST is used rather than nudging in the uncoupled ocean analysis, at least in any data-sparse regions. Where SST nudging is used in conjunction with one-directional coupling of separate ocean and atmosphere analyses, the gains in forecast skill due to reductions in initialisation shock following the implementation of a coupled DA system similar to CERA may, based on these results, be small. This is more a statement of the acceptable degree of balance achieved in the U1 initialisation than a criticism of coupled DA. Additionally, coupled DA may result in a more accurate analysis than uncoupled assimilation (Laloyaux et al. 2015), which could lead to further gains in forecast skill, separate to any achieved through reductions in initialisation shock, although this was not the case in the precipitation results shown here.

With regard to the relative merits of a more strongly coupled DA method (one involving the modelling of cross-covariances to spread information between the two model components), while this offers the potential to produce a more balanced initial state than is produced by CERA, which should in itself lead to better forecasts, it seems unlikely that forecast skill will be further improved specifically by a reduction in initialisation shock, judging by the similarity in skill between C1 and U1 (Fig. 8).

To mitigate the shocks that can result from the use of bias correction in the ocean analysis (Fig. 9(a)), it can be argued that the bias correction term, estimated during the initialisation phase, should be maintained during the forecast itself. This would not only reduce the overall initialisation shock, but would also slow the model drift. However, this is not possible in forecasts using uncoupled initialisation methods (such as U1, M1 and M2), due to the different biases present in the forced ocean model compared to the coupled forecast model (and potentially differences between the analysis and forecast models themselves). Such a method would be possible in a forecast of type C1, however, and the viability and usefulness of this approach should be investigated.

A further consideration, not described so far in this paper, is that large adjustments in the upper ocean (evidence of which was seen in Fig. 6) could generate shock signals that propagate beyond the 10-day duration of the forecasts shown here, due to the longer dynamical timescale of the upper ocean. Several exploratory 7-month forecasts, which are described in the Supplementary Information, have shown evidence of spurious Rossby waves propagating westward in the equatorial Pacific, following a change in ocean model version between analysis and forecast. Significant differences in the upper ocean between forecasts of type M1 and M3 were seen at lead times of up to 7 months (see Figs. S5 & S6). The impact on seasonal forecasts of using non-native ocean models for initialisation is a possible area for further study.

The results of this work should serve as useful guidance for medium-range and seasonal forecasting at operational centres. Besides finding hints of a slight increase in atmospheric forecast skill when using coupled DA rather than uncoupled assimilation methods for initialisation, we have also shown that initialisation shock can be generated through the use of non-native models for the creation of initial conditions. Depending on the resources available to an operational centre, using initial conditions derived from an older version of the operational forecast model, possibly at lower resolution, or from another model entirely, may be the most practical option for seasonal forecasting. Even if not the case for the forecast itself, this may be more common in performing the set of calibration hindcasts (e.g., MacLachlan et al. 2014) that forms an essential component of a seasonal forecast (and is also valuable at shorter ranges (Hamill et al. 2004)). The hindcasts are used to compute a posteriori bias correction terms, so it is important that the temporal evolution of bias in the hindcasts is as similar as possible to the development of bias in the forecast (as discussed by Hamill et al. 2004). In either case, it has been shown that using non-native analyses for forecast or hindcast initialisation does result in substantial initialisation shocks in both atmosphere and ocean.

Various studies have declined to use non-native atmospheric analyses directly as initial conditions for coupled forecasts, preferring to nudge towards these analyses (e.g., Hudson et al. 2011) or to initialise a model atmosphere by forcing with observed SSTs (e.g., Alessandri et al. 2010). The results above confirm that there is indeed good reason to avoid direct use of a non-native analysis (even when derived from the same model 'family') in initialisation, in the ocean as well as in the atmosphere, if possible. The detrimental aspect of nudging a forecast model towards such an analysis lies in the production of initial conditions that may lie further from the truth, and the optimal nudging strength — one which balances accuracy in the initial state with minimisation of shock, so as to produce the most skilful forecast — is likely to be strongly model dependent. For example, we have not investigated whether or not a forecast system initialised from ERA-Interim and ORAS4, and comprised of model versions 31r2 and 3.0 (see Table 1), outperforms M1 in forecast skill by removing a major component of the initialisation shock at the expense of using deprecated, and inferior, forecast models. The decision over whether or not to use the operational forecast system without also generating initial conditions using the same system will depend on the degree of improvement offered by the newer system in comparison to the one which generated the initial conditions that are already available. Our results do suggest that, where possible, initial conditions for both forecasts and hindcasts should be obtained through analyses using the same models.

5. Conclusions

We have identified initialisation shocks in sets of coupled forecasts for which the initial conditions were obtained using uncoupled data assimilation in the atmosphere and ocean. Three distinct sources of shock, with varying degrees of impact on the forecasts, have been identified:

1. A lack of balance between the atmospheric and oceanic components of the initial state exerts an influence on the forecast drift, as seen through the comparison of forecast types C1 and U1. Initialisation shocks of this type occur most strongly in regions of large SST temporal variability. Their impact on forecast skill, measured by ACC in total precipitation rates, appears to be neutral. This source of shock may be atypically weak in the present case due to the use of SST nudging in the ocean analysis, which limits the size of atmosphere-ocean imbalances that can form in the initial conditions.

2. A change in model version from analysis to forecast, which occurs in the atmosphere in M3 and in both ocean and atmosphere in M1 and M2, leads to larger and more widespread initialisation shocks. These occur due to differences between model attractors, and are particularly strong in the equatorial ocean, in the present case. Oceanic shocks have the potential to exert an influence on the seasonal timescale.

3. The use of bias correction during the ocean analysis, and its removal during the forecasts, can impart further initialisation shocks in the upper ocean, at least when different model versions

are used for analysis and forecast. These shocks are generally less substantial than those caused by the change in model.

These results strengthen the case for operational seasonal forecasting centres to perform new ocean and atmosphere reanalyses, and consistent sets of calibration hindcasts, whenever the operational model is upgraded. The benefit to forecasting of aiming to minimise initialisation shock by using coupled data assimilation to produce these analyses, rather than performing uncoupled assimilation using the operational models, is less clear, but may emerge more strongly if assimilation of SST is used in preference to nudging towards a gridded product, during the ocean analysis.

Acknowledgments. We thank three anonymous reviewers for their comments, which have improved the clarity of this manuscript. This work was funded by the UK Natural Environment Research Council (ERGODICS project), the European Space Agency (Data Assimilation Project) and the European Commission (ERA-CLIM2 FP7 project). The work was accomplished through an ECMWF Special Project (spgbhain).

References

Alessandri, A., A. Borrelli, S. Masina, A. Cherchi, S. Gualdi, A. Navarra, P. Di Pietro, and A. F. Carril, 2010: The INGV-CMCC seasonal prediction system: improved ocean initial conditions. *Mon. Weather Rev.*, **138 (7)**, 2930–2952.

Alves, O., Y. Yin, L. Shi, R. Wedd, D. Hudson, P. Okely, and H. Hendon, 2014: A coupled ensemble ocean data assimilation system for seasonal prediction and its comparison with other state-of-the-art systems. *EGU General Assembly Conference Abstracts*, EGU General Assembly Conference Abstracts, Vol. 16, 9487.

Anderson, D. L. T., J. Sheinbaum, and K. Haines, 1996: Data assimilation in ocean models. *Rep. Prog. Phys.*, **59 (10)**, 1209.

Arribas, A., and Coauthors, 2011: The GloSea4 ensemble prediction system for seasonal forecasting. *Mon. Weather Rev.*, **139 (6)**, 1891–1910.

Balmaseda, M., and D. Anderson, 2009: Impact of initialization strategies and observations on seasonal forecast skill. *Geophys. Res. Lett.*, **36 (1)**.

Balmaseda, M. A., D. Dee, A. Vidard, and D. L. T. Anderson, 2007: A multivariate treatment of bias for sequential data assimilation: Application to the tropical oceans. *Q. J. Roy. Meteor. Soc.*, **133 (622)**, 167–179.

Balmaseda, M. A., K. Mogensen, and A. T. Weaver, 2013: Evaluation of the ECMWF ocean reanalysis system ORAS4. *Q. J. Roy. Meteor. Soc.*, **139 (674)**, 1132–1161.

Balmaseda, M. A., and Coauthors, 2009: Ocean initialization for seasonal forecasts. *Oceanography*, **22 (3)**, 154.

Bell, M. J., M. J. Martin, and N. K. Nichols, 2004: Assimilation of data into an ocean model with systematic errors near the equator. *Q. J. Roy. Meteor. Soc.*, **130 (598)**, 873–893.

Daley, R., 1991: *Atmospheric data analysis*. Cambridge University Press.

Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Q. J. Roy. Meteor. Soc.*, **137 (656)**, 553–597.

Fu, X., B. Wang, D. E. Waliser, and L. Tao, 2007: Impact of atmosphere-ocean coupling on the predictability of monsoon intraseasonal oscillations. *J. Atmos. Sci.*, **64 (1)**, 157–174.

Gemmill, W., B. Katz, and X. Li, 2007: Daily real-time global sea surface temperature — High resolution analysis at NOAA/NCEP. Tech. rep., NOAA/NWS/NCEP/MMAB Office Note Nr. 260, 39 pp (available at: http://polar.ncep.noaa.gov/sst/).

Goddard, L., and Coauthors, 2013: A verification framework for interannual-to-decadal predictions experiments. *Clim. Dyn.*, **40 (1-2)**, 245–272.

Hamill, T. M., J. S. Whitaker, and X. Wei, 2004: Ensemble reforecasting: Improving medium-range forecast skill using retrospective forecasts. *Mon. Weather Rev.*, **132 (6)**, 1434–1447.

Hudson, D., O. Alves, H. H. Hendon, and G. Wang, 2011: The impact of atmospheric initialisation on seasonal prediction of tropical pacific SST. *Clim. Dyn.*, **36 (5-6)**, 1155–1171.

Huffman, G. J., D. T. Bolvin, and R. F. Adler, 2012: GPCP Version 1.2 1-Degree Daily (1DD) Precipitation Data Set. WDC-A, NCDC, Asheville, NC. NASA/GSFC, Data set accessed June 2014 at http://precip.gsfc.nasa.gov.

Janssen, P., and Coauthors, 2013: Air-sea interaction and surface waves. Technical Memorenda 712, ECMWF.

Klingaman, N. P., P. M. Inness, H. Weller, and J. M. Slingo, 2008: The importance of high-frequency sea surface temperature variability to the intraseasonal oscillation of Indian monsoon rainfall. *J. Climate*, **21 (23)**, 6119–6140.

Laloyaux, P., M. Balmaseda, D. Dee, K. Mogensen, and P. Janssen, 2015: A coupled data assimilation system for climate reanalysis, submitted to *Q. J. Roy. Meteor. Soc.*

Lea, D., I. Mirouze, M. Martin, A. Hines, C. Guiavarch, and A. Shelly, 2014: The Met Office Coupled Atmosphere/Land/Ocean/Sea-Ice Data Assimilation System. *EGU General Assembly Conference Abstracts*, Vol. 16, 12097.

MacLachlan, C., and Coauthors, 2014: Global Seasonal Forecast System version 5 (GloSea5): a high resolution seasonal forecast system. *Q. J. Roy. Meteor. Soc.*.

Magnusson, L., M. Alonso-Balmaseda, S. Corti, F. Molteni, and T. Stockdale, 2013: Evaluation of forecast strategies for seasonal and decadal forecasts in presence of systematic model errors. *Clim. Dyn.*, **41 (9-10)**, 2393–2409.

Marshall, A. G., D. Hudson, M. C. Wheeler, H. H. Hendon, and O. Alves, 2011: Assessing the simulation and prediction of rainfall associated with the MJO in the POAMA seasonal forecast system. *Clim. Dyn.*, **37 (11-12)**, 2129–2141.

Molteni, F., and Coauthors, 2011: *The new ECMWF seasonal forecast system (System 4)*. European Centre for Medium-Range Weather Forecasts.

Rahmstorf, S., 1995: Climate drift in an ocean model coupled to a simple, perfectly matched atmosphere. *Clim. Dyn.*, **11 (8)**, 447–458.

Reynolds, R. W., N. A. Rayner, T. M. Smith, D. C. Stokes, and W. Wang, 2002: An improved in situ and satellite SST analysis for climate. *J. Climate*, **15 (13)**, 1609–1625.

Saha, S., and Coauthors, 2006: The NCEP climate forecast system. *J. Climate*, **19 (15)**, 3483–3517.

Saha, S., and Coauthors, 2010: The NCEP climate forecast system reanalysis. *B. Am. Meteorol. Soc.*, **91 (8)**, 1015–1057.

Shelly, A., P. Xavier, D. Copsey, T. Johns, J. M. Rodríguez, S. Milton, and N. Klingaman, 2014: Coupled versus uncoupled hindcast simulations of the Madden-Julian Oscillation in the Year of Tropical Convection. *Geophys. Res. Lett.*, **41 (15)**, 5670–5677.

Smith, D. M., R. Eade, and H. Pohlmann, 2013: A comparison of full-field and anomaly initialization for seasonal to decadal climate prediction. *Clim. Dyn.*, **41 (11-12)**, 3325–3338.

Stark, J. D., C. J. Donlon, M. J. Martin, and M. E. McCulloch, 2007: OSTIA: An operational, high resolution, real time, global sea surface temperature analysis system. *OCEANS 2007-Europe*, IEEE, 1–4.

Vitart, F., and Coauthors, 2008: The new VAREPS-monthly forecasting system: A first step towards seamless prediction. *Q. J. Roy. Meteor. Soc.*, **134 (636)**, 1789–1799.

Wang, C., L. Zhang, S.-K. Lee, L. Wu, and C. R. Mechoso, 2014: A global perspective on CMIP5 climate model biases. *Nature Clim. Change*, **4 (3)**, 201–205.

Zhang, S., 2011: A study of impacts of coupled model initial shocks and state-parameter optimization on climate predictions using a simple pycnocline prediction model. *J. Climate*, **24 (23)**, 6210–6226.

Zhang, S., M. J. Harrison, A. Rosati, and A. Wittenberg, 2007: System design and evaluation of coupled ensemble data assimilation for global oceanic climate studies. *Mon. Weather Rev.*, **135 (10)**, 3541–3564.

LIST OF TABLES

TABLE 1. Details of the various analyses (atmosphere, ocean or coupled) that are used for forecast initialisation and as reference fields for forecast evaluation in this paper. The gridded SST product used is either the Operational Sea surface Temperature and sea Ice Analysis (OSTIA) (Stark et al. 2007) or one of two National Centers for Environmental Prediction (NCEP) products (Reynolds et al. 2002; Gemmill et al. 2007), depending on the time period (during 2008–2010) in question. The name 'CERA' is used to denote both its atmosphere and ocean components.

Name	Atmosphere/ocean	Model version	Resolution	SST treatment
CERA	Atmosphere and ocean	40r1 and 3.4	T159L137 and ORCA1	OSTIA/NCEP (nudged)
U_atmos	Atmosphere	40r1	T159L137	OSTIA/NCEP (prescribed)
ERA-Interim	Atmosphere	31r2	T255L60	OSTIA/NCEP (prescribed)
U_ocean	Ocean	3.4	ORCA1	OSTIA/NCEP (nudged)
ORAS4	Ocean	3.0	ORCA1	OSTIA/NCEP (nudged)
ORAS4_nobiascrtn	Ocean	3.0	ORCA1	OSTIA/NCEP (nudged)

TABLE 2. Description of forecast sets described in the text. All forecasts use the same operational coupled ocean-atmosphere model system (model versions 40r1 and 3.4 for IFS and NEMO respectively), but types differ in the model versions and settings used for their initialisation (refer to Table 1). The sources of shock considered are the three listed in Section 1.

Name	Details	Resolution	Atmos IC	Ocean IC	Sources of shock
C1	Coupled DA	T159L137/ORCA1	CERA	CERA	Baseline
U1	Uncoupled analyses, consistent models	T159L137/ORCA1	U_atmos	U_ocean	Surface imbalance
M1	Uncoupled analyses, change in models	T159L91/ORCA1	ERA-Interim	ORAS4	Surface imbalance, model version change, bias corr. removal
M2	Uncoupled analyses, change in models	T159L91/ORCA1	ERA-Interim	ORAS4_nobiascrtn	Surface imbalance, model version change
M3	Uncoupled analyses, change in atm. model	T159L91/ORCA1	ERA-Interim	U_ocean	Surface imbalance, model version change, bias corr. removal

LIST OF FIGURES

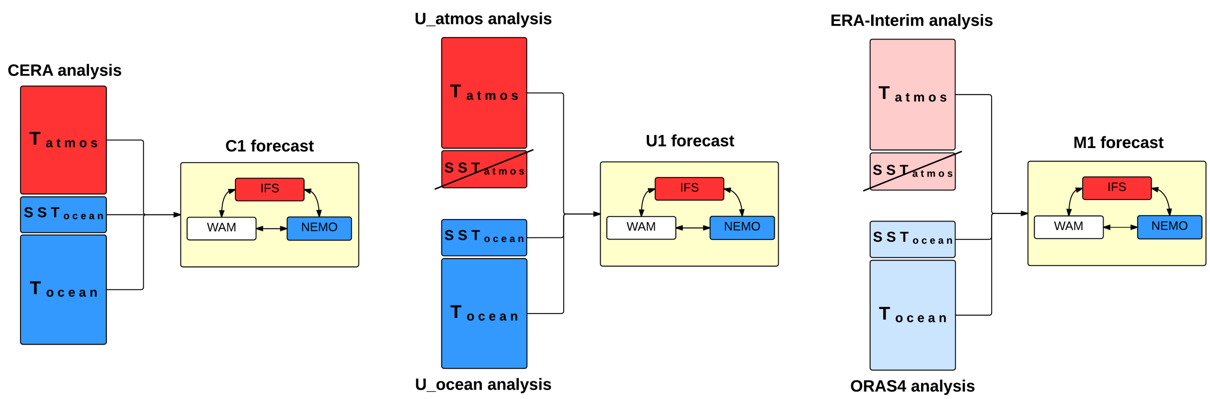


FIG. 1. The initialisation (analysis) methods used for forecast sets C1 (left), U1 (middle) and M1 (right). Colour coding indicates differences in model version, and elements of the analyses that are not used in forecast initialisation are marked with a diagonal line. (Forecast model components IFS, WAM and NEMO are the atmospheric, wave and ocean components respectively.)

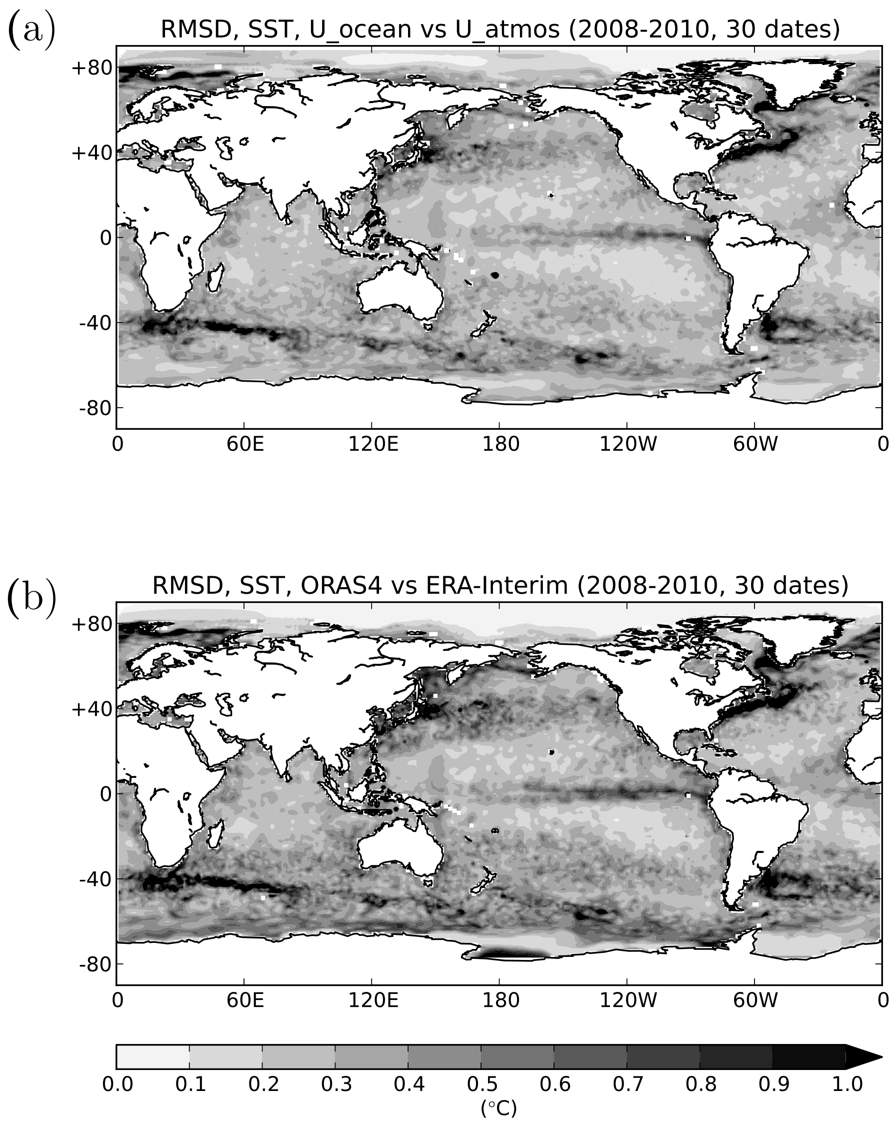


FIG. 2. (a) Root-mean-square difference between U_ocean SST and the SST used as forcing by U_atmos, at the beginning of the forecasts, showing the imbalance present in the initialisation of forecasts U1; (b) the same for ORAS4 and ERA-Interim, showing one of the sources of imbalance in the initialisation of forecasts M1.

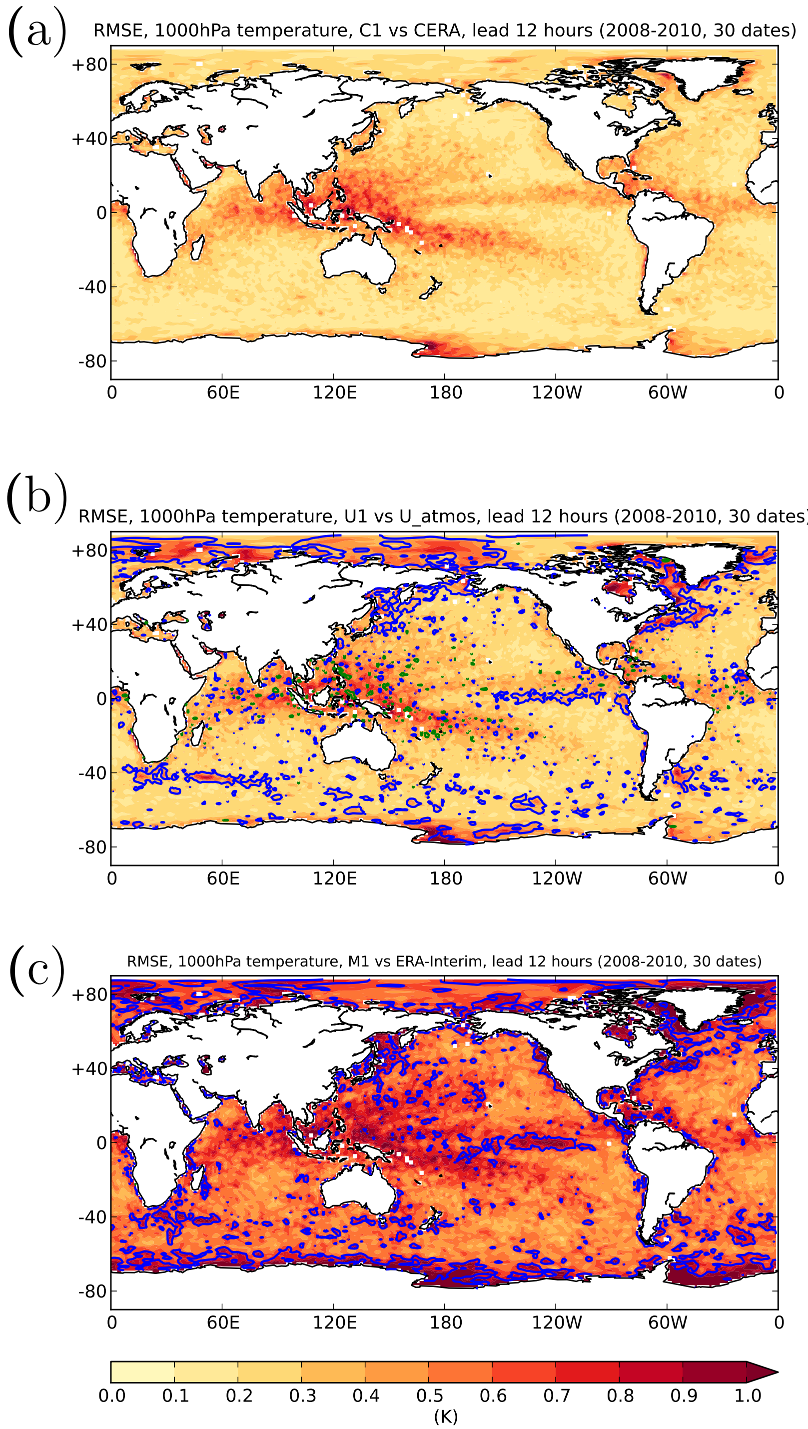


FIG. 3. 1000 hPa temperature forecast RMSE, relative to the analysis used as the initial conditions, for C1 (a), U1 (b) and M1 (c), at 12 h lead time. Land areas are masked out, as the focus is on atmosphere-ocean imbalance. Contours in (b) and (c) show differences in RMSE relative to C1, with blue (green) contours marking increased (decreased) RMSE in U1 and M1. Contours are drawn at differences of 0.15°C in (b), and at differences of 0.5°C in (c). Only differences that are significant at the 90% level, estimated using the bootstrapping method, are contoured.

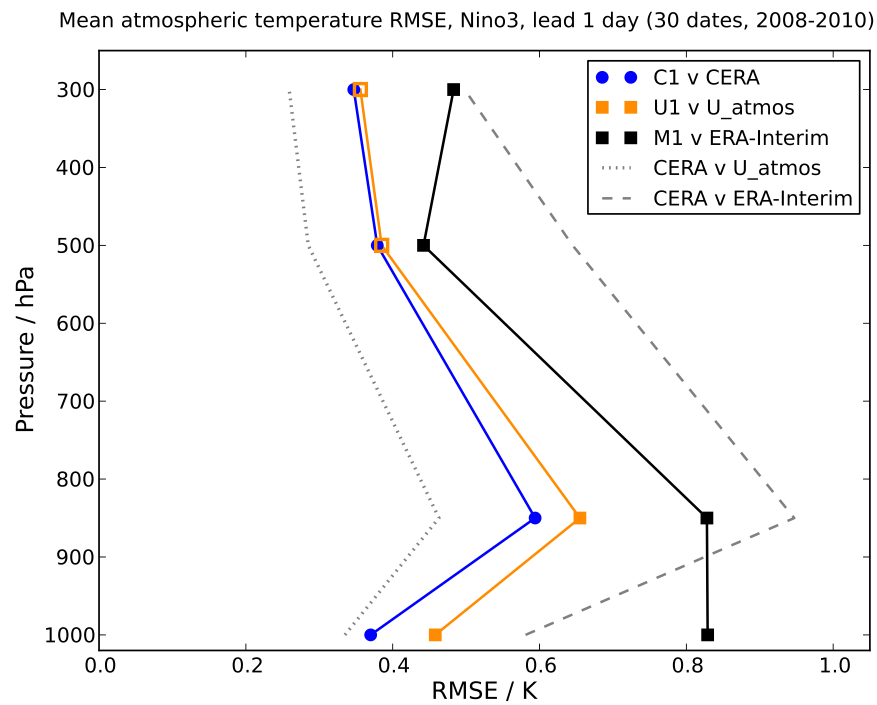
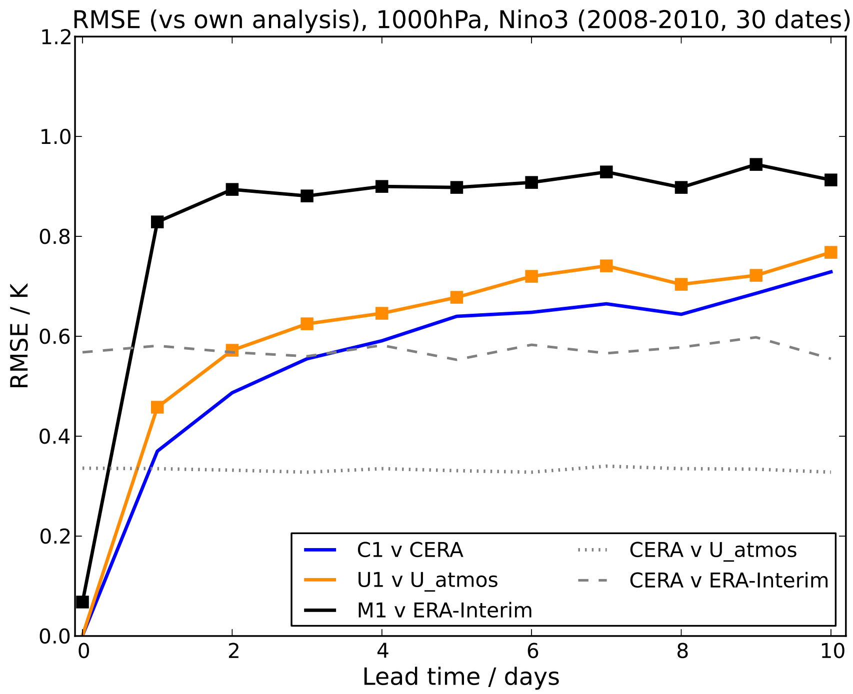
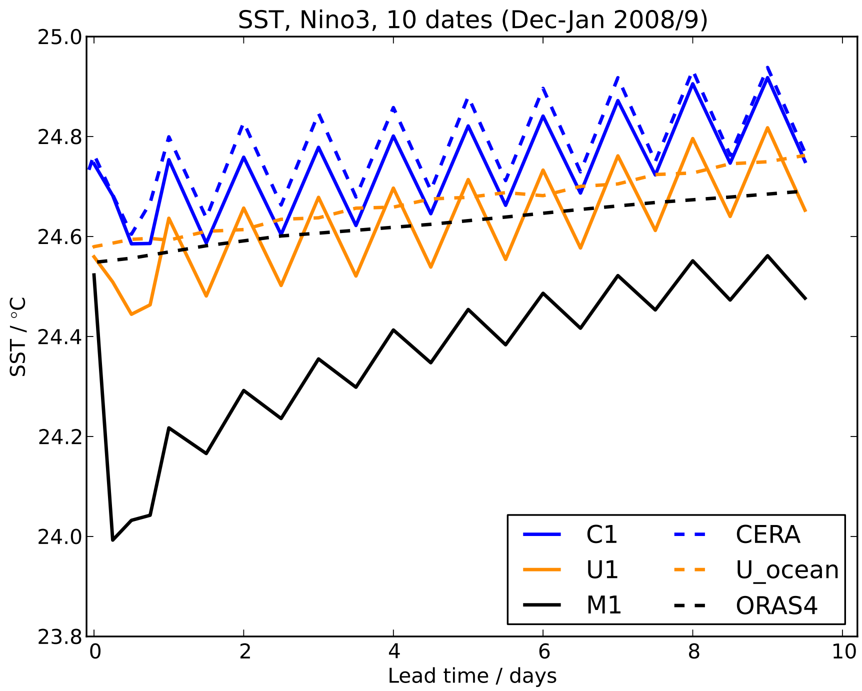


FIG. 4. Air temperature RMSE profiles averaged over the Niño3 region (150–90°W, 5°N–5°S), for C1 (blue), U1 (orange) and M1 (black), evaluated against CERA, U_atmos and ERA-Interim respectively, and RMSD profiles between CERA and the other two analyses (gray dashed and gray dotted). Filled (open) squares mark output pressure levels where the RMSE difference between U1 or M1 and C1 is significant (not significant) at the 90% level, estimated using the bootstrapping method.



738 FIG. 5. 1000 hPa temperature forecast RMSE averaged over the Niño3 region for C1 (blue), U1 (orange) and M1 (black) each evaluated against their own corresponding analysis, as labelled. RMSD between CERA and the other two analyses are shown for comparison (gray dashed and gray dotted). Squares mark where points in the U1 and M1 series are different from C1 at the 90% significance level, using confidence intervals calculated via the bootstrapping method.



743 FIG. 6. SST forecast and analyses time series for the 10 start dates in Dec–Jan 2008/9, averaged over the
744 Niño3 region. Forecast series are plotted at (0, 6, 12, 18, 24) hours, and every 12 hours thereafter; analysis series
745 for CERA and U_ocean are plotted at the same frequency (U_ocean features a very weak diurnal cycle), but only
746 daily means are plotted for ORAS4 (which also has a very weak diurnal cycle, not shown). Across the 10 start
747 dates, the magnitude of the drop from 0 to 6 hours in the M1 series ranges from 0.44 to 0.62°C.

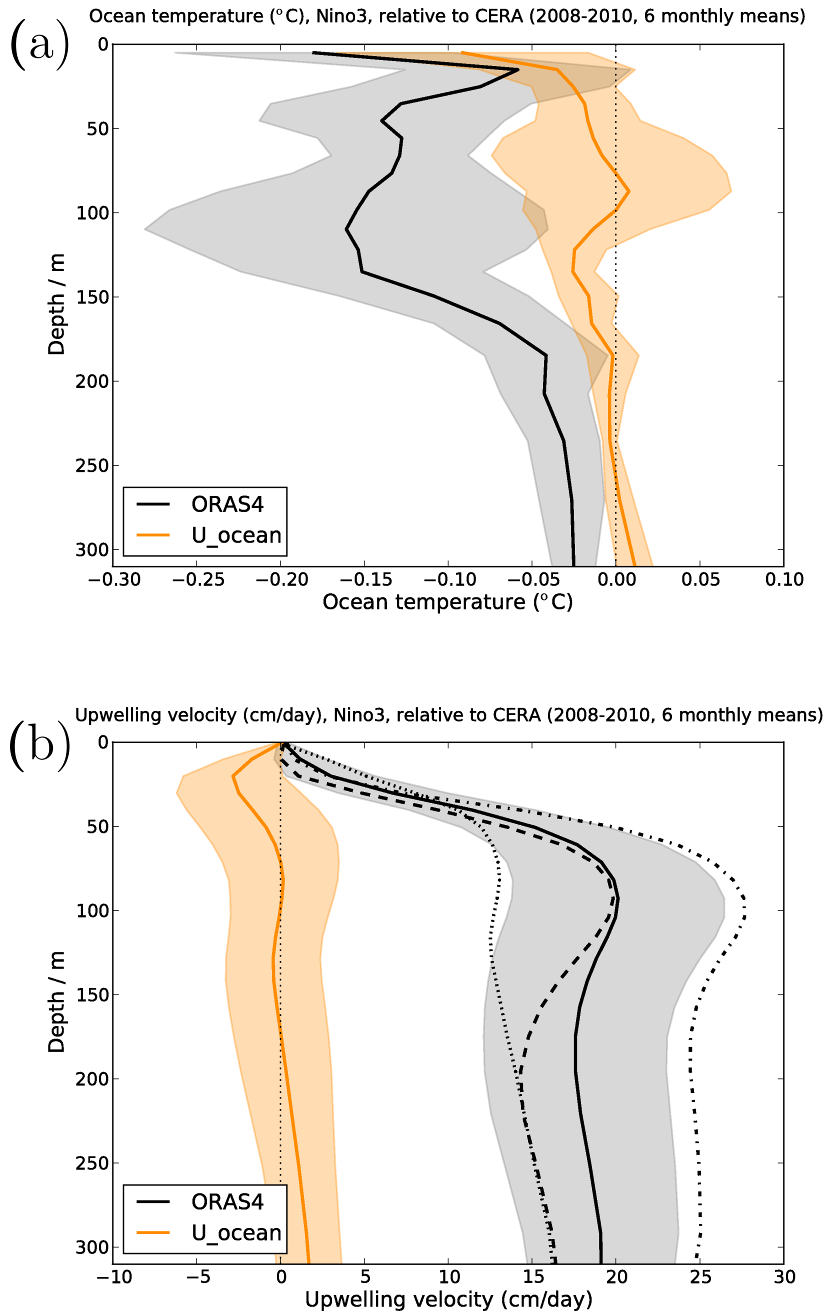


FIG. 7. Niño3 ocean temperature (a) and upwelling velocity (b) profiles from the ocean analyses U_ocean (orange) and ORAS4 (black), relative to CERA, constructed using monthly means for the 6 months in 2008–2010 during which forecasts were performed. Shading shows ± 1 standard deviation of the 6-month ensemble. Upwelling velocity profiles for each of the three forecast periods are also shown explicitly for ORAS4 (dotted: Apr–May 2008; dashed: Dec–Jan 2008/9; dash-dotted: Aug–Sep 2010).

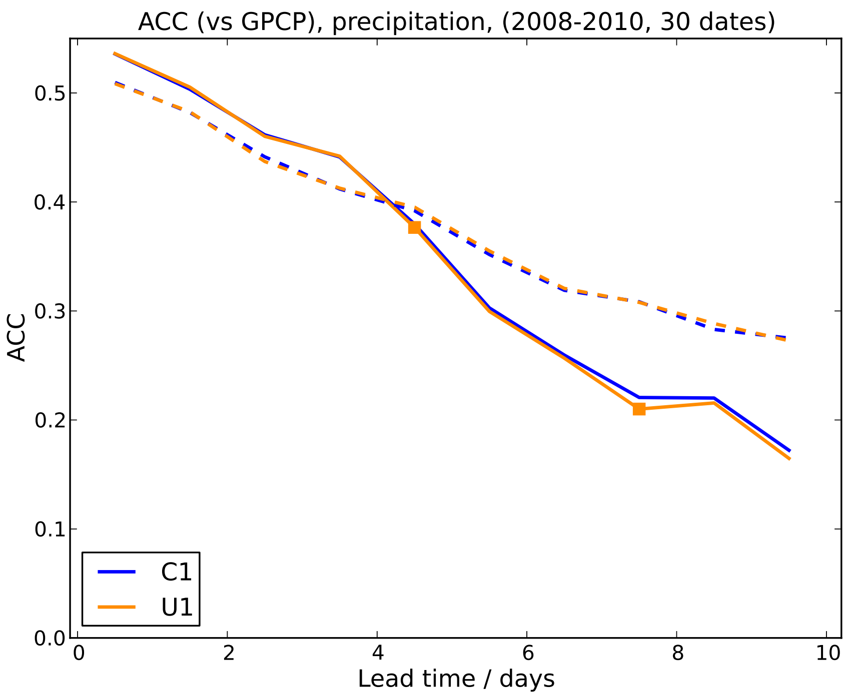


FIG. 8. Anomaly correlation coefficient for precipitation, evaluated against GPCP daily averages, in the tropics (20°N–20°S, dashed) and the northern extratropics (20–60°N, solid). Squares mark where points in the U1 series are different from C1 at the 90% significance level, using confidence intervals calculated via the bootstrapping method. Anomalies are calculated with respect to climatologies taken as the mean of the forecast period 2008–2010, which includes three different seasons, so some of the skill shown here is simply due to seasonal variability.

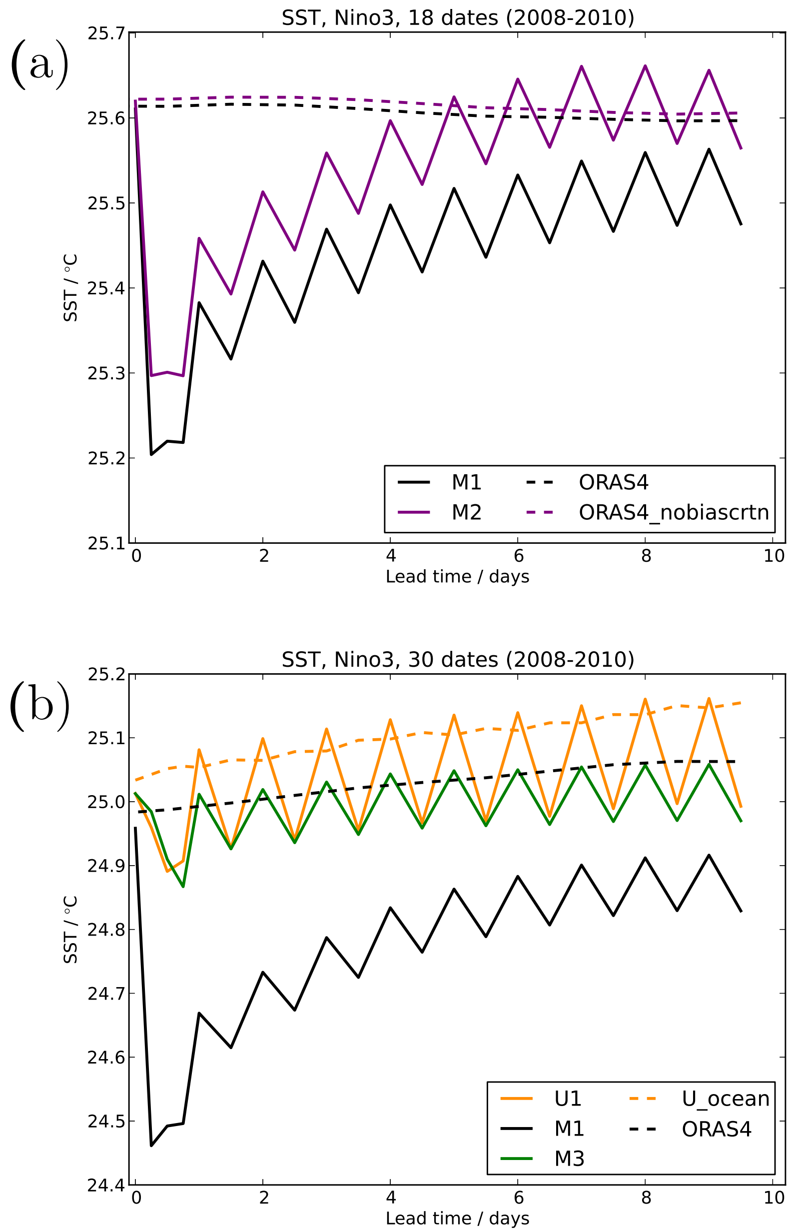


FIG. 9. (a) SST series for M1 and M2, and analyses ORAS4 and ORAS4_nobiascrtn, in the Niño3 region (where the largest shocks are produced in M1, M2 and M3); (b) SST series for U1, M1 and M3, and analyses ORAS4 and U_ocean, again in Niño3, averaged over the ensemble of 18 dates used for the M3 experiment. Forecast series are plotted at (0, 6, 12, 18, 24) hours, and every 12 hours thereafter; analysis series U_ocean is plotted at the same frequency, but only daily means are plotted for ORAS4 and ORAS4_nobiascrtn.