

Evaluation of forecast strategies for seasonal and decadal forecasts in presence of systematic model errors

Linus Magnusson · Magdalena Alonso-Balmaseda ·
Susanna Corti · Franco Molteni · Tim Stockdale

Received: 16 July 2012 / Accepted: 12 November 2012 / Published online: 28 November 2012
© The Author(s) 2012. This article is published with open access at Springerlink.com

Abstract This study discusses and compares three different strategies used to deal with model error in seasonal and decadal forecasts. The strategies discussed are the so-called full initialisation, anomaly initialisation and flux correction. In the full initialisation the coupled model is initialised to a state close to the real-world attractor and after initialisation the model drifts towards its own attractor, giving rise to model bias. The anomaly initialisation aims to initialise the model close to its own attractor, by initialising only the anomalies. The flux correction strategy aims to keep the model trajectory close to the real-world attractor by adding empirical corrections. These three strategies have been implemented in the ECMWF coupled model, and are evaluated at seasonal and decadal time-scales. The practical implications of the different strategies are also discussed. Results show that full initialisation results in a clear model drift towards a colder climate. The anomaly initialisation is able to reduce the drift, by initialising around the model mean state. However, the erroneous model mean state results in degraded seasonal forecast skill. The best results on the seasonal time-scale are obtained using momentum-flux correction, mainly because it avoids the positive feedback responsible for a strong cold bias in the tropical Pacific. It is likely that these results are model dependent: the coupled model used here shows a strong cold bias in the Central Pacific, resulting from a positive coupled feedback between winds and SST.

At decadal time-scales it is difficult to determine whether any of the strategies is superior to the others.

Keywords Forecast strategies · Seasonal forecasting · Decadal forecasting · Anomaly initialisation · Flux correction

1 Introduction

Systematic model error is a difficult problem for seasonal forecasting and climate predictions. Systematic model error means that the climatology of the model is different to observed climatology. The term “climatology” refers to the probability density function of the climate, which is often characterised by its mean (the mean of a variable over a long period) and the variability around this mean state. The climatologies could therefore differ either in the mean and/or the variability. We will use the term climatology as the subspace of the phase space covered by the model trajectories over a long period of time (sometimes referred to as the attractor of the system). In a non-linear system, the different moments of the probability distribution are linked, and errors in the mean state could affect the variability of the system.

Systematic model error leads to difficulties in the forecasting process. Particular problems happen when transferring information between observation space and model space, namely the initialisation and the issuing of the forecast. At the initialisation stage, information needs to be transferred from observations to model space. When issuing the forecast, the model output needs to be calibrated using reliable information about the real world. In numerical weather prediction (NWP) the forecast covers typically the range 1–15 days, and, because of the

L. Magnusson (✉) · M. Alonso-Balmaseda · S. Corti ·
F. Molteni · T. Stockdale
ECMWF, Shinfield Park, RG2 9AX Reading, UK
e-mail: Linus.Magnusson@ecmwf.int

S. Corti
ISAC-CNR, Bologna, Italy

relatively short forecast time, the difference between model and observed climatologies can be ignored (i.e. the model error is neglected). At longer lead times (monthly, seasonal and decadal time-scales) the systematic model error cannot be ignored. Due to the systematic model errors, the state of the model will drift away from the real-world attractor towards its own attractor, leading to a biased state. In these cases a strategy for accounting for the model bias is needed. In this study we present and compare different forecast strategies to cope with model bias.

Figure 1 illustrates the concepts behind the forecast strategies. In full initialisation (red), the model is initialised from an analysis that is close to the actual state, and can be assumed to be on the "real-world attractor". After initialisation, the state of the model drifts towards the model attractor. Therefore, before issuing the forecast, the model output needs to be calibrated. In monthly and seasonal forecasting, the commonly used calibration is a simple a-posteriori bias correction [i.e. correction of the mean only, assuming no interaction between mean state and variability, (Stockdale 1997)]. The bias is corrected by applying a lead-time dependent bias correction in the post-processing. For this correction a large data set of hindcasts (retrospective forecasts) is needed.

For anomaly initialisation (purple), the aim is to initialise the model on its own attractor by adding the observed anomalies relative to observed climate (estimated from a set of re-analyses) to the model climatology (estimated by multiyear forecasts). Then by construction, the model drift should be avoided. This strategy is popular to initialise decadal predictions (Smith et al. 2007 among others).

The third alternative to be discussed here is flux correction (blue). The aim of this strategy is to avoid (or limit) the model drift by adding a correction term to the model during the simulation that pushes the model solution towards the climatology of nature. Although flux correction was widely used in early work with coupled global circulation models, it has largely been considered "taboo" by the scientific

community since the paper by Neelin and Dijkstra (1995), where they argue that flux corrections could lead to non-natural variability patterns by disturbing the feedbacks operating in a free dynamical system. Indeed, flux correction should be avoided if the aim is the study of coupled feedbacks, and can be misleading for model development. In this report we will discuss the results from a forecasting perspective, i.e. whether flux correction can deliver an improved forecast, which is a pragmatic point of view.

In this comparative study we discuss the practical difficulties and advantages of the different forecast strategies, with particular focus on seasonal and decadal forecasts. These three methodologies are applied to the ECMWF coupled model. In a companion paper (Magnusson et al. 2012b), the ENSO variability and its dependence on the model state is discussed, which have implications for the choice of forecast strategy.

This paper is organised as follows: the model system and experimental setup is described in Sect. 2; the three different forecast strategies in Sect. 3; the results from the different strategies regarding model drift in Sect. 4 and in terms of quality scores in Sect. 5. Finally the findings are discussed in Sect. 6.

2 Model setup and experiments

The model used for this study is the ECMWF IFS model (model version 36r1) coupled with the NEMO ocean model version 3.0 (Madec 2008). The atmospheric resolution is T159 (corresponds to an horizontal resolution of 150 km) and 91 vertical levels. The ocean uses the ORCA1 grid, a tripolar grid with a horizontal resolution of about 1 degree at mid latitudes, with a finer meridional resolution (approximately 0.3 degrees) at the equator. There is no prognostic sea-ice model, instead the sea-ice concentration is prescribed from observed states, randomly selected from any of the 5 years previous to the forecast starting date [for details see Molteni et al. (2011)]. The model runs include increased greenhouse gases following observed values. Tropospheric and stratospheric aerosols are included in the model only as fixed climatologies, so no account is taken of volcanic eruptions and changes in anthropogenic "pollution". The model has not been specifically tuned to perform (near-term) climate simulations, as is the case for EC-EARTH (Hazeleger et al. 2010), which uses a similar model system.

The initial conditions for the atmosphere are provided by the ERA-40 reanalysis (Uppala et al. 2005) for starting dates prior to 1989, after which the ERA Interim reanalysis (Dee et al. 2011) is used. The ocean initial conditions are from a reanalysis based on NEMOVAR (Balmaseda et al. 2010b) oceanic reanalysis. The ocean reanalysis uses fluxes from the ERA-reanalyses as well as sub-surface

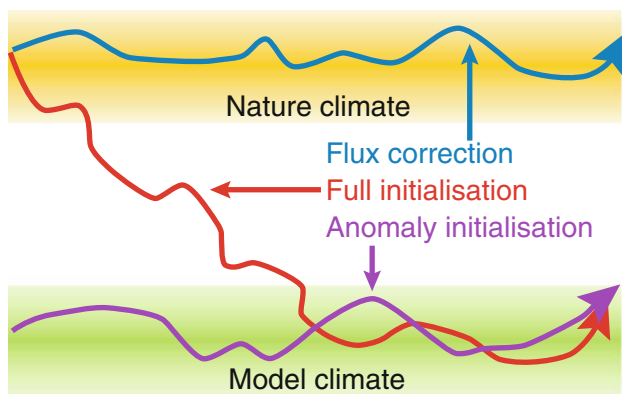


Fig. 1 Conceptual model of the forecast strategies

observations. The forecast ensemble is constructed by applying stochastic physics (SPPT scheme) to simulate model uncertainties in the atmosphere (Palmer et al. 2009).

Table 1 (model climate experiments) and Table 2 (hindcast experiments) show a summary of the experiments that have been undertaken. To obtain an estimate of the model climate, 3-member ensembles initialised in 1965, 1975 and 1985 have been run for 25 years (referred to as Control in what follows). These simulations are used to calculate the model climate for the anomaly initialisation (see below) as well as for diagnostics. An additional set of 25-year (1965, 1975) and 23-year (1985) forecasts was conducted where the SST (sea-surface temperature) were strongly constrained to observations (StrongRelax), using a relaxation strength of 1,000 W/K. The resulting atmospheric fields are equivalent to those obtained by an AMIP run (atmospheric only simulation forced by observed SST). This methodology has been used to initialise coupled models (Keenlyside et al. 2008), but here it will be used for the calculation of the momentum-flux correction. The SST data used for the relaxation is the same as for ERA-40 up to 1981 and after that Reynolds version 2 (Reynolds et al. 2002). To estimate the heat-flux correction needed when the momentum-flux correction is applied, an experiment with a weak SST relaxation (WeakRelax) has been undertaken (see below).

In order to evaluate the forecast strategies, one set of hindcast experiments (Table 2) has been run on a decadal time-scale while another set is run on an annual time-scale with an increased number of start dates and ensemble members in order to get more reliable forecast statistics.

Table 1 Experiment configurations for model climate estimations and flux-correction calculations

Name	Forecast length (years)	Members	Flux correction	Number of initial dates
Control	25	3	None	3
StrongRelax	25 (23)	3	None	3
WeakRelax	25	3	Momentum	2

Table 2 Configurations for hindcast experiments

Name	Forecast length (months)	Members	Initialisation	Flux correction	Number of initial dates
FullIni	120	7	Full	None	9
AnoIni	120	7	Anomaly	None	9
Ucorr	120	3	Full	Momentum	9
UHcorr	120	7	Full	Heat and momentum	9
FullIni	14	10	Full	None	30
AnoIni	14	10	Anomaly	None	30
Ucorr	14	10	Full	Momentum	30
UHcorr	14	10	Full	Heat and momentum	30

3 Methods

In this section we will explain the different forecast strategies used in this paper. As a background to the motivation of the different strategies, we will start with highlighting some of the systematic errors in the model.

Figure 2 shows the sea-surface temperature bias for the Control forecast averaged for forecast year 14–23, where we have discarded the first 13 years in order to let the coupled model drift to its climate. The bias has been calculated with respect to ERA Interim. The modelled SSTs are in general too cold compared to the reanalysis. Exceptions from the cold bias occur in the southern ocean and in the vicinities of the western boundary currents and around the southern tip of Greenland. The coldest bias is found in the mid of the northern Atlantic with a bias of more than 6 Kelvin. This bias is believed to be due to poor representation of the path of the Gulf Stream caused by low model resolution.

In the tropical Pacific the cold bias is pronounced with a too intense cold tongue. This systematic error is of importance for this study due to its connection to the ENSO. In the western part of the tropical Pacific a warm bias is present at depth, extending all across the Equatorial Pacific at the depth of the thermocline (not shown). The pattern of the errors (cold surface and warm subsurface) is indicative of too diffuse thermocline. The warm bias is stronger in the western Pacific, as if the thermocline in the coupled model is not only more diffuse, but also has a strong zonal gradient. This error in the slope of the thermocline is the fingerprint of too strong zonal winds.

Figure 3a shows the bias in the zonal component of the 10-m wind speed for the coupled model. Generally, the bias is less than 1 m/s, with a few exceptions. The largest bias appears in the western tropical Pacific, with a bias up to 3 m/s. The bias is of the same order of magnitude as the wind speed in the atmospheric reanalysis, meaning that the wind speed in the model is about twice that in the reanalysis.

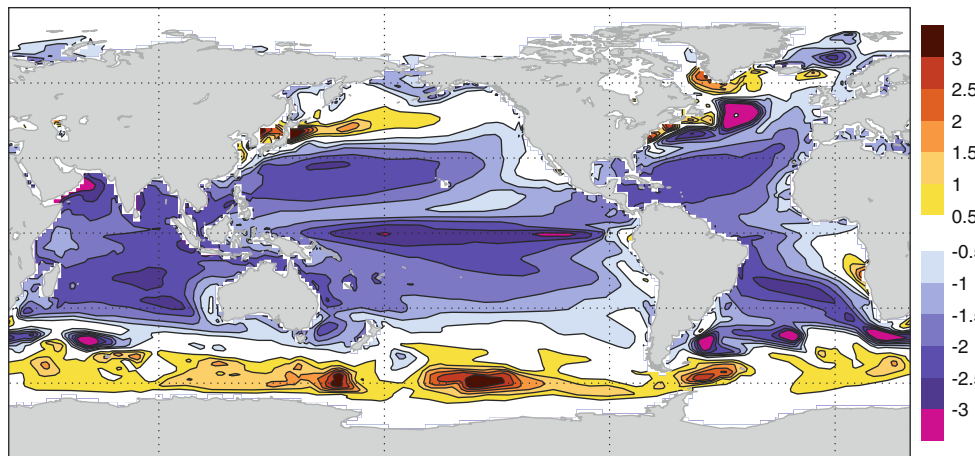


Fig. 2 Bias in sea-surface temperature for the long control simulation, forecast year 14–23

Fig. 3 Bias in zonal 10-m wind. Forecast year 14–23

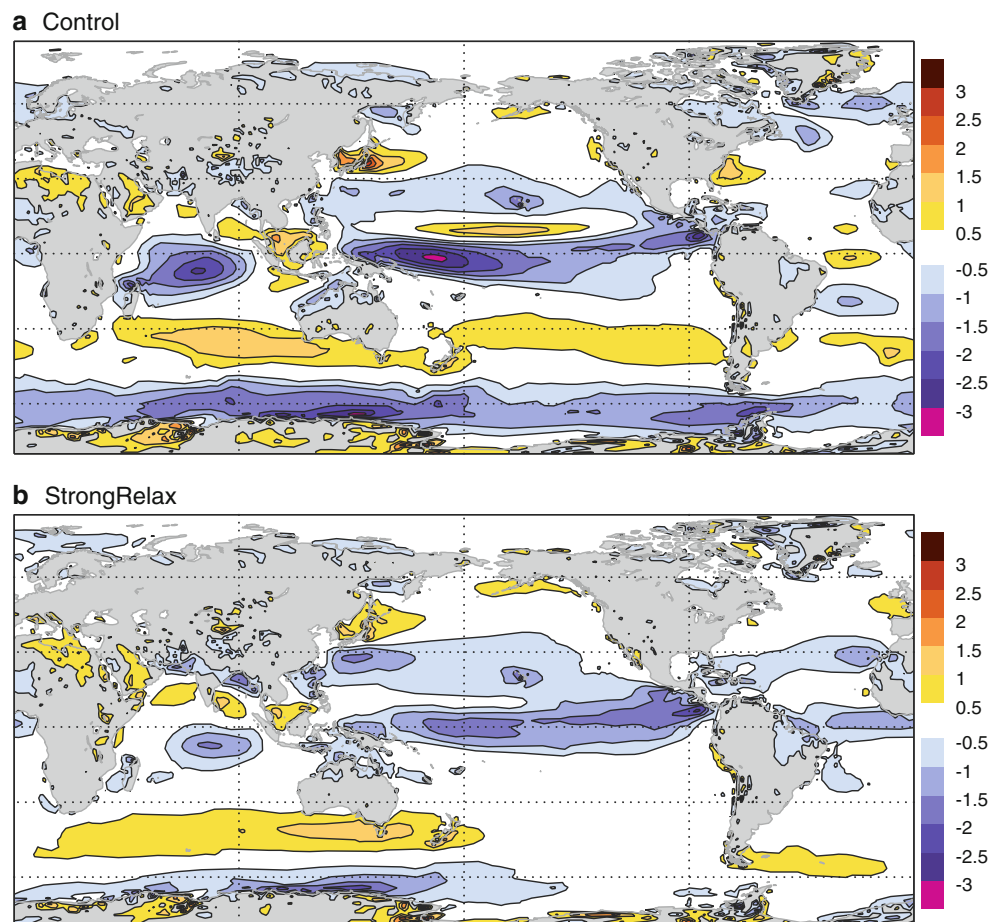


Figure 3b shows the wind bias in the StrongRelax experiment, which is reduced with respect to the free coupled model. This result clearly shows that if the SST bias is removed the wind bias in the tropical Pacific is reduced. The impact of the SST bias is especially strong in the western part of the basin where the wind bias is reduced by 50 % in the StrongRelax experiment compared with the

Control. Note, though, that substantial wind biases are present even in the absence of SST bias.

The diagnostics of the structure of the temperature bias in the tropical Pacific together with the wind bias suggest that at least a part of the cold bias in the region results from a positive coupled feedback between winds and SST: too strong winds lead to an excess of upwelling, producing colder SST, which

in turn produces stronger zonal winds. Disrupting this positive feedback is an important motivation for using momentum-flux correction in this study (see below).

3.1 Full initialisation

In numerical weather prediction, the normal procedure is to initialise the model from an analysis performed via data assimilation. The analysis is a combination of the latest observations together with a short-range forecast. By continuously using the information from the observations the analysis state is kept close to the real-world attractor (although in poorly observed areas, a difference could still be present). During the model integration, the state of the model will diverge from the true state both due to the loss of predictability and the development of systematic errors. In NWP the systematic error is assumed small compared to non-systematic errors and often the model output is not calibrated. At longer lead times (monthly, seasonal and decadal time-scales), the model will drift away from the real-world attractor towards its own attractor. In these cases the model bias cannot be neglected and the strategy for accounting for the model bias is the a-posteriori removal of it.

The bias is corrected by applying a lead-time dependent bias correction when post-processing. The bias correction made is also dependent on the seasonal cycle. This is the strategy commonly used in monthly and seasonal forecasts (Stockdale 1997). For example, in an operational seasonal forecast issued every month with a typical lead-time of 7 months, the estimation of $7 \times 12 = 84$ bias correction terms is needed to account for all lead times and all starting dates. The robust estimation of this large number of bias fields requires a large data set of hindcasts (retrospective forecasts). This strategy will fail if the bias is non-stationary, and can lead to sub-optimal forecast skill. The non-stationarity of the bias may be due to non-stationary errors on the initial conditions (Kumar et al. 2012), or to flow-dependent bias arising from the non-linear nature of the system (Balmaseda et al. 2009). Generally speaking, if the systematic error bias is large enough, the non-linear terms will become non-negligible and therefore a mere linear calibration process will be insufficient.

The full initialisation strategy may also give rise to so-called initialisation shock, a term referring to the rapid adjustment processes in the initial phases of the forecast, that can produce non-monotonic behaviour in the model drift. The consequence of the initialisation shock is that at short lead times the error can be larger than at longer lead times. The cause of the initialisation shock is an imbalance between the initial conditions and the dynamics of the model.

In this paper the experimentation using full initialisation but without any flux correction will be referred to as *FullIni*.

3.2 Anomaly initialisation

Due to the difference in mean climate of the analyses and the model, a forecast initialised from an analysis will drift towards the model climate. The drift towards the model attractor could result in an initialisation shock or overshooting of a model drift further away than the model climate, before stabilising at the model climate. The idea of using anomaly initialisation is to avoid the model drift and initialisation shock (but not the model error). The procedure of using anomaly initialisation is to calculate anomalies in the analysis with respect to the analysis' climatology and add such anomalies to the climate of the model. The method has previously been used in several studies e.g. Schneider et al. (1999), Pierce et al. (2004) and Smith et al. (2007). The rationale is to avoid an initialisation shock due to an initial state being far from the model attractor. Strategies for initialising imperfect models are also discussed in Toth and Peña (2007) in a simple model framework.

It is often emphasised that the advantage of the anomaly initialisation is the avoidance of initialisation shock. This is by no means guaranteed, since the structure of the observed anomaly may not be consistent with the model mean state. For instance, the largest anomalies in the observations are associated with displacement of sharp fronts or gradients. If, for example, the systematic error of the model is highly correlated with the misplacement of these fronts/gradients, simply adding an anomaly where it is not supposed to be found is not the same that initialising the model around its attractor. A sharp inconsistency is found also when the placement of the anomalies is associated with vertical displacements of the Equatorial thermocline. Another clear example is the application of an observed sea-ice anomaly in regions where the model never has sea-ice.

A more interesting advantage of the anomaly initialisation, which is often not discussed, is the avoidance of model drift. By avoiding model drift, the a-posteriori correction of the forecast does not require the bias dependence on the forecast lead time (so typically only the 12-month climatology of the bias is required), and the bias estimators can be more robust. This is more relevant for decadal forecast ranges, when it is also more expensive to conduct the calibrating hindcasts. The procedure requires however a long integration to estimate the model climatology.

The procedure of the anomaly initialisation is not without problems. First of all, if the non-linearities are strong, the calibration of the forecasts will face the same problems as the full initialisation, or even stronger, since the mean error is fully developed during the coupled model integrations. It is not guaranteed that the initialisation shock is removed, since it depends on how the anomaly is assimilated into the model (as discussed above). The other

drawback is that the estimation of the anomaly requires the knowledge of the observed climatology. This introduces two kind of difficulties. On the one hand, it is important that the sampling period used for the observed climatology is consistent with that used for the model climatology (for instance, a model climatology estimated for the pre-industrial era should not be used for the anomaly initialisation of decadal forecast post 1960's, with an observed climatology estimated during the period 1970–2005). The other kind of problem is related with defining the climatology of new or sporadic observations. For instance, some regions like the Southern Ocean had not been observed prior to advent of Argo (Roemmich and Owens 2000). Most of the deep ocean has only been observed sporadically with cruise data, and there is not enough information to extract a long term climatology. To avoid this problem, the anomaly initialisation strategy of the ocean often uses gridded fields from existing ocean reanalysis. In this way, it turns an initial weakness into a good advantage, since it means that different coupled modelling groups can initialise their decadal forecasts with external ocean reanalysis, without the need of having to develop data assimilation systems for their own models.

In our experiments, the anomalies are added to the full model state vector instantaneously at initial time, instead of assimilating only temperature and salinity anomalies with a certain time-scale, usually by means of relaxation techniques, as in e.g. Pohlmann et al. (2009). One potential problem with using the instantaneous full state vector is that the new state could also not lie on the model attractor, creating instabilities in the adjustment process that may lead to the quick disappearance of the anomalies. But none of the results from these experiments suggest that this might be the case.

To obtain a model climate for the anomaly initialisation procedure, the Control experiment was conducted, consisting of a set of 25-year forecasts from 3 different initial dates and 3-ensemble member each (i.e. a total of 6 25-year forecasts initialised in 1965 and 1975 and 3 23-year forecasts initialised in 1985). The ensemble mean of these Control integrations was used to compute the model climate. The first 10 years of the simulations have been discarded in order to let the model drift to its own climatology. This may still be a short period for the drift in the deep ocean to be fully developed, but it is sufficient for obtaining a stable state in the upper ocean, which is the prime objective of this study. An observed ocean climatology has been calculated from ocean reanalysis, spanning the same time period used in the estimation of the model climate. Using the same period yields data sets with the same impact of greenhouse gases. The model and analysis climate is calculated for the actual day of the year used for the initialisation (November 1st in our case).

The forecasts using anomaly initialisation will be referred as *AnoIni*.

3.3 Flux correction

It is clear from a variety of studies that strong non-linear interactions (e.g the relationship between SST and atmospheric convection) between mean state and anomaly are at play in the coupled model forecasts. For instance, Balmaseda et al. (2010a) show that the atmospheric response to a given sea-ice anomaly depends on the atmospheric mean state, which in turn is conditioned by the underlying SST. In their study they show that correcting the SST in the North-Atlantic, where the model has errors due to the wrong position of the Gulf Stream, has large impact on the atmospheric mean state and in the atmospheric response to the Arctic sea-ice anomaly. Results along these lines are documented by Scaife et al. (2011) and Keeley et al. (2012).

Model improvement is the ultimate way of reducing model biases. However this is a slow process, especially if the systematic errors are related to model resolution (as in the case of the correct Gulf Stream). A temporary solution, until the problems in the model are detected and solved, is to compensate for the systematic errors by applying empirical corrections.

One specific correction is the so-called flux correction, applied only in the coupling between the atmosphere and the ocean. The aim of the strategy is to avoid (or limit) the model drift by adding a correction term to the model during the simulation that pushes the model solution towards the observed climatology. The aim is to keep the forecast errors as linear as possible, to be able to apply simple calibration techniques. In this strategy the empirical correction of the forecast is done during the model integration rather than only in the final calibration phase. Ideal candidates for this strategy are those situations where the model exhibits a very clear mean error, difficult to correct by improving the model formulation but relatively easily by applying empirical correction. In these cases flux correction may be a successful forecast strategy. The use of flux correction has recently been discussed in Spencer et al. (2007) and Manganello and Huang (2009).

In this study we will investigate two options, namely of using only momentum-flux correction (*Ucorr*) and a combination of momentum and heat-flux correction (*UHcorr*). The flux correction has been calculated in two steps. First, the strong SST relaxation simulations initialised in 1965 and 1975 have been used in order to calculate the wind stress errors when the SST is constrained to the reanalysis (see Fig. 3b). The momentum-flux correction has been calculated as the difference between the Strong-Relax forecasts and the surface stresses from the reanalysis.

The first 5 years of each SST-relaxation simulation have not been used in order to let the atmospheric model drift. In a second step, a set of 25-year forecasts (WeakRelax) has been run applying the momentum-flux correction and a weak SST-relaxation (40 W/K), in order to calculate the heat-flux correction required to correct the coupled SST when the momentum-flux correction applied. The heat-flux correction is calculated from the SST relaxation coefficients. This strategy yields a heat-flux correction suitable to be used together with momentum-flux correction and that partly accounts for the feedback effects.

The corrections are added to the fields passed from the atmosphere to the ocean: the wind stress seen by the ocean has been modified by the momentum-flux correction, and the heat-correction has been applied to the non-solar component of the heat flux received by the ocean.

As examples of the applied flux corrections, Fig. 4 shows the correction for zonal wind stress component (upper panels) and the heat flux (lower panels). The corrections applied have a seasonal variability and here the January and September values are plotted. These months have been selected because the heat-flux correction has its tropical minimum in January and its tropical maximum in September, and the momentum-flux correction is close to the maximum (August) and minimum (April). For the full seasonal cycle in the tropics, see Magnusson et al. (2012b). Comparing the momentum-flux corrections (upper panels) for January and September we see a strong seasonality, especially in the tropical Pacific where the required correction is strongest in September in order to decrease the too strong easterlies in the model climate. We also see that the corrections are strongest in the winter hemispheres.

The lower panels in Fig. 4 shows the heat-flux correction required with the momentum-flux correction applied, for January (Fig. 4c) and September (Fig. 4d). Positive correction means that heat is added to the ocean and the global annual mean is 3.9 W/m^2 . Regionally, positive corrections are needed for the winter hemispheres in the subtropics. Large heat-flux corrections are also needed in the vicinity of the western boundary currents in both the Atlantic and the Pacific. In the Southern Ocean a negative flux correction is needed due to the warm bias in the model. Regarding the tropical Pacific, the heat-flux correction is weak in January, while there is a strong positive correction during September, especially in the eastern part. Close the south-American coast the correction is negative for both January and September.

4 Model drift

In the previous section, the model climate and its systematic errors were introduced. In this section we will discuss

the evolution of the systematic errors as a function of forecast lead-time, usually referred to as the model drift. We will discuss the global temperature drift as well as focusing on the drift in the tropical Pacific.

Figure 5a shows the evolution of the global mean SST for the reanalysis and forecast data from 4 initial dates based on 3 ensemble members, using a 12-months running mean filter. For the reanalysis SSTs, a warming trend of about 0.5 Kelvin during the 40 year period is apparent. The trend is modulated by interannual variability, dominated by the ENSO events (the El Niño event in 1997 caused the highest global SSTs in the time-series). In Fig. 5b the bias averaged over all initial dates is plotted, in order to visualise the mean model drift.

For the *FullIni* experiment (red), the SST drifts towards colder values as expected from the cold bias for the model climate. The drift has different time-scales; most of the drift appears during the first two years with a slower drift acting on the decadal time-scale. For the *AnoIni* experiment (pink), a substantial part of the model drift is avoided by initialising the model close to its attractor, which is the primary aim of the strategy. The *AnoIni* experiment has a slightly colder mean state than the *FullIni* even after 8 years, which is a sign of a drift acting on long time-scales.

The results for the *Ucorr* (green) experiment indicate a delayed drift compared to *FullIni*, which can be related to the tropical Pacific (see below). For the longest forecast range, the SST bias is comparable to *FullIni*. The *UHcorr* experiment shows a much improved model climate, which is an expected result from using heat-flux correction. The small difference between the reanalysis and the *UHcorr* experiment is believed to be due to the differences in the SST data set used to calculate the flux correction and the SST in the reanalysis, but it can also be an artifact of sampling errors and non-stationarity of the record.

In Fig. 5c, the ensemble mean of bias-corrected forecasts is plotted, using a lead-time dependent correction similar to the bias in Fig. 5b. After applying the bias correction, it is difficult from this figure to determine whether any forecast method is better or not. The forecast quality will be discussed in Sect. 5.

In order to further exploit the 3-dimensional development of the temperature bias in the ocean, vertical cross-sections of the global bias development are plotted in Fig. 6 as a function of lead time and depth of the ocean. The development of the integrated heat content in the upper 700 m is plotted in Fig. 7a for 8 forecasts and the development of the bias in Fig. 7b.

Studying the results for *FullIni* (Fig. 6a) and *AnoIni* (Fig. 6b), the general features are a cold bias in the upper-ocean and a warm bias at depth. For the *FullIni* experiment, the cold bias appears on a much faster time-scale than the

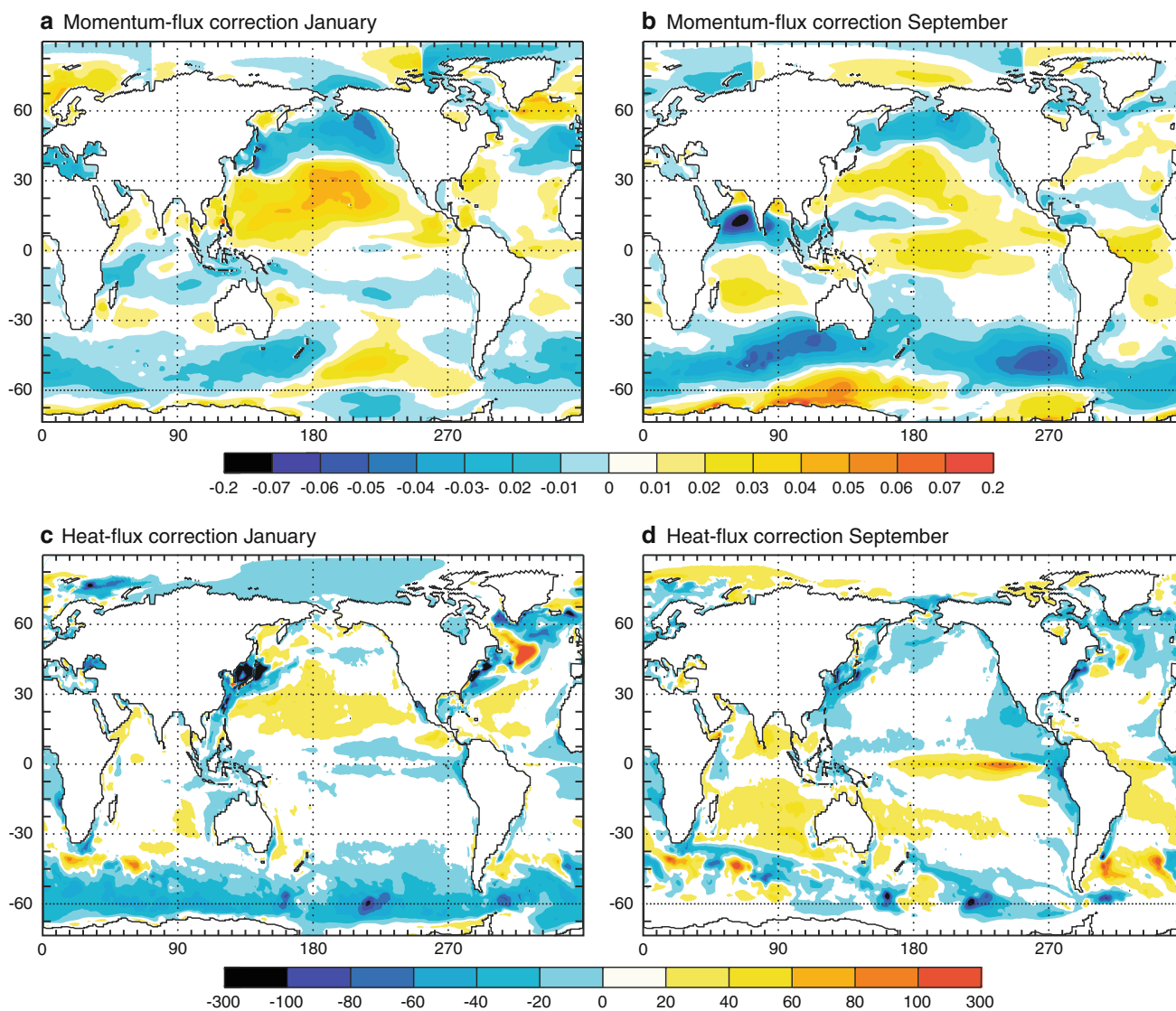


Fig. 4 Examples of flux corrections. The units are N/m^2 (top panels) and W/m^2 (bottom panels)

warm bias at deeper levels. Magnusson et al. (2012a) argue that the fast development of the surface bias is believed to be related to the initial imbalance in the radiation at the top of the atmosphere together with a fast development of the enhanced cold tongue in the Tropical Pacific ocean surface. The slow component of the drift could be due to a too strong vertical mixing in the ocean. The location for the warm bias is generally in the extra-tropics. Because of the different time-scales of the cold and the warm bias, the integrated heat-content decreases during the first years of integration for *FullInI*, while the bias for *AnoInI* is much more stable.

For the *AnoInI* experiment, as expected, the initial conditions already show the bias structure of the coupled model, but the warm bias continues to develop. One possible explanation is that the model climate has been

calculated before the coupled model reached equilibrium, and therefore the drift in the deep ocean continues after 25 years, which would be consistent with the slow time-scale of errors associated with vertical mixing in the ocean interior. Another possible explanation is the non-stationarity of the model error during the period considered: it may be possible that the response in the model to greenhouse radiative forcing is too strong, which, combined with the errors in the ocean vertical mixing, results in too strong warming of the deep ocean in the model. This would explain why the *AnoInI* decadal simulations exhibit a correct warming trend in SST, while the trend in the 700-m integrated heat content is stronger than in the ocean reanalysis (Fig. 7a, pink curve). This error is particularly apparent in the *AnoInI* decadal simulation, where at initial time the cold bias in the upper ocean is already developed.

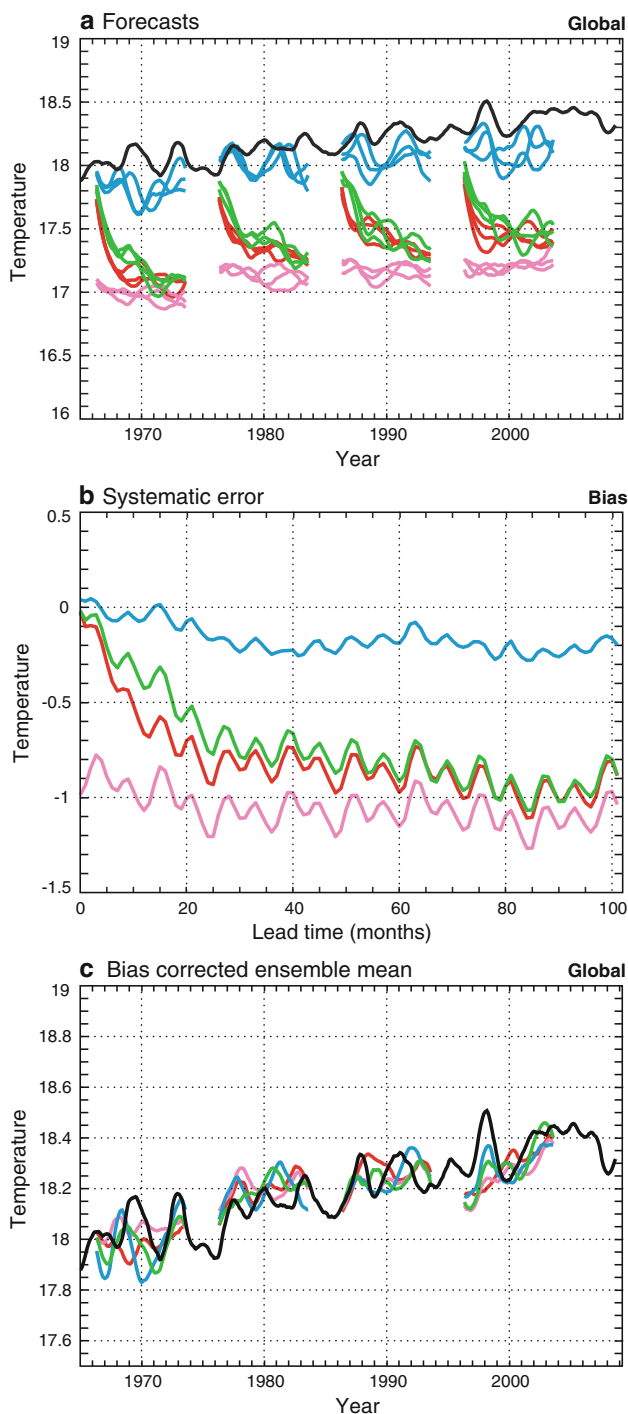


Fig. 5 Time-series of the global mean sea-surface temperature and its systematic error. *FullInI* (red), *Ucorr* (green), *UHcorr* (blue), *AnolInI* (pink) and reanalysis (black). 12-month running mean applied for the forecasts. The evaluation is based on 3 ensemble members

A third explanation is the poor available sample of decadal forecast start dates, insufficient for a robust estimation of model drift, in particular in the presence of outliers such as the failure to predict the strong post-Pinatubo cooling of the ocean apparent in the ocean reanalysis. This third

possibility should not be discarded, especially since the verifying ocean reanalysis and radiative forcing can also be in error.

For the *Ucorr* experiment, the cold bias at the surface is similar in magnitude to *FullInI*, while the warm bias below is somewhat weaker. This makes the integrated heat content even colder than for the *FullInI* experiment, which has a stronger compensation of errors. Comparing *FullInI* and *Ucorr*, the impact of the momentum-flux correction is mainly in the Pacific (not shown). By changing the speed of the gyre, the *Ucorr* experiment has reduced biases at this depth, especially around the western boundary current in the northern Pacific.

The *UHcorr* experiment shows a reduced bias at the surface as expected from the flux correction design. However, the *UHcorr* experiment also shows a weaker warm bias at depth compared with *FullInI*, while a simple linear argument would have suggested the opposite since the heat-flux correction is adding heat to the ocean. This is a clear example of the non-linear processes at play in the coupled system. A warmer ocean surface would result in stronger stratification, then reduced mixing and less heat uptake. There is also the possibility that as a result of the warmer surface, more heat is released back into the atmosphere.

Figure 7c shows the forecast time-series of the 700-m heat content for the ensemble mean with a bias correction applied. Studying the bias-corrected forecasts (Fig. 7c), the general feature is a faster warming in the latter part of the time-series than in the early part. However, all forecasts miss the cooling of the oceans in 1992 and the fast warming started in year 2000.

The systematic model errors in the Tropical Pacific and its effect on the variability was extensively discussed in Magnusson et al. (2012b). The main finding was that the ENSO variability is clearly suppressed by the systematic errors in the model and by using flux correction the variability was increased. In order to evaluate ENSO forecasts, the average SST for different areas are commonly used. In this study we will refer to Niño3 (150°W–90°W, 5°N–5°S) in the eastern part of tropical Pacific; Niño3.4 (170°W–120°W, 5°N–5°S) in the central part and Niño4 (160°E–150°W, 5°N–5°S) in the central-western part of the basin.

Figure 8a and b shows the SST model drift as a function of lead time in the Niño3 and Niño4 area respectively, calculated from the 14-months simulations, consisting of 10 ensemble members. For these figures, it is clear that *AnolInI* experiment is initialised on the biased state (by design), while the other experiments are initialised on the observed (analysed) state. Studying the *FullInI* experiment, the drift starts immediately and a major part of the drift takes place in the first 6 months. For Niño3, one outstanding feature is that the *FullInI* experiment drifts even

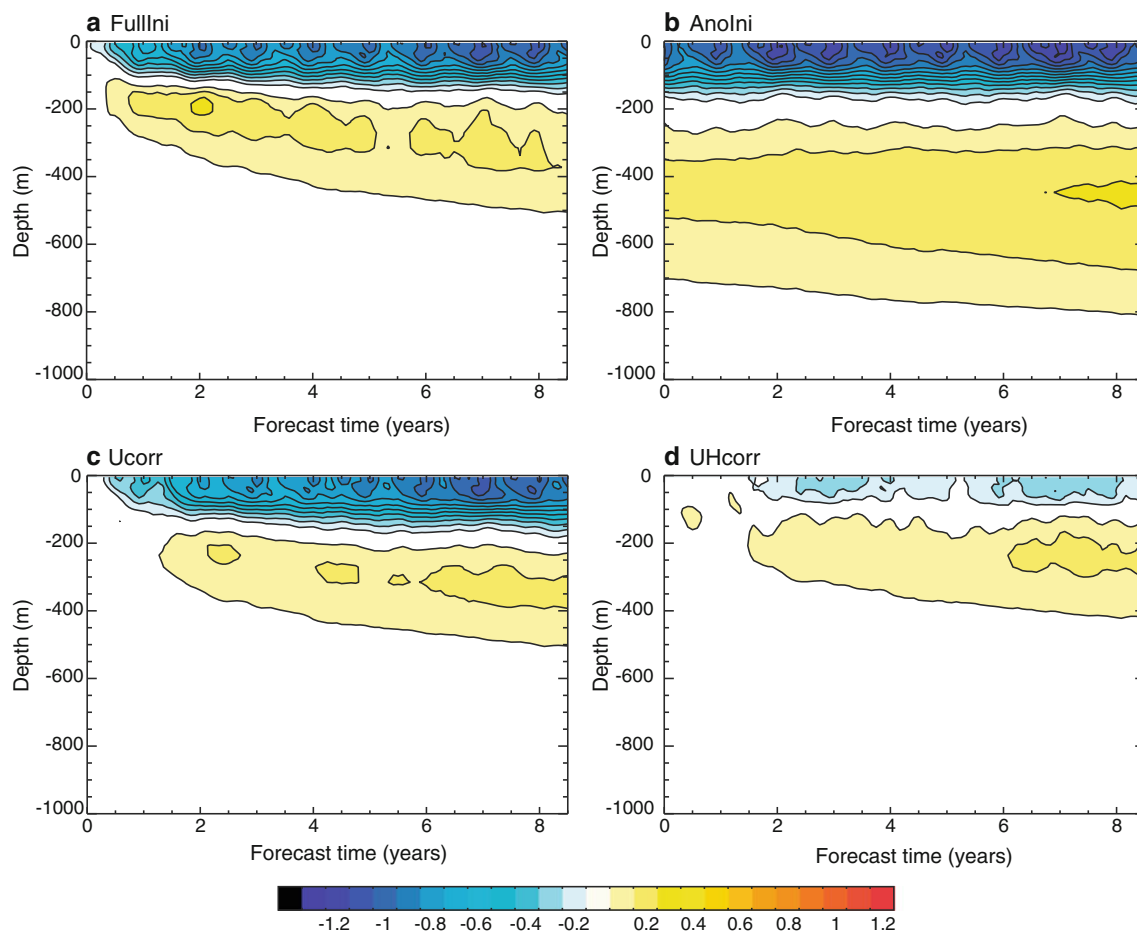


Fig. 6 Vertical cross-section of the development of the global temperature bias. The evaluation is based on 3 ensemble members. Contour line intervals 0.1 K

colder than for the *Anolni* experiment, which is a sign of "over-shooting" (the systematic error is larger than the error of the model climate during a transient period). The over-shooting is believed to be connected to the different time-scales for the development of the surface cold bias (fast) and the sub-surface warm bias (slow). Once the waters at thermocline level are warmer, the cooling contribution of the upwelling will be reduced, and so will be the amplitude of the surface cold bias.

For the *Ucorr* experiment, the bias during the first year is clearly reduced, compared to *FullIni*. The reduction is clearest in the eastern part where the applied momentum-flux correction is due to the change in the amount of upwelling cold water. In the western part, the additional heat-flux correction plays a role as *Ucorr* develops a stronger bias than *UHcorr*.

In order to evaluate the model drift on a longer time-scales, the model bias as a function of lead time has been calculated from the decadal simulations, using 3 ensemble members. Figure 8c and d shows the model drift for the first 5 years of these simulations for Niño3 and Niño4

respectively. Comparing with the 14-month simulations, the data from the decadal simulations is much noisier, due to a reduced number of initial dates and ensemble members. However, the same features stand out as in the 14-month data sets. The overshooting for *FullIni* for Niño3 is also present here and remains into the second year. For longer forecast ranges (year 2 and onwards), the sub-surface bias structure has developed in the *FullIni* experiment and the period of over-shooting is over. After that the *Anolni* is somewhat colder than *FullIni* due to the generally colder ocean as seen in Fig. 5b.

For *Ucorr* and *UHcorr* the positive bias after one year is much more pronounced in the decadal simulations. For the *UHcorr* experiment the bias is about 2 Kelvin. For the following year a corresponding cold bias is present. One could speculate that it could be a sign that the flux correction is over-doing the correction for the first year, which can not be discarded since the flux correction was estimated from a long coupled integration (i.e., with errors fully developed). It can be also an artifact of poor sampling, the results being affected by some outliers (false alarms for

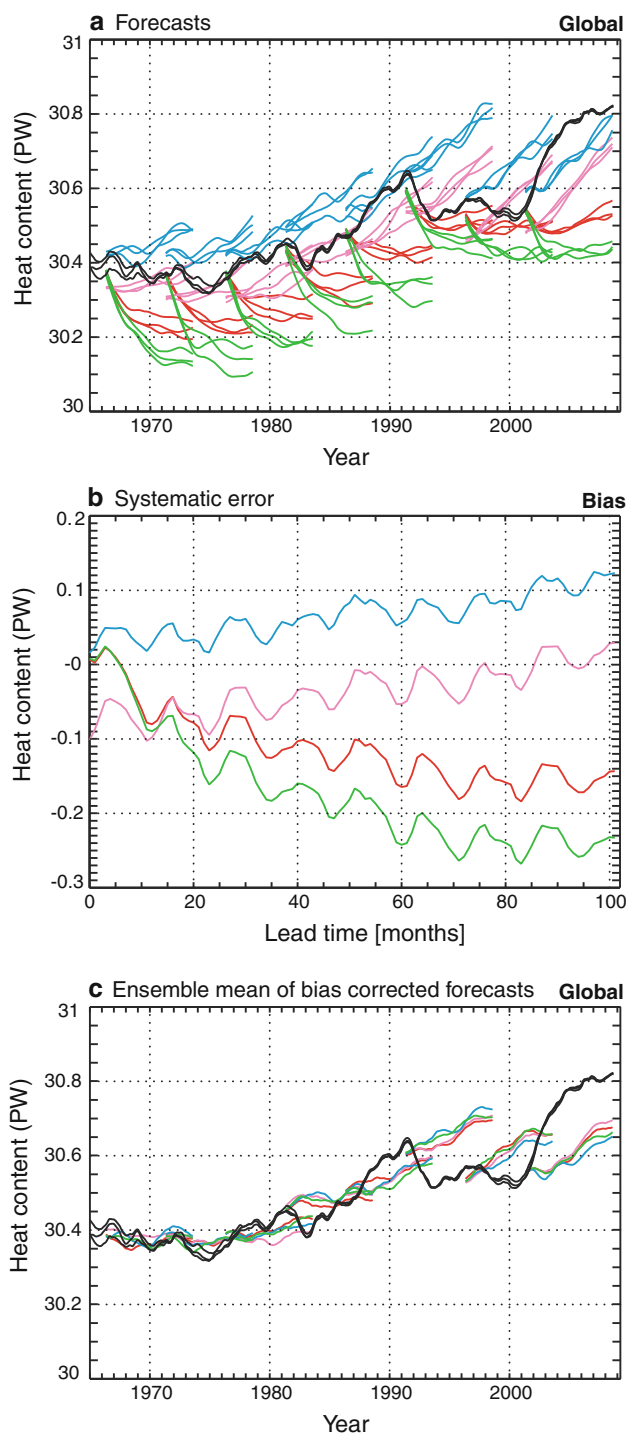


Fig. 7 Time-series of the global mean ocean heat content in the upper 700 m and its systematic error. *Fullni* (red), *Ucorr* (green), *UHcorr* (blue), *Anolni* (pink) and reanalysis (black). 12-month running mean applied for the forecasts. The evaluation is based on 3 ensemble members

El Niño event or a missed La Niña). For longer lead times the mean state for *Ucorr* is colder than *UHcorr* but still not as biased as *Fullni*.

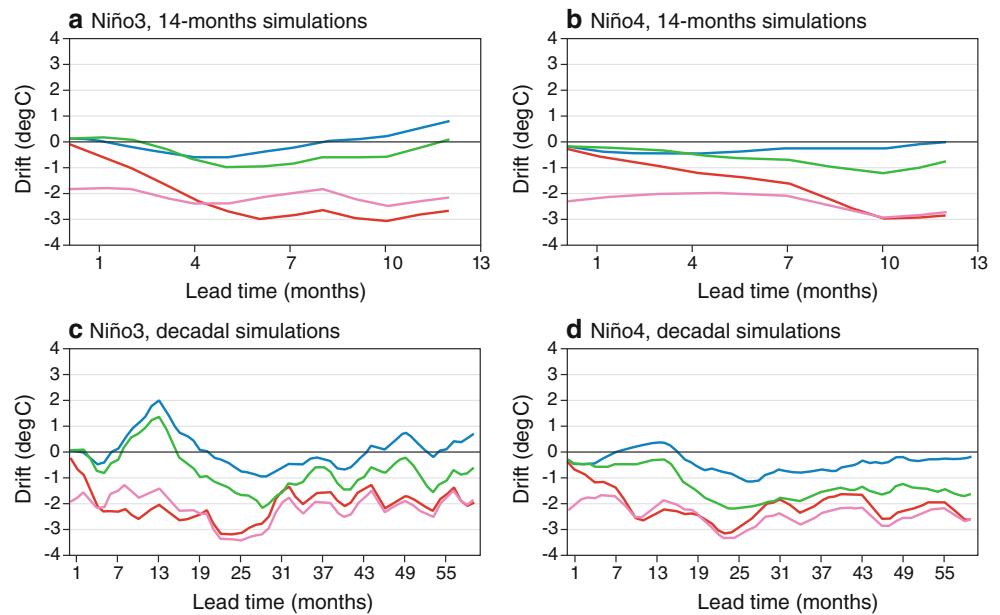
5 Forecast quality

In this section, the forecast quality on seasonal and decadal time-scales will be discussed, starting with the predictability of ENSO on seasonal time-scales. The predictability of ENSO is a key for the predictability of other regions due to ENSO’s strong teleconnections. For all scores presented here the bias has been removed by applying a lead-time dependent bias correction. For the decadal time-scale, results from *Ucorr* will not be discussed due to only 3 available ensemble members. In the evaluation of the decadal experiments, the data has been detrended hence the main interest in this study is the effect of the forecast strategy and not the response of increased greenhouse gases.

Figure 9a, b and c shows the anomaly correlation coefficient (ACC) for the Niño3, Niño3.4 and Niño4 respectively. For Niño4, *Ucorr* shows the best scores while *Anolni* clearly shows the worst results; even worse than the persistence. This is an example of forecast improvement obtained by correcting the mean state and disrupting the development of positive feedbacks between cold tongue and trade winds [discussed in e.g Neelin and Dijkstra (1995)], which are strongest in this area (Niño4 is the Central-Western Pacific). For Niño3 (in the Eastern part of the Pacific basin) the flux-corrected experiments show most noticeable advantage compared to the other forecast strategies in the forecast range 3–5 months. The score for *Fullni* is slightly better than *Anolni*. This is a region affected by remote forcing via propagation of Kelvin waves generated in the Western Pacific. So the improvement seen here is likely a combination of local improvements (better thermocline depth) as well as a response to remote improvements in the western part of the basin. The Niño3.4 region, located in the middle of the basin, is a combination of a part of Niño3 and Niño4. Therefore it is not a surprise that this area shares some features with the other areas. In this area *Anolni* is worst but is still better than the persistence.

Figure 10 shows the amplitude ratio of anomalies compared to the observed amplitude. The outstanding feature here is the period of large amplitude in the *Fullni* experiment, which peaks roughly at the same time as the main drift towards cold bias (month 5 of the forecasts, see Fig. 8). The coupled model fails to capture the seasonal relaxation of the trade winds, which in nature happens during late spring / early summer. During this time of the year, the amplitude of the observed SST anomalies in the Eastern Pacific is at its lowest, since the SST variability is somewhat decoupled from the thermocline variability. Any anomaly in the initial conditions remains in the thermocline, but it does not translate into an SST anomaly. This is also the reason for the seasonal predictability barrier. In the

Fig. 8 Model drift in SST in the tropical Pacific. *FullIni* (red), *Ucorr* (green), *UHcorr* (blue) and *AnoIni* (pink). The evaluation is based on 10 ensemble members



model however, the trade winds fail to relax, and the thermocline-SST feedback continues to be active during the spring-summer season, giving rise to an overestimation of the SST anomaly. This problem in the model will be more obvious in situations when the anomaly in the initial conditions is large. We are in a situation that illustrates the difficulties of dealing with flow-dependent, seasonally dependent and lead-time dependent bias. The amplitude of the *FullIni* SST anomalies decays at longer forecast ranges, and by month 10 they are weaker than the observed anomalies. For decadal forecasts the ENSO amplitude is clearly damped due to strong biases in SST and zonal winds (see Magnusson et al. (2012b) for further discussion).

For *AnoIni*, which has only a small model drift, the amplitude ratio is close to 1. This shows that the strategy implemented here for anomaly initialisation is able to retain the information for about 6–7 months, after which the amplitude of the anomaly converges to the same values as seen for *FullIni*, and probably for the same reasons (a too cold mean state favouring weak ENSO variability). These results indicate that even with anomaly initialisation the features of the coupled model depend on the forecast lead time: although the mean state is relatively stable, and therefore there is not drift in the mean, there is a drift in the variability, which is larger at the initial time (when the anomalies are from the “real-world”), since the model attractor will not contain the same sort of anomalies.

For *Ucorr* and *UHcorr*, an overestimation of the amplitude is present around forecast month 6, but not as strong as in the *FullIni*: the seasonal momentum-flux correction will enforce the trade wind relaxation. For *UHcorr*, the amplitude ratio increases after one year. This manifests

itself in the overprediction of El Niño events at forecast ranges longer than 1 year.

Figure 11 shows the ACC for the precipitation in the Niño3.4 area. Here we see large differences between forecast strategies. The worst results by far are obtained by the *AnoIni* experiment, which has an erroneous mean state throughout the simulation. Even if *AnoIni* produces skillful ENSO forecasts of SST after bias correction (better than persistence), the precipitation forecasts are poor, substantially worse than persistence until month 5. The precipitation rate is dependent on the absolute value of the SST and therefore the precipitation will be negatively affected by the cold SST bias. This is an example of the non-linear relation between SST and atmospheric convection. The best performance is shown by the *Ucorr* strategy, followed by *UHcorr*. One possible reason for this improvement is the better prediction of SST, but also the improved mean state of the atmosphere, which may respond more effectively to a given anomaly. The *FullIni* experiments show reasonable results for the first few months, when the mean state of the SST still is close to the observed one.

Figure 12 shows a summary of the ACC for different areas for 2-m temperature, for various lead times and averaging periods. All diagnostics are based on 7 ensemble members. For forecast year 2–5, maps for ACC is shown in Fig. 13. Here statistically significant points are marked with a dot (different from zero with a 5 % significance level).

For global temperatures (Fig. 12a), skill is present throughout the first year. This is mainly due to persistence of initialised anomalies and prediction of ENSO, which has an influence on the global mean temperature. For longer time-scales the effect of increased greenhouse gases plays

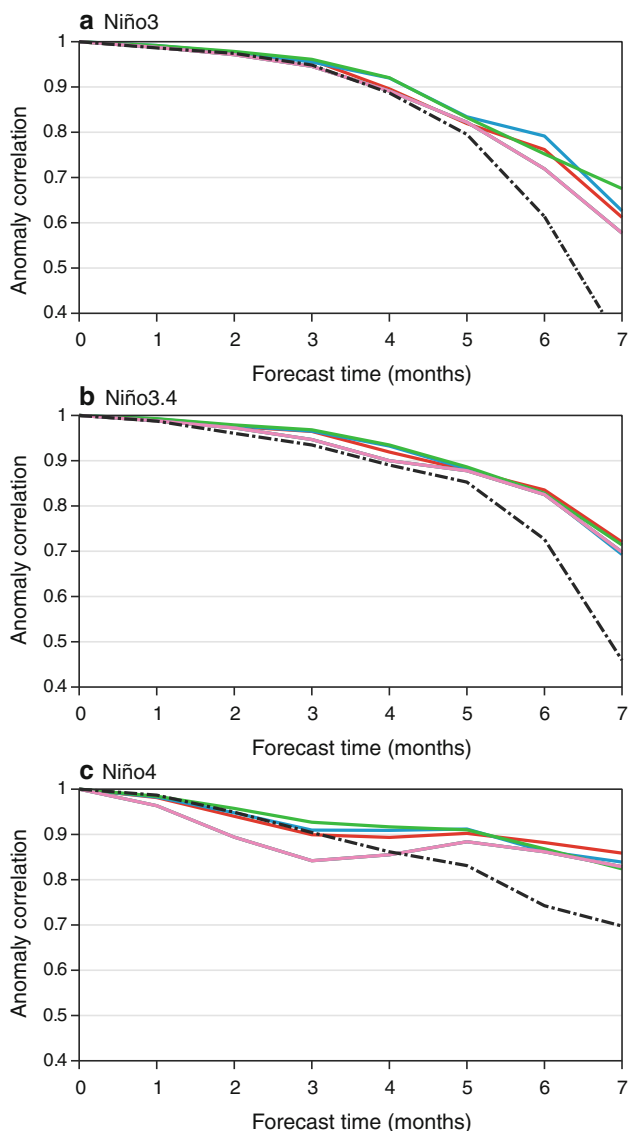


Fig. 9 Anomaly correction coefficient for SST forecasts in El Niño areas. *FullInI* (red), *Ucorr* (green), *UHcorr* (blue), *AnolNi* (pink) and persistence (black). The evaluation is based on 10 ensemble members

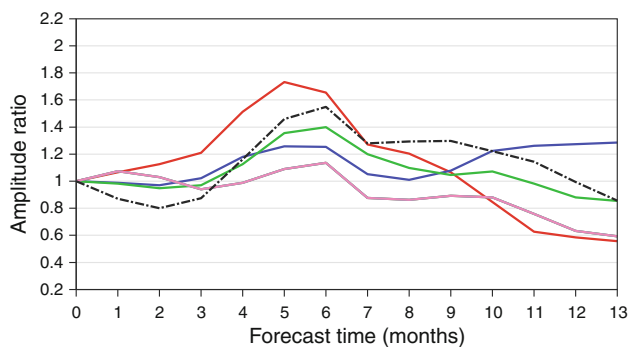


Fig. 10 Amplitude ratio for SST anomalies compared with observed anomalies in Niño3 area. *FullInI* (red), *AnolNi* (pink), *Ucorr* (green) and *UHcorr* (blue). Persisted forecast in (black). The evaluation is based on 10 ensemble members

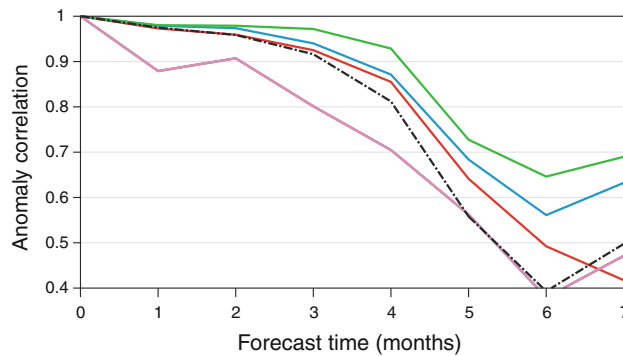


Fig. 11 Anomaly correlation coefficient for precipitation in the Niño3.4 area. *FullInI* (red), *AnolNi* (pink), *Ucorr* (green) and *UHcorr* (blue). Persisted forecast in (black). The evaluation is based on 10 ensemble members

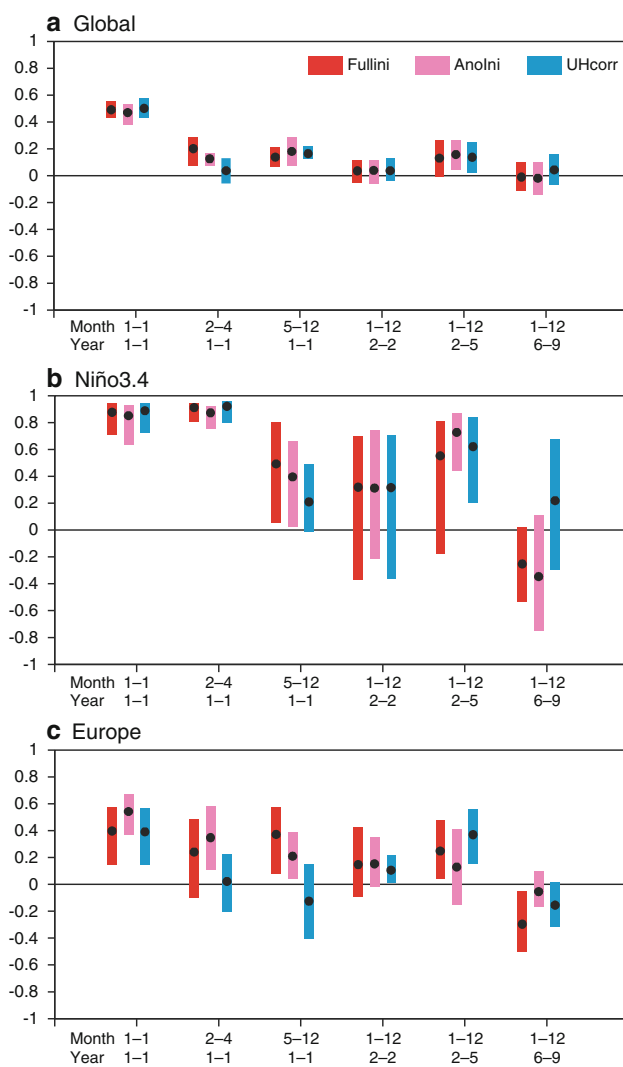
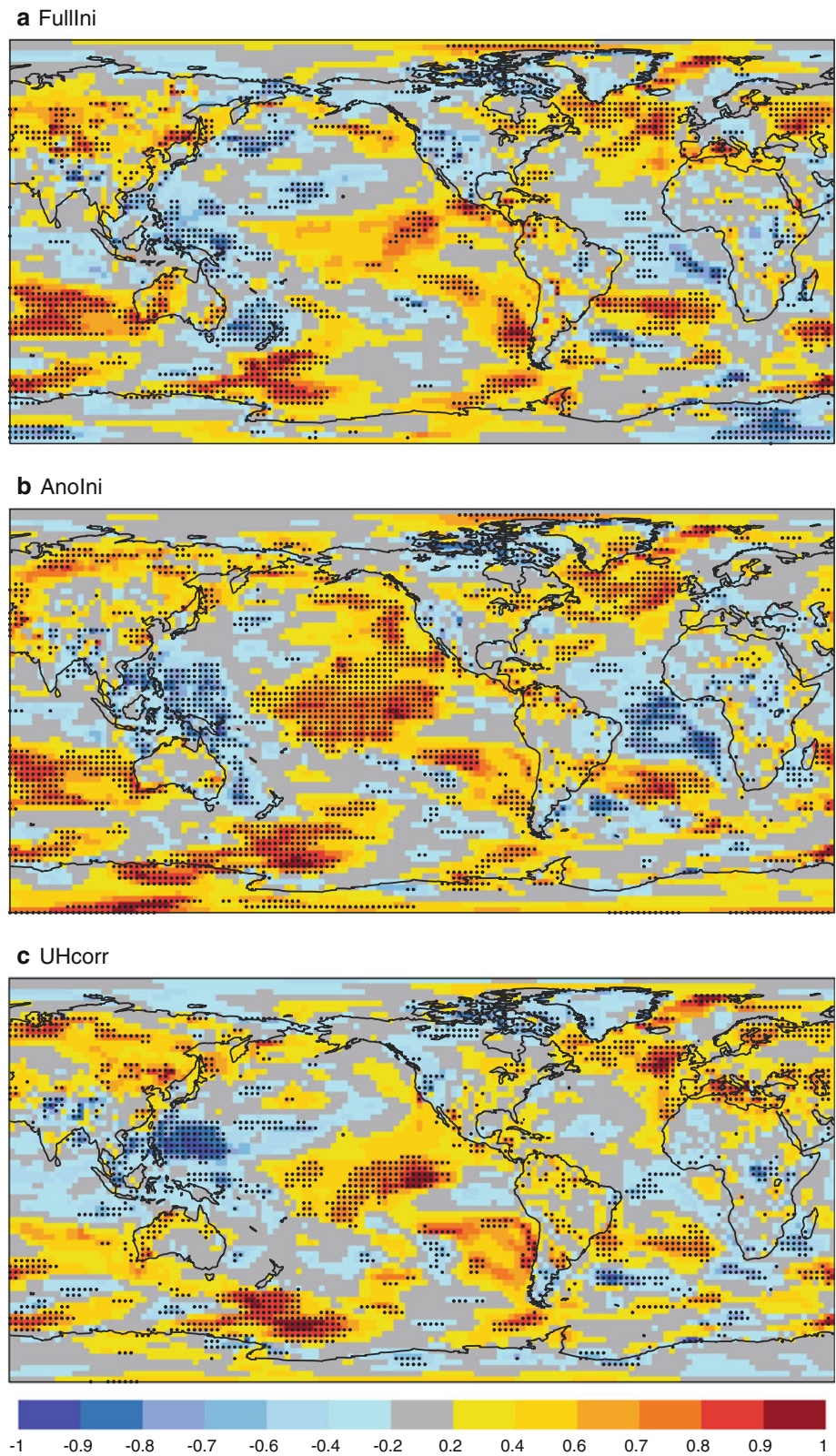


Fig. 12 ACC for 2-m temperature averaged over different periods (from left): forecast year 1, month 1; year 1, month 2–4; year 1, month 5–12; year 2, month 1–12; year 2–5, month 1–12; year 6–9, month 1–12. *FullInI* (red), *AnolNi* (pink) and *UHcorr* (blue). The evaluation is based on 7 ensemble members

Fig. 13 ACC for 2-m temperature, forecast month 1–12, year 2–5. *Black dots* for values significantly different from zero with 95 % confidence. The evaluation is based on 7 ensemble members



an important role (Hawkins and Sutton 2009), and has been removed by linear detrending in the calculation of these scores.

The region with the highest predictability for month 2–4 is the tropical Pacific that benefits from the predictability of ENSO (see above). For the second half of the first year,

UHcorr shows the worst performance for Niño3.4 (Fig. 12b). This could be related to this strategy producing too much ENSO variability at longer lead times, as indicated by a model drift and a high amplitude ratio (see above).

Figure 13 shows the ACC for forecasts of 2-m temperature on a multiyear time-scale averaged over the forecast period year 2–5. This could be seen as a measure of skill of decadal forecasts, even if the period is only four years. One area with enhanced predictability on this time-scale is the North Atlantic, in agreement with other studies (e.g. Pohlmann et al. 2004; Mochizuki et al. 2010; van Oldenborgh et al. 2012). Although the figures show that the *UHcorr* and *FullIni* perform somewhat better than *AnoIni*, it would be dangerous to draw conclusions without a better understanding of the causes. One could speculate that this is due to a more realistic strength of the MOC on this time-scale, but in general the MOC is not well represented in the coupled model, nor in the initial conditions. Some skill is surprisingly present in the southern Indian Ocean, where other coupled models show little skill after the linear detrending (e.g. van Oldenborgh et al. 2012). In fact, it seems that the positive skill extends to the south-eastern Pacific and southern Atlantic. These results would suggest that the higher skill at 2–5 years is not exclusive to the North Atlantic, but more generally of the mid latitudes. The decadal predictability in the Indian Ocean is further discussed in Guemas et al. (2012). Some skill for year 2–5 is also present in the Tropical Pacific for all experiments (the highest skill for *AnoIni*); the pattern of the positive skill (latitudinally broad horseshoe, with little amplitude in the Eastern Pacific) is reminiscent of the decadal ENSO signal (Power et al. 1999, 2006).

For year 6–9, the *UHcorr* experiment seems to have higher skill in the Niño3.4 area than *FullIni* and *AnoIni*. It is difficult to judge whether this signal is robust and a deeper investigation is needed to find the reason for this increased predictability.

6 Summary and discussion

In the presence of systematic model errors, a forecast strategy needs to be applied to deal with bias. This study discusses different forecasts strategies, and compares the results when applied to seasonal and decadal forecasting. The standard forecast strategy for numerical weather prediction, monthly and seasonal forecasts is full initialisation, where the model is initialised from a state close to the real-world attractor. In the presence of systematic model errors, the model, once initialised, will drift towards the model attractor. For short lead times, as for medium-range forecasts, the systematic error is usually ignored, since it is

considered small compared to the error growth of the initial conditions. However, at longer lead times (monthly and seasonal) this is no longer the case. If the difference between the model and the real-world attractor is large, a different forecast strategy is needed: the direct model output needs to be calibrated in order to issue the forecasts. Initialisation and calibration can be considered two different aspects of the forecast strategy. In this study we assume the simplest calibration strategy (the removal of the mean bias a posteriori) and compare different forecast strategies, focusing on their effect on the forecasting quality.

We have compared full initialisation, anomaly initialisation, momentum-flux correction and heat and momentum-flux correction. While the full initialisation simulations are initialised close to the real-world attractor, anomaly initialisation aims to initialise the model on its own attractor by attaching observed anomalies to the model climate. The purpose of this strategy is to avoid the model drift and possible non-linear effects of the model drift such as over-shooting and initial shocks. Another strategy to avoid or reduce model drift is to apply flux correction in the coupling between the atmosphere and ocean. This will act as an artificial energy and/or momentum source or sink with a seasonal cycle.

The model system used for this study (ECMWF atmospheric model version 36r1 + NEMO ocean model version 3.0) shows a general cold bias. The bias is due to imbalance in the energy flux at the top of the atmosphere. A part of the atmospheric temperature bias can also be attributed to a strong uptake of heat by the ocean. One sign of this is a warm bias in the oceans below 200 m depth. The coupled model also has an enhanced cold bias in the tropical Pacific caused by too strong easterlies yielding a positive feedback to the Walker circulation (Bjerknes 1969). The cold bias has a strong influence on the ENSO variability, which is discussed for the present model system in Magnusson et al. (2012b).

Comparing the forecast quality for ENSO prediction (SST indices) on seasonal time-scales, the best results are obtained using momentum-flux correction. The worst results are found with the anomaly initialisation, especially in the western part of the tropical Pacific. For the second half of the first year, the worst scores are found for the heat and momentum-flux corrected experiment. This could be related to erroneously triggering of El Niño events, for which we have found evidence by studying the model drift of this experiment.

Comparing precipitation scores for the tropical Pacific, we see a clear disadvantage for the anomaly initialisation. By avoiding the model drift, the mean state for the anomaly initialisation experiment is always in a cold state. The SST bias leads to a strong suppression of convective precipitation. Even with a warm SST anomaly, the SST is too cold

to trigger convection. Here we have a clear advantage for the flux-corrected experiment, which is closer to the observed mean state and has a better precipitation response to warm events (c.f. discussion in Magnusson et al. 2012b).

Looking at decadal time-scales and comparing full initialisation, anomaly initialisation and heat and momentum-flux correction, even with a clear difference in mean climate, the differences in scores are small and uncertain due to a limited set of hindcasts. For the detrended data, some skill is present for the northern Atlantic, tropical Pacific and southern Indian Ocean for year 2–5. However, it is hard to verify any systematic differences between the experiments on this time-scale. This result is in line with the findings in Smith et al. (2012), where full initialisation and anomaly initialisation were compared on the decadal time-scale using a different model system and more initial dates for the hindcast.

All methods investigated in this report have advantages and disadvantages, both with regard to results and from a practical point of view. By using full initialisation the model will drift from the attractor of the analysis to the attractor of the model. During this drift the properties of the forecasts change with lead time, both in the sense of mean and variability. In order to correct for this one needs to apply a lead-time dependent bias correction. The development of the bias might well depend on the state (i.e. be conditional), and optimally one should account for this, although sampling considerations are an obstacle. In this study we found evidence of "over-shooting" model drift in the eastern Pacific at certain forecast ranges.

The anomaly initialisation removes most of the model drift in the global SST and a large part of the model drift in the ocean heat content. The skill in the ENSO forecasts indicates that the anomalies are initialised correctly, although the scores are worse than using full initialisation. By using the anomaly initialisation the model is always in an erroneous state. It severely affects interactions in the climate system, as seen for the ENSO variability in Magnusson et al. (2012b) and the precipitation scores for Niño3.4 in this report. For a practical point of view, long simulations are needed to obtain the model climate. The main problem is, however, the limited period for which the analysed climate can be defined, due to limitations in past ocean observations.

The momentum-flux correction is mainly aimed to compensate for the wind bias in the tropical Pacific, even if the correction is applied globally. For the mean climate and the skill scores for the tropical Pacific we see a clear improvement. This is also documented in Molteni et al. (2011), for a slightly different model configuration. Using the combined heat and momentum-flux correction, the mean climate as well as the variability is well improved. However, it is difficult to see improvement in the scores.

The practical downside of flux correction is the estimation of the corrections. This is not straight-forward, especially for a combination of heat and momentum-flux correction.

In this study we carried out retrospective forecasts (hindcasts), covering the same period that the model climate for anomaly initialisation and the flux-corrections are estimated for. In a world where the climate is changing due to long time-scale variability and/or increase of greenhouse gases this approach is suboptimal. All three methods compared here are affected by the non-stationarities associated to a changing climate. It can be argued that in this assessment the flux-correction and anomaly initialization strategies have not been fully tested in cross-validation mode, since the flux correction, model climate and analysis climate were estimated using data included in the validation period. And even if the bias correction is run in cross-validation mode, a change in climate could affect the bias correction because the bias correction is based on the hindcast period. Therefore all three methods are affected by a non-stationary climate and it is difficult to evaluate which method is most sensitive to a changing climate.

In this study we have not found clear evidence for any universal and easy solution for forecasting in the presence of systematic error. Flux correction (especially momentum-flux correction) has a positive effect on the seasonal time-scale for the model system used here, but this result is highly dependent on the type of systematic errors in the model and may not hold true for other models. For the decadal time-scale only small differences are present between the strategies and no strategy shows a clear advantage. This could be related to the limited sample of starting dates and the limited scope for predictability on these scales, once the effect of increased greenhouse gases has been removed. To conclude, it is still an open question whether the choice of forecast strategy matters for the decadal time-scale.

Acknowledgments We would like to acknowledge Kristian Mogensen for help during this project. We would also like to thank Rob Hine and Anabel Bowen for help with the preparation of the figures for this report. This work was supported by the EU-funded projects COMBINE (FP7/2007-2013) under grant agreement 226520, and THOR (FP7/2007-2013) under grant agreement 212643.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

- Balmaseda M, Ferranti L, Molteni F (2009) Air-sea interaction in seasonal forecasts: some outstanding issues. ECMWF workshop on ocean-atmosphere interaction

- Balmaseda M, Ferranti L, Molteni F, Palmer T (2010a) Impact of 2007 and 2008 Arctic ice anomalies on the atmospheric circulation: implications for long-range predictions. *Q J R Met Soc* 136:1655–1664
- Balmaseda MA, Mogensen K, Molteni F, Weaver AT (2010b) The NEMOVAR-COMBINE ocean re-analysis. COMBINE Technical Report 1, FP7 COMBINE project
- Bjerknes J (1969) Atmospheric teleconnections from the equatorial Pacific. *Mon Wea Rev* 97:163–172
- Dee DP et al. (2011) The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q J R Met Soc* 137:553–597
- Guemas V, Corti S, Garcia-Serrano J, Doblas-Reyes FJ, Balmaseda M, Magnusson L (2012) The Indian Ocean: the region of highest skill worldwide in decadal climate prediction. *J Clim*. doi:10.1175/JCLI-D-12-00049.1
- Hawkins E, Sutton R (2009) The potential to narrow uncertainty in regional climate predictions. *Bull Am Meteor Soc* 90:1095–1107
- Hazeleger W et al. (2010) EC-Earth: a seamless earth-system prediction approach in action. *Bull Am Meteor Soc* 91:1357–1363
- Keeley SPE, Sutton RT, Shaffrey LC (2012) The impact of North Atlantic sea surface temperature errors on the simulation of North Atlantic European region climate. *Q J R Met Soc* 668:1774–1783
- Keenlyside NS, Latif M, Jungclaus J, Kornbluh L, Roeckner E (2008) Advancing decadal-scale climate prediction in the North Atlantic sector. *Nature* 453:84–88
- Kumar A, Chen M, Zhang L, Wang W, Xue Y, Wen C, Marx L, 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
- Madec G (2008) NEMO reference manual, ocean dynamics component: NEMO-OPA. Note du Pole de Modelisation 27, Institut Pierre-Simon Laplace (IPSL), France
- Magnusson L, Balmaseda MA, Corti S, Molteni F, Stockdale T (2012a) Evaluation of forecast strategies for seasonal and decadal forecasts in presence of systematic model errors. Technical Memorandum 676, ECMWF
- Magnusson L, Balmaseda MA, Molteni F (2012b) On the dependence of ENSO simulation on the coupled model mean state. *Clim Dyn*. doi:10.1007/s00382-012-1574-y
- Manganello JV, Huang B (2009) The influence of systematic errors in the Southeast Pacific on ENSO variability and prediction in a coupled GCM. *Clim Dyn* 32:1015–1034
- Mochizuki T et al. (2010) Pacific decadal oscillation hindcasts relevant to near-term climate prediction. *PNAS* 107:1833–1837
- Molteni F et al. (2011) The new ECMWF seasonal forecast system (System 4). Technical Memorandum 656, ECMWF
- Neelin JD, Dijkstra HA (1995) Ocean-atmosphere interaction and the tropical climatology. part i: the dangers of flux correction. *J Clim* 8:1325–1342
- Palmer T, Buizza R, Doblas-Reyes F, Jung T, Leutbecher M, Shutts J, Steinheimer GM, Weisheimer A (2009) Stochastic parametrization and model uncertainty. Technical Memorandum 598, ECMWF
- Pierce DW, Barnett TP, Tokmakian R, Semtner A, Maltrud M, Lysne J, Craig A (2004) The ACPI project, element 1: initializing a coupled climate model from observed conditions. *Clim Change* 62:13–28
- Pohlmann H, Botzet M, Latif M, Roesch A, Wild M, Tschuck P (2004) Estimating the decadal predictability of a coupled AOGCM. *J Clim* 17:4463–4472
- Pohlmann H, Jungclaus JH, Kohl A, Stammer D, Marotzke J (2009) Initializing decadal climate predictions with the GECCO oceanic synthesis: effects on the North Atlantic. *J Clim* 22:3926–3938
- Power S, Casey T, Folland C, Colman A, Mehta V (1999) Interdecadal modulation of the impact of ENSO on Australia. *Clim Dyn* 15:319–324
- Power S, Haylock M, Colman R, Wang XD (2006) The predictability of interdecadal changes in ENSO activity and ENSO teleconnections. *J Clim* 19:4755–4771
- Reynolds RW, Rayner TM, Smith DC, Stokes DC, Wang W (2002) An improved in situ and satellite SST analysis for climate. *J Clim* 15:1609–1625
- Roemmich D, Owens WB (2000) The Argo Project: global ocean observations for the understanding and prediction of climate variability. *Oceanography* 13:45–50
- Scaife AA et al. (2011) Improved Atlantic winter blocking in a climate model. *Geophys Res Lett* 38:L23703
- Schneider EK, Huang B, Zhu Z, DeWitt DG, Kinter JL, Kirtman BP, Shukla J (1999) Ocean data assimilation, initialisation, and predictions of ENSO with a coupled GCM. *Mon Wea Rev* 127:1187–1207
- Smith D, Cusack A, Colman A, Folland C, Harris G (2007) Improved surface temperature predictions for the coming decade from a global circulation model. *Science* 317:796–799
- Smith DM, Eade R, Pohlmann H (2012) A comparison of full-field and anomaly initialization for seasonal to decadal climate prediction. Submitted to *Clim Dyn*
- Spencer H, Sutton R, Slingo JM (2007) El nino in a coupled climate model: sensitivity to changes in mean state induced by heat flux and wind stress corrections. *J Clim* 20:2273–2298
- Stockdale T (1997) Coupled ocean-atmosphere forecast in the presence of climate drift. *Mon Wea Rev* 125:809–818
- Toth Z, Peña M (2007) Data assimilation and numerical forecasting with imperfect models: the mapping paradigm. *Physica D* 230:146–158
- Uppala SM et al. (2005) The ERA-40 re-analysis. *Q J R Meteor Soc* 131:2961–3012
- van Oldenborgh GJ, Doblas-Reyes FJ, Wouters B, Hazeleger W (2012) Decadal prediction skill in a multi-model ensemble. *Clim Dyn* 38:1263–1280. doi:10.1007/s00382-012-1313-4