TECHNICAL MEMORANDUM

404

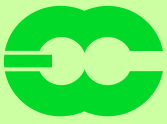
Comparison of the ECMWF seasonal forecast Systems 1 and 2, including the relative performance for the 1997/8 El Nino

David Anderson, Tim Stockdale, Magdalena Balmaseda, Laura Ferranti, Frederic Vitart, Paco Doblas-Reyes, Renate Hagedorn, Thomas Jung, Arthur Vidard, Alberto Troccoli, Tim Palmer

Research Department

SAC report September 2002

April 2003



European Centre for Medium-Range Weather Forecasts
Europäisches Zentrum für mittelfristige Wettervorhersage
Centre européen pour les prévisions météorologiques à moyen terme

For additional copies please contact

The Library
ECMWF
Shinfield Park
Reading
RG2 9AX
library@ecmwf.int

Series: ECMWF Technical Memoranda

A full list of ECMWF Publications can be found on our web site under:
http://www.ecmwf.int/publications/

Abstract

The skill of the recently-introduced seasonal forecast system (S2) is compared with that of the original system (S1) over the period 1987-2001. S2 differs from S1 in that it uses more recent cycles of both the atmospheric and oceanic models. Although the overall skill of the two systems is comparable, they differ markedly in their forecasts of the 1997/8 El Nino. A series of hybrid coupled experiments is performed in which different versions of the ocean model are coupled to different versions of the atmosphere model in order to quantify the effect of the changes to the atmospheric and oceanic models. It is found that changes to the atmospheric model are more important than changes to the oceanic model. To identify differences in the two atmospheric models a series of uncoupled integrations is performed for comparison with the coupled integrations. These show that many of the systematic errors of the coupled system are present in the uncoupled integrations and are therefore not a consequence of coupling.

A detailed analysis is made of the 1997/8 El Nino to try to identify the causes of the differences of the predictions for this time. This highlights the difference in the response of the atmosphere model winds to SST anomalies as being important, but also makes clear that errors in the ocean mean state are significant.

1 Introduction

Initial comparisons of System 1 (S1) and System 2 (S2) used for ECMWF seasonal forecasting showed that although S2 has many desirable features, the Nino 3 SST forecasts started in the latter part of 1997 were worse than those from S1. Because the 1997 El Nino was such an important event in terms of its impact on climate anomalies around the globe, it is important to understand the difference in predictions, and in particular why S2 was apparently inferior to S1 in this case. The Nino 3 SST forecasts for S1 were rather good so it was unlikely that any model change would lead to improved forecasts and there was a risk of change leading to inferior forecasts, as in fact happened. This will be discussed in the following sections. First a description of the two coupled models will be given in section 2. In section 3 a comparison of the results of the two systems will be presented, together with the results of a series of hybrid experiments performed to assess the relative contributions of the changes in atmospheric and oceanic model components. In section 4 the atmospheric response in the coupled models will be compared with the response in uncoupled experiments to show that much of the systematic error is of atmospheric origin, and that it is established within the first month of integration. In section 5 we will consider the origin of errors in the SST forecasts in 1997, to determine what part of these errors may originate from errors in the ocean model. A summary and conclusions are given in section 6.

2 Description and comparison of the two coupled systems S1 and S2.

2.1 The Ocean model of S1

The ocean model used is based on HOPE (Hamburg Ocean Primitive Equation model) version 2 (Latif et al. 1994, Wolff et al. 1997). The model is global and has 20 vertical levels. Horizontal discretisation is on an Arakawa E grid with a variable grid spacing: the zonal resolution is 2.8 degrees and the meridional resolution varies from 0.5 degrees in the equatorial region (within 10 degrees of the equator), smoothly increasing to 2.8 degrees polewards of 30 degrees. One difference to the basic version of HOPE is the use of a pseudo ice model to constrain the model solution over the polar regions and a second is the use of a slightly different bathymetry.

2.2 The ocean data assimilation scheme of S1

An ocean data assimilation scheme is used to introduce sub-surface observations into the ocean model. It is based on the statistical interpolation scheme described by Smith et al. (1991). This is essentially a univariate temperature optimum interpolation carried out on overlapping sub-domains of the model horizontal grid. Where domains overlap, the analyses are blended together. The division of the globe into sub-domains depends on the observation distribution and is done so that the maximum number of observations within the domain is less than 200. This reduces the cost of matrix inversion, (see Smith et al.1991), though with present computing resources this number could be increased. The optimum interpolation equations are solved on each level of the model independently over the top 1000 m. There is no temperature assimilation in the top model level; instead the model SST is relaxed to SST fields derived from the NCEP analysis, with a relaxation time-scale of 3 days. (See Reynolds and Smith 1994, and Smith and Reynolds 1998). A weaker relaxation (time-scale $\sim$ 30 days) was applied to the surface salinity which was relaxed to monthly-mean Levitus climatology. These constraints are not active during coupled model forecasts. The model is forced at the surface with specified fluxes of heat, momentum and fresh water. The fluxes are taken from ERA15 until the end of 1993 and from the operational atmospheric analysis system thereafter. The solar radiation penetrates below the surface layer exponentially.

The model background errors are represented by gaussian functions which are anisotropic and inhomogeneous. The values follow Smith et al. (1991). Within 4 degrees of the equator the correlation length scale in the E/W direction is 1500 km while in the N/S direction it is 200 km. In the sub-tropics and high latitudes, polewards of 15 degrees, the correlation length scale is 400 km in all directions. Between the equatorial strip and the sub-tropics there is a smooth transition in correlation scales. Observation errors are assumed to be correlated in space and time, with a spatial correlation function with length scale of 2 degrees and a time correlation scale of 3 days. Since the OI scheme is univariate, the magnitudes of the observation and background errors are not relevant, only the relative ratio which is 1.

The observations that were used are from the GTSPP (Global Temperature Salinity Pilot Project) at NODC (National Oceanographic Data Center). These include data from XBTs (eXpendable Bathy Thermographs), TAO moorings and drifting buoys. The analysis itself is carried out on each level independently. The analyzed temperature field is assimilated into the oceanic model as follows. Every 10 days the model is stopped and the model state is used as the background for an OI analysis using observations which span a window five days either side of the model background. An increment to the background is calculated. To avoid exciting gravity waves, and to allow the model dynamics to adjust gradually to the changes in the density field, this increment is added slowly over the subsequent 10 days, after which a new background field is available, and the cycle repeated. No change is made to the salinity field during the OI assimilation.

2.3 The Coupled model of S1

The atmospheric component of the coupled model is the ECMWF NWP IFS (Integrated Forecast System) model version 15R8, with a T63 spectral horizontal resolution and 31 vertical sigma levels. Cycle 15R8 was used for ECMWF operational medium-range forecasts from Jan – May 1997. A coupler, OASIS (Ocean Atmosphere Sea Ice Soil coupler) developed by Terray and Thual (1995) is used to interpolate between oceanic and atmospheric grids at coupling times (once per day). The initial conditions for the atmospheric model are taken from ECWMF re-analyses (ERA-15) for start dates up to Dec 1993 and from the ECMWF operational analyses for start dates thereafter as ERA-15 only extends to the end of 1993. Land surface conditions such as soil moisture and snow cover are similarly obtained from the atmospheric analysis system. The same ocean model is used for the coupled forecasts as was used to generate the ocean initial conditions. For the real time forecasts starting in December 1996, one coupled integration is made each day. An ensemble is made up of a

set of successive forecasts, which use initial conditions from successive days. For the 1991-96 forecasts, all of the coupled integrations start on the 1st of the month, and the ensemble is built up by making small perturbations to the initial SST fields. The period 1991-96 is taken as the model calibration period, and anomalies are calculated relative to this.

2.4 Comparison of S2 with S1

S2 is broadly similar to S1, but there are a number of differences in the ocean, the ocean data assimilation, the atmospheric model and the method of ensemble generation. These are outlined below.

2.4.1 The ocean model

The S2-ocean has 29 levels in the vertical compared to 20 in S1. Near the surface the level separation is 10m compared with 20m in S1. The horizontal resolution in the meridional direction is doubled at mid latitudes, and near the equator, is increased from 0.5 degrees in S1 to 0.3 degrees in S2. The resolution in the zonal direction is doubled at all latitudes. The mixing coefficients have been reduced in the higher resolution S2 ocean broadly in line with their scale dependence. In addition, the form of vertical mixing has been changed. Whereas in S1 a form of PP (Pacanowski and Philander 1981) mixing is used, in S2 a form of PGT (Peters, Gregg and Toole 1988) mixing is used. In order to ease comparison with (and later assimilation of) sea level from altimetry, the pressure gradient is evaluated at mid level. In the original version of HOPE as used in S1, the pressure gradient is evaluated at the bottom of the model levels. In forecast mode, the sea-ice is strongly relaxed to a damped-persistence anomaly. The damped persistence is defined such that after 60 days the sea ice field being relaxed to is purely climatology. One major technical difference concerns the barotropic mode, though it probably has little impact on the seasonal results. In S1, the time evolution of the baratropic mode is solved directly. This approach is quite efficient but requires large memory allocation. With the increased resolution in S2, this approach is no longer feasible. Instead, the barotropic component is multi-stepped with a time step of 15 seconds compared with 1 hour for the baroclinic part.

The data assimilation scheme remains OI as in S1, but the background error covariances have been changed. In particular the decorrelation scales have been reduced. A substantial change has been made to the way salinity is handled in the assimilation process. In S1, no change is made to salinity following an OI correction to T, whereas in S2, salinity is changed beneath the surface layer in order to preserve the T-S relationship (Troccoli et al 2001). This has lead not only to a marked improvement in the Atlantic sea level analysis but to an improved salinity analysis in the Pacific. Monthly climatology of river runoff is now included. There is still a relaxation to Reynolds climatological salinity but it is weaker: at the surface the relaxation time is about 6 months, and in the subsurface it is 18 months. A weak relaxation to Levitus thermal analysis is also included. There is a geostrophic correction applied to velocity following a density change (Burgers et al 2002). In both system 1 and system 2 the weight given to the data relative to the weight given to the background field varies with depth, to account for an increase in uncertainty associated with the large gradients near the thermocline. In system 1, the background and the observations are given the same weight, except in the thermocline, where the observations are given twice the weight of the background. In system 2, the observations are given half the weight of the background, except in the thermocline, where they are given twice the weight of the background.

One other difference relates to how the model is forced to create ocean initial conditions. In S1 we used the daily 0-24 averaged forecast stresses from the atmospheric NWP system (ERA15 and then operations). Model changes and changes to the analysis system were reflected in these stresses creating a low frequency variation that was unphysical and undesirable. By using analyzed winds rather than forecast stresses and calculating the stresses offline using a bulk formula, some of this variability could be reduced. The forcing fields used to create

ocean initial conditions as part of the ocean assimilation system come from ERA15 for the period 1987 to 1993 and from Operations from 1994 onwards.

The ocean initial conditions are provided not from a single ocean analysis but from a 5-member ensemble of ocean analyses. The analyses differ in that a measure of uncertainty in the surface winds used to force the ocean is taken into account. In the absence of ocean data assimilation, the uncertainty in ocean state is relatively large, but in the presence of ocean data assimilation is much smaller. The SSTs to which the ocean model is relaxed are different. The SST for S1 are the same as those used by the atmospheric analysis system, based on Reynolds OI version 1. For S2 we use the improved SST analyses from Reynolds OI version 2.

2.4.2 The atmospheric model

The atmospheric model is cycle 23r4. This is the same cycle as is used in ERA 40, except that there are 40 levels in the vertical, compared to the 60 used in ERA40. There are many differences between cycle 15r8 and 23r4: one of the more significant may be the set of physics changes introduced in cycle 18r3 (Gregory et al., 2000). This certainly had an impact on the coupled system drift to which we will return later.

The atmospheric initial conditions, including land conditions come from ERA15 for the period 1987 to 1993 and from Operations from 1994 onwards. The ensemble generation strategy is somewhat different to S1. In S2, the real-time ensemble set consists of 40 members, and the calibration set consists of 5 members spanning the 15-year period 1987-2001, so creating a calibration PDF of 75 members. Each of these ensembles has a single start date, and the ensemble is constructed by combining the 5 member ensemble ocean analysis with SST perturbations and the activation of stochastic physics. S2 thus has a greater consistency between the calibration and real-time ensembles than was obtained in S1.

3 Overview of coupled model SST forecasts.

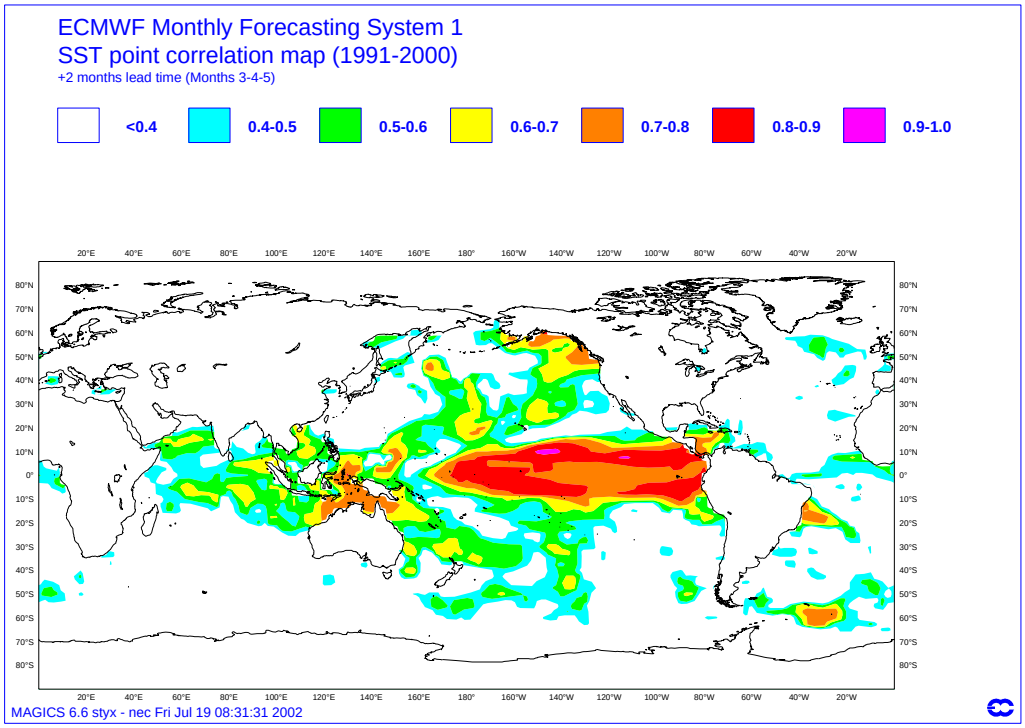
To allow a comparison between S1 and S2 over a longer period, S1 integrations were extended back to 1987 (for operational use S1 back integrations began in 1991). For this purpose it was necessary to recreate S1 for research experimentation (a major task) and run additional integrations. In particular a 5-member ensemble for each month spanning the years 1987-1990 was created. Most of the statistics presented in this section refer to the 1987-2001 period. However, for technical reasons, some diagnostics were not possible over the extended period, and in those cases the period for validation will be indicated.

3.1 General characteristics of S2 compared to S1.

Before concentrating on the 1997 El Nino predictions, it is necessary to describe general properties of S2 SST forecast skill. A broad overview can be obtained by considering correlation maps though scatter plots will later be used to give more quantitative assessments. In Fig 1a and 1b we show the SST point correlation maps for S1 and S2. These are created by calculating and plotting the temporal correlation of forecast and observed SST anamolies at each point, and the plots show very similar behaviour between the two systems. There are some differences in detail in the Pacific: in particular S2 may be better in the North Pacific, though this might not be particularly robust given that the record spans only 11 years (1991-2001). The patterns in the Atlantic are remarkably similar though the general level of skill is low. Skill in the Indian ocean is also low.

As well as correlation, the amplitude of anomalies is important. In Fig 2 a, b we plot the amplitude of the SST anomaly predicted relative to that observed for both the region Nino 3 and the region Nino 4 for forecasts

a)



b)

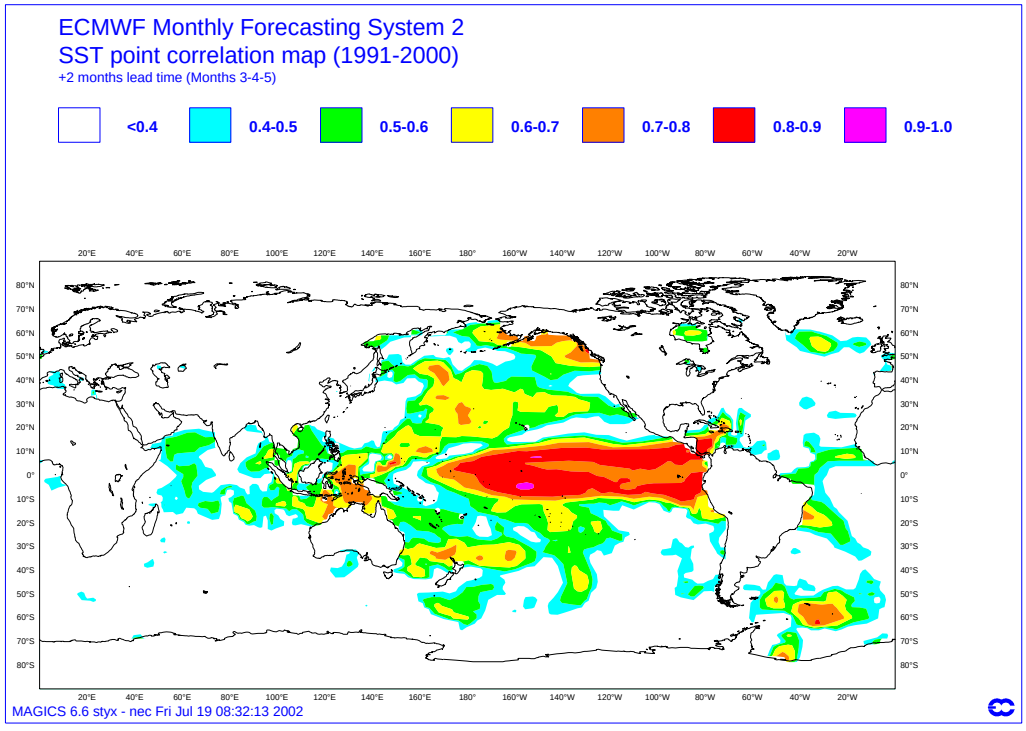


Figure 1: Map of the temporal anomaly correlation between forecasts and observations of SST anomaly, for the forecasts for months 3-5 and for the years 1991-2001 for a) S1 and b) S2.

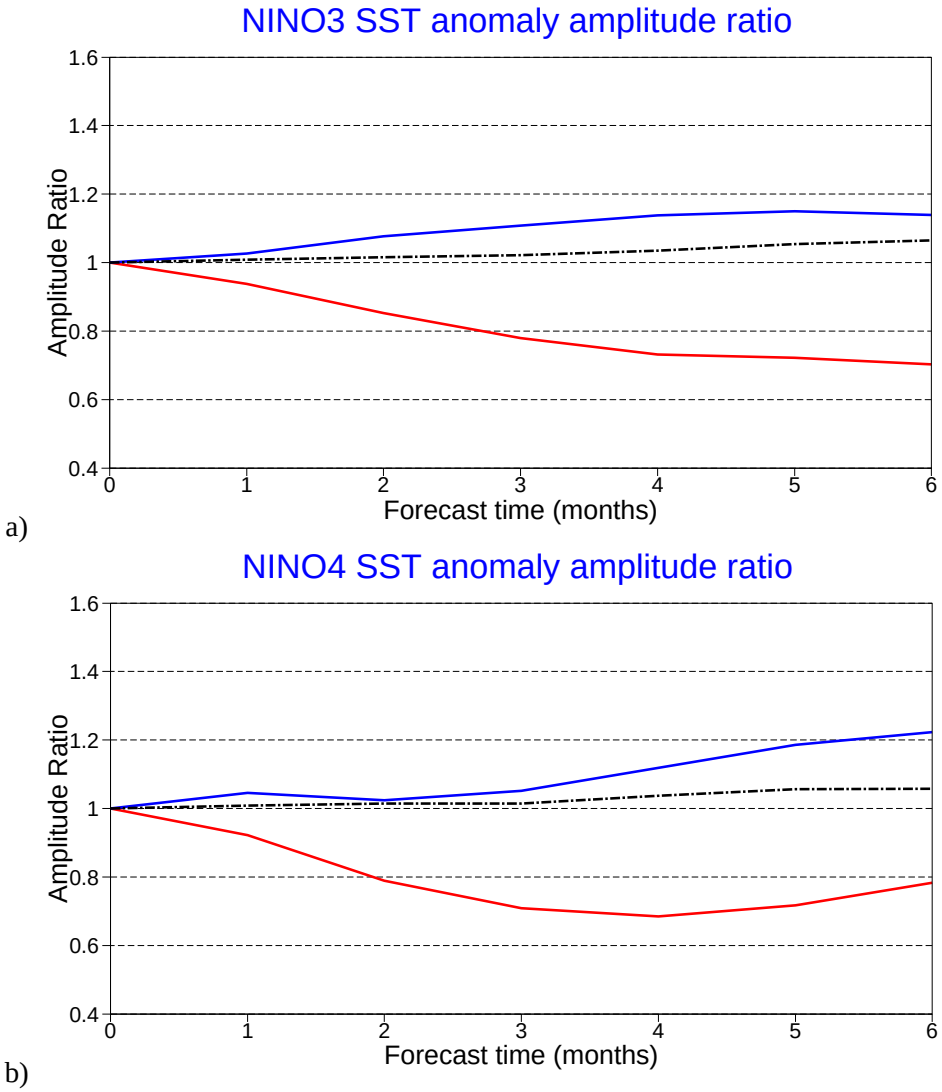


Figure 2: Amplitude ratio of the SST anomalies in Nino 3 and Nino4 relative to observed. S1 (blue), S2 (red) persistence (dashed). Results are based on the period 1987-2001. In both regions S1 is slightly over active but S2 is damped. For a definition of the regions see Appendix.

spanning the period 1987-2001. This shows that in both regions, S2 is damped. In fact, in the later months of the forecast, the anomalous SSTs are typically only 0.7 of what they should be. By contrast S1 is a little too active in both regions.

The causes of SST variability might well be different in the two regions, but wind variability is likely to be important for both. It is therefore instructive to check the amplitude of the wind variability, not in the Nino 3 region, but further to the west since it is wind variability around and to the west of the date line that most often influences SSTs further east. In Fig 3 the variability in the wind in the Nino 4 and EQ2 regions is plotted for both S1 and S2. The period covered here is 1991-6 which is shorter than for fig 4 but the signal is clear. Both systems are damped but the damping in S2 is much more rapid than in S1. The wind amplitude drops rapidly in the first month, suggesting, but at this stage not proving, that the lack of wind response is an atmospheric problem rather than a coupled one: the SST amplitude (Fig 2) drops more slowly in S2 than the wind variability suggesting that the anomalous SST is responding to the winds with some delay.

In the later stages of the forecast the wind variability in S1 is damped but the SST variability is not, suggesting that the ocean model may be too active (i.e. has too strong an SST response for a given wind anomaly). The fact that the wind variability drops only after the second month into the forecast suggests that the drift in the coupled system may contribute to this.

To quantify the magnitude of error in the two systems we plot in Fig 4 the mean absolute error over the 6 month forecast for every month from 1987 to 2001 for a) Nino3 and b) Nino4. The results are based on the 5-member ensemble for this period and are shown in red for S2 and in blue for S1.

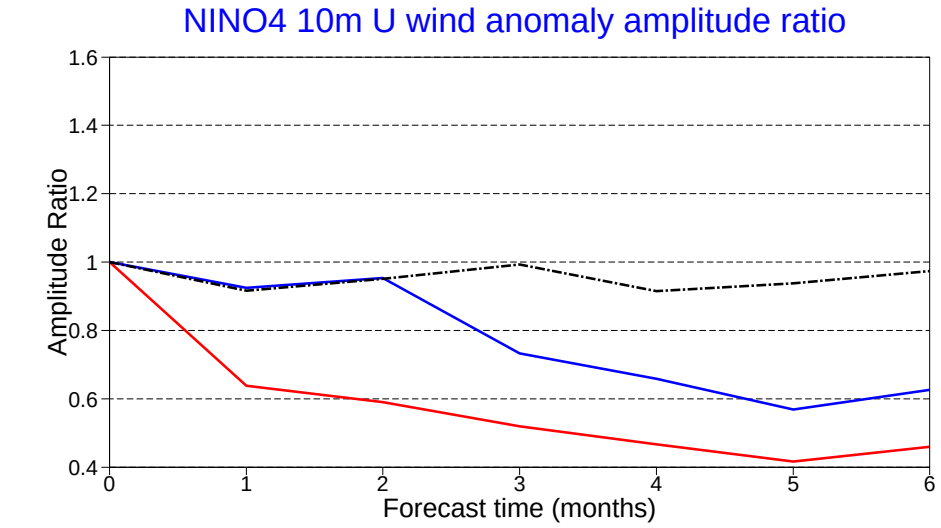
Considering Nino 3 first, in general, S2 has lower errors than S1: the average error is 0.40K for S2 compared to 0.45K for S1. However, it is also apparent that the errors in S2 are larger than in S1 during part of the 97 El Nino. If one compares the Nino 4 errors, one gets a different impression. In this case the errors are generally larger for S2 than for S1 but during part of the 1997 event, the opposite is true. Thus the picture is rather different with respect to general errors and with respect to errors during a specific period.

The relative performance of the two systems can also be compared by plotting scatter diagrams of the errors. These are shown for both Nino 3 and Nino 4 in Fig 5. The large diagonal cross marks the centroid of the cloud of small crosses while the length of its arms gives a measure of the reliability. If the centroid is displaced away from the 45 degree line by an amount greater than the cross line, then this indicates that using a Student-t test and assuming the forecast errors are independent for different dates, the displacement is significant at the 95% level. Fig 5 confirms the general superiority of S2 in the Nino 3 region and the superiority of S1 in the Nino 4 region. Very marked in both figures is the large spread indicating that the two systems behave very differently. We will return to this point later when we compare results of hybrid experiments and of S2 with the Met Office model. Both the atmospheric and oceanic components of the coupled system are different between S1 and S2. Which contributes most to the different performance shown above?

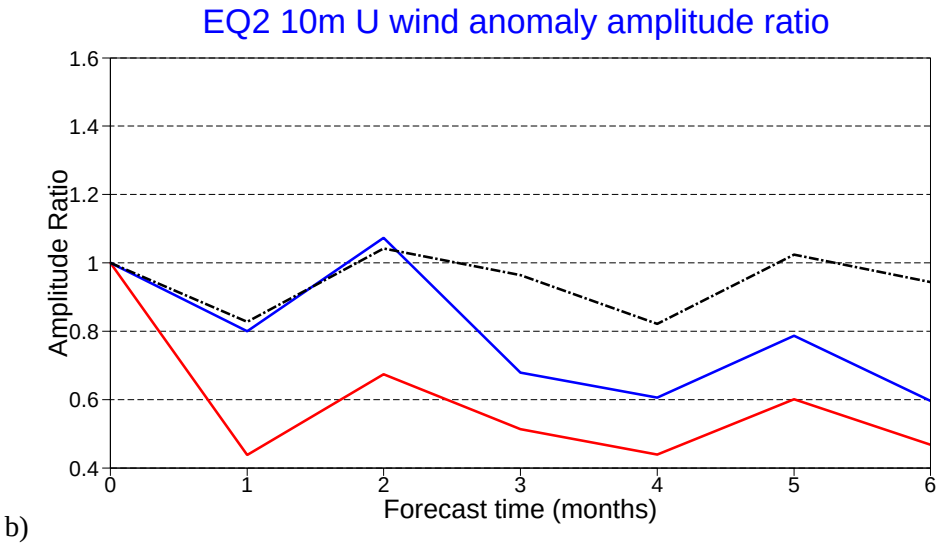
This can be answered in part by performing hybrid experiments. Let us define O1 as the version of the ocean model used in S1, O2H as the ocean model used in S2. The H indicates that this is high resolution. It is also possible to run the revised version of the ocean model at the same resolution as O1 which we will denote as O2. The two atmospheric versions are denoted A1 and A2. The resolution for these varies since A1 is T63 while A2 is T_L95 but both exchange information with the ocean at the same resolution, since T_L95 has the same grid as T63. A1 is cycle 15r8 and A2 is cycle 23r4.

A matrix of experiments is possible.

* Only with considerable effort was it possible to resurrect cycle 15r8 for research experiments. The priority was to concentrate on the low resolution experiments. With further effort it should be possible to run the combination A1_O2H, but for the moment we will concentrate on the first 4 components of the matrix.



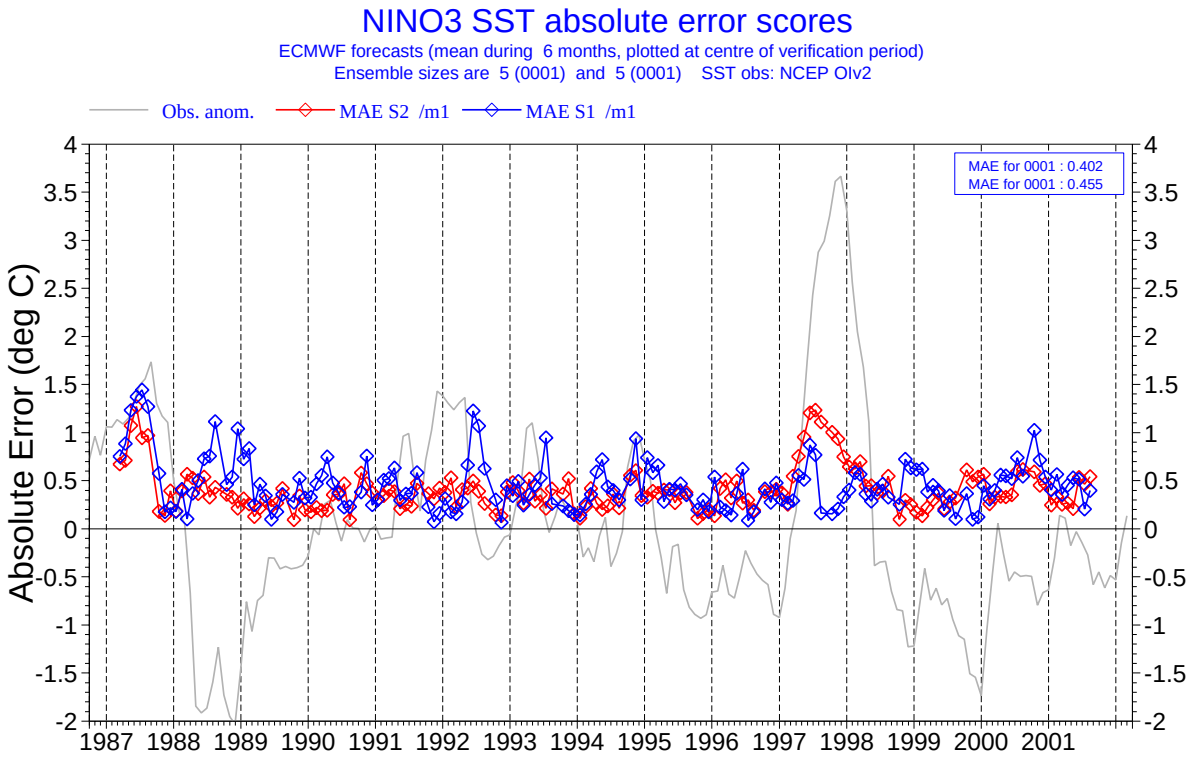
a)



b)

Figure 3: Amplitude ratio of the 10m wind in the Nino4 and EQ2 regions of the central west Pacific. S1 (blue), S2 (red) persistence (dashed). Both S1 and S2 are damped but the damping is much more acute in S2, and develops rapidly.

a)



b)

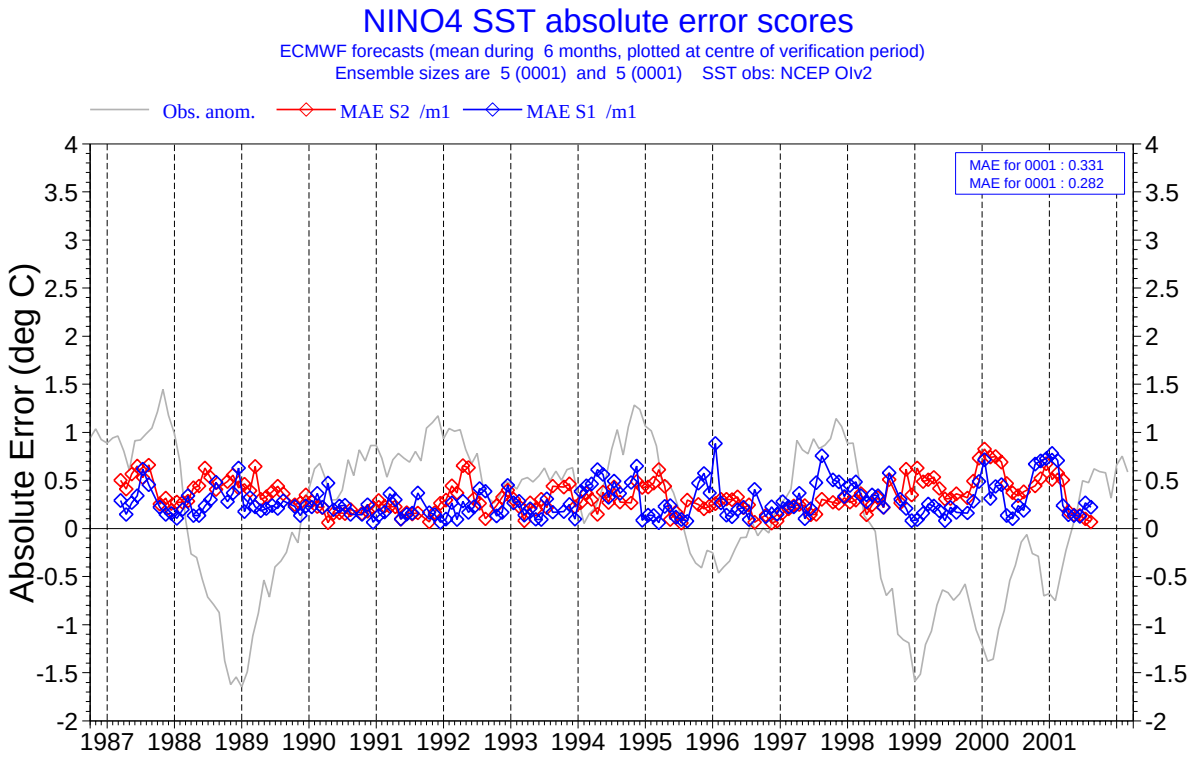
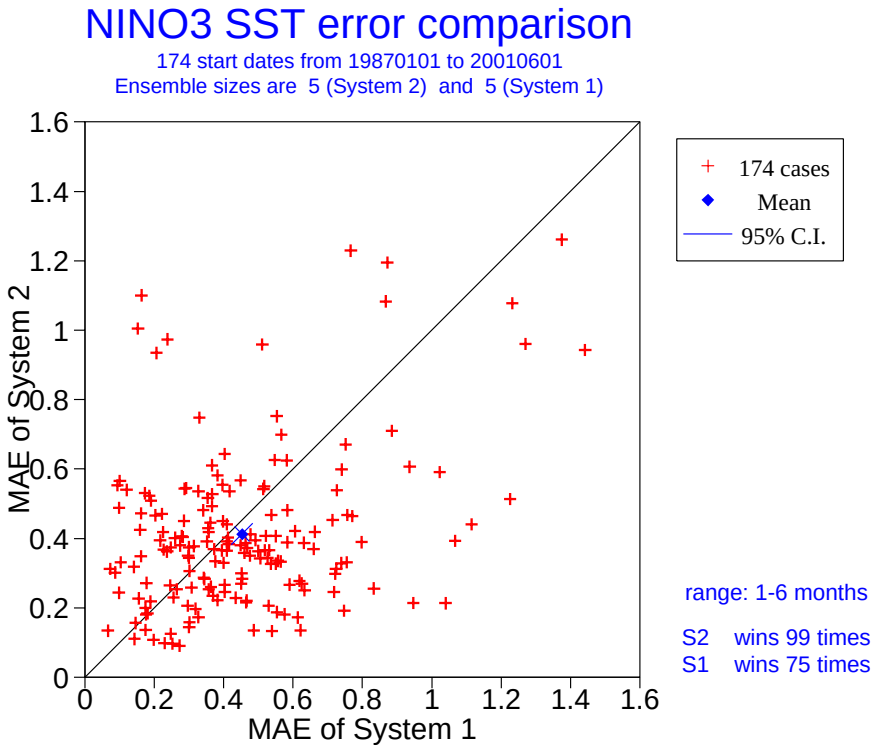


Figure 4: Mean absolute error for a 6-month forecast for a) Nino3 and b) Nino4. Red indicates S2 and blue S1. The grey curve shows the amplitude of the observed anomaly.



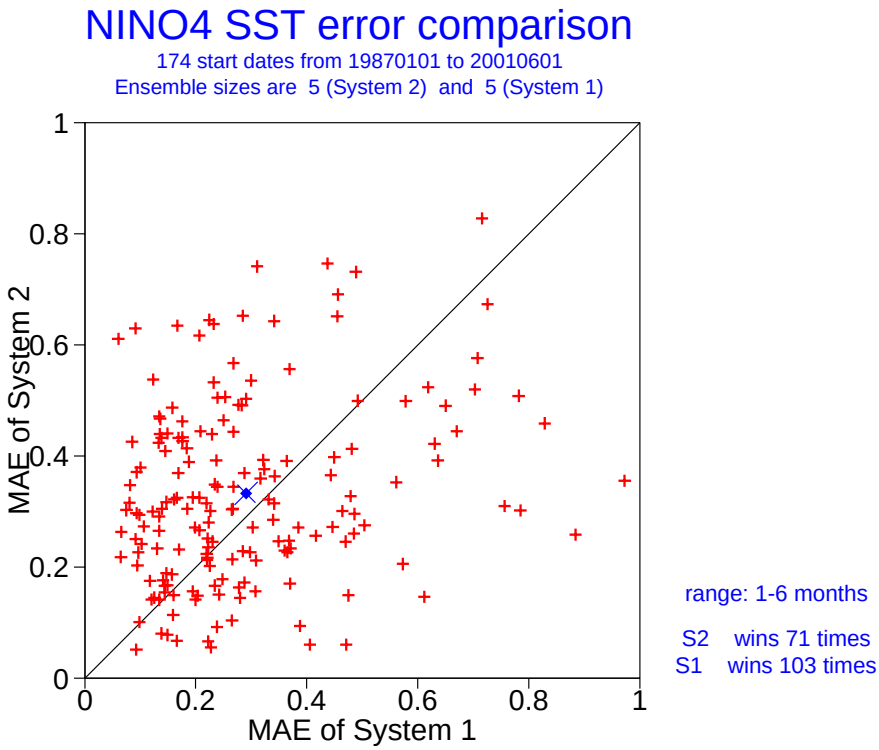


Figure 5: Scatter plot of the mean absolute error for forecasts covering the period 1987- 2001 for S2 v S1: a) Nino3, b) Nino4. The blue star marks the centroid, with the arms of the cross indicating the 95% confidence interval for the position of the centroid.

Ocean atmosphere	A1	A2
O1	S1	e906
O2	e92q	e6xp
O2H	*	S2

Table 1: Summary of experiments showing the different atmospheric and oceanic components

3.2 Results of hybrid experiments

3.2.1 Comparison of forecasts using the different ocean models O1 and O2.

By comparing A1_O1 (S1) with experiment A1_O2, we can test the influence of the changes made to the ocean physics and data assimilation in O2 on the forecast results. In both experiments the atmosphere is 15r8 and the resolution is the same. The resolution of the ocean is also the same in the two experiments. A scatterplot of mean forecast error is shown in fig 6 for both Nino3 and Nino4. In both regions the results of A1_O2 are better than those for A1_O1 though in Nino4 the improvement in results is not significant at the 95% level. The ratio of forecast SST to observed is shown in fig 7 for both Nino3 and Nino4. In both regions O2 reduces the level of activity suggesting that O2 is less active than O1. In Nino3 it is close to 1.0 in experiment A1_O2, suggesting that the A1_O2 combination might be about right. In Nino4, O2 is overactive though not as overactive as O1.

A similar behaviour is obtained when we compare the relative performance of O1 and O2 but this time coupled to the A2 atmosphere (not shown). To summarise, the skill of A2_O2 is greater than that of A2_O1, in all Nino regions as measured by scatterplots, rms and mean error growth and anomaly correlation, but the new (O2) ocean is more damped than the original (O1) ocean. Since the effect of A2 is also to damp, this means that A2_O2 is too damped.

3.2.2 Comparison of forecasts using the two atmospheric models A1 and A2.

We will first compare A1_O1 (S1) with A2_O1. Fig 8 shows the scatter plot of SST error in the two experiments. The skill in the Nino4 region has decreased as a result of using A2. This also seems to be true in Nino3 though the results are not significant at the 95% level. Fig 9 shows the amplitude ratio of the SSTs in Nino3 and Nino4 compared to those from S1. It is clear that S1 is more active than A2_O1 in both regions. In Nino3 this means the coupled model A2_O1 is not sufficiently active. In Nino4 it has about the right level of activity.

A further comparison of the role of the atmosphere can be made by comparing A1_O2 with A2_O2. The skill scores are shown in fig 10. They show that the greatest effect is in the Nino4 region and that the use of A2 leads to worse results there. The effect of using the A2 atmosphere leads to a damping of the variability (not shown).

3.2.3 A comparison of O2 and O2H

By comparing A2_O2 and A2_O2H (S2) we can gauge the impact of increasing the ocean model resolution. The changes are not just direct resolution changes, however, as the choices made in the data assimilation will also have an effect. The net effect of the changes from O2 to O2H is to slightly dampen the response in Nino3 but to dampen it further west as in eg Nino4 or EQ3. The reason for this damping is not totally clear but is likely to be related to the reduced westward extension of the oceanic cold tongue in the high resolution ocean. In the low resolution ocean the cold tongue extends too far westward. The high resolution ocean has a more realistic cold tongue structure.

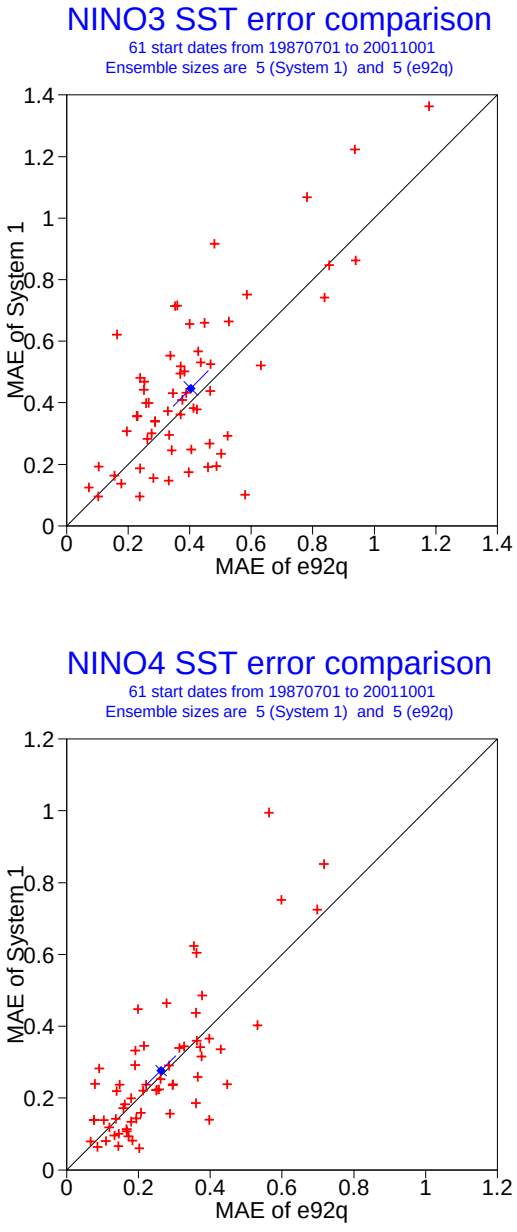
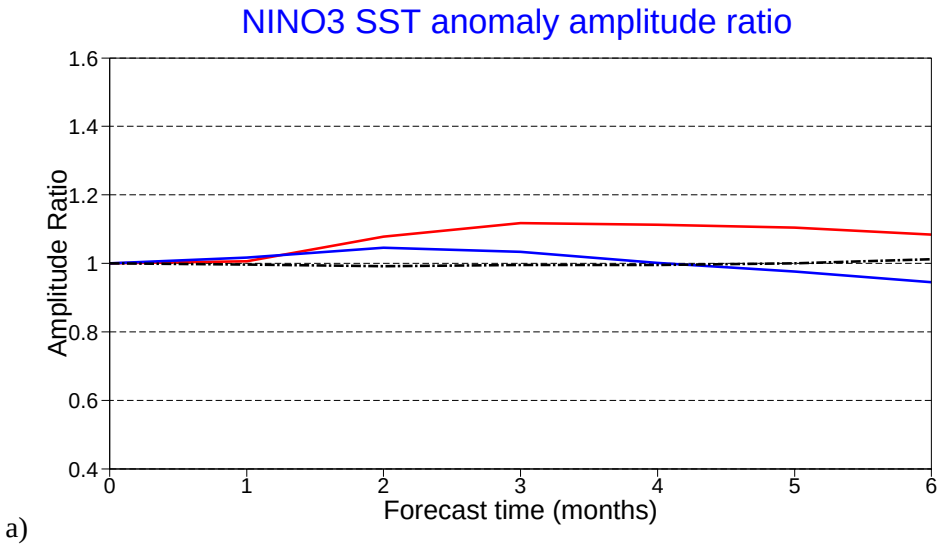
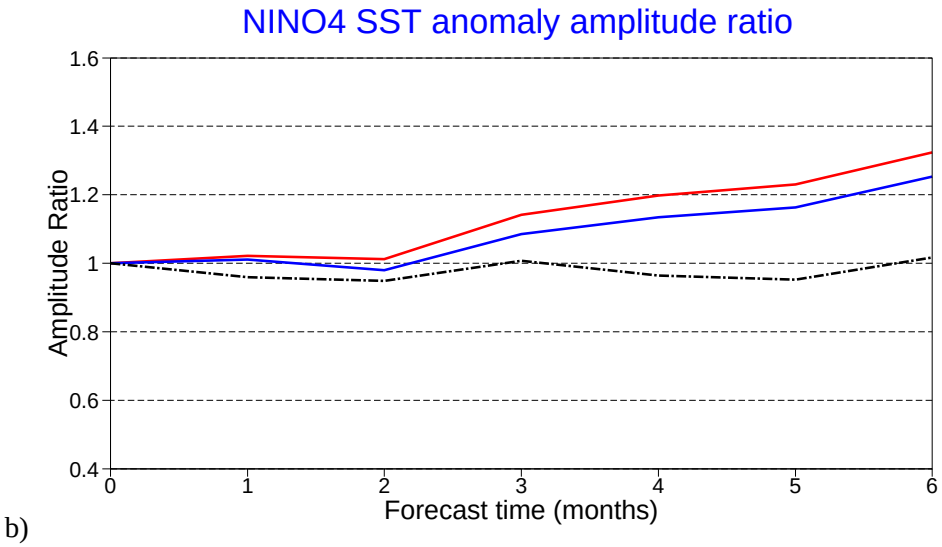


Figure 6: Similar to Fig 5 but comparing experiments S1 with A1_O2 (e92q). This comparison shows the effect of changes to the ocean model and the ocean data assimilation. The atmospheric component is the same in both.



a)



b)

Figure 7: Similar to Fig 2 but for S1 (red) and A1_O2 (blue).

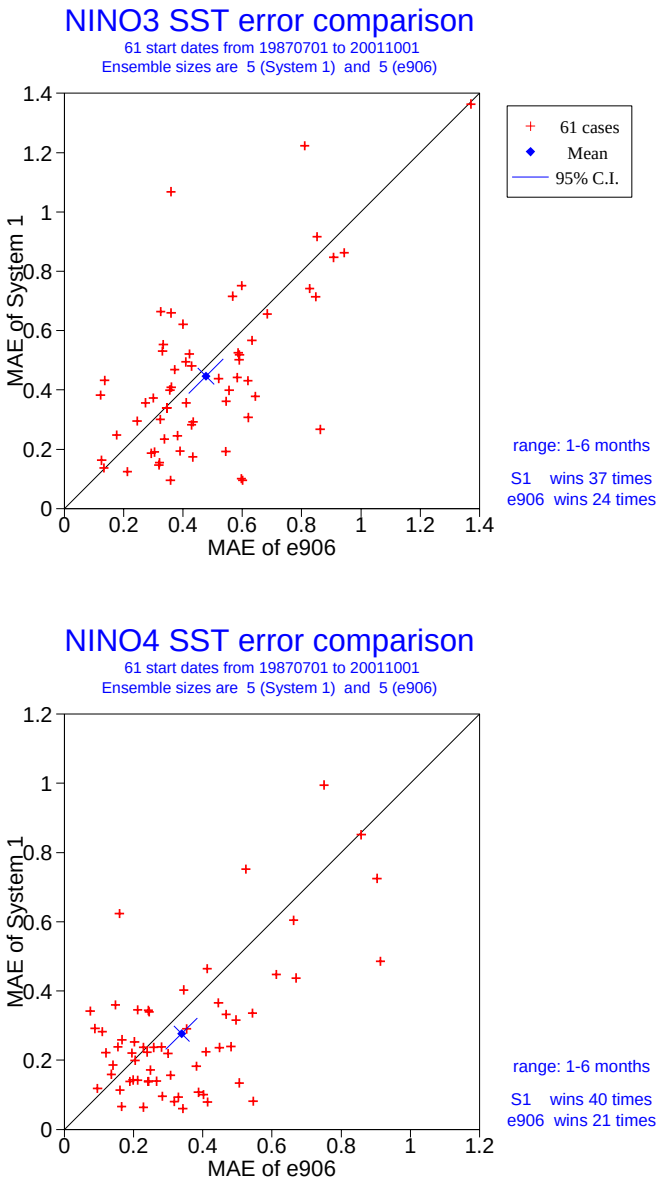
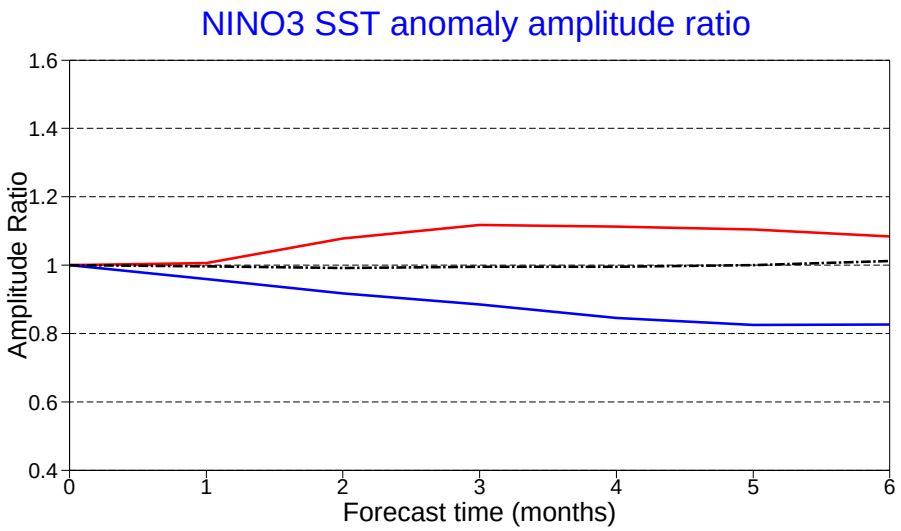
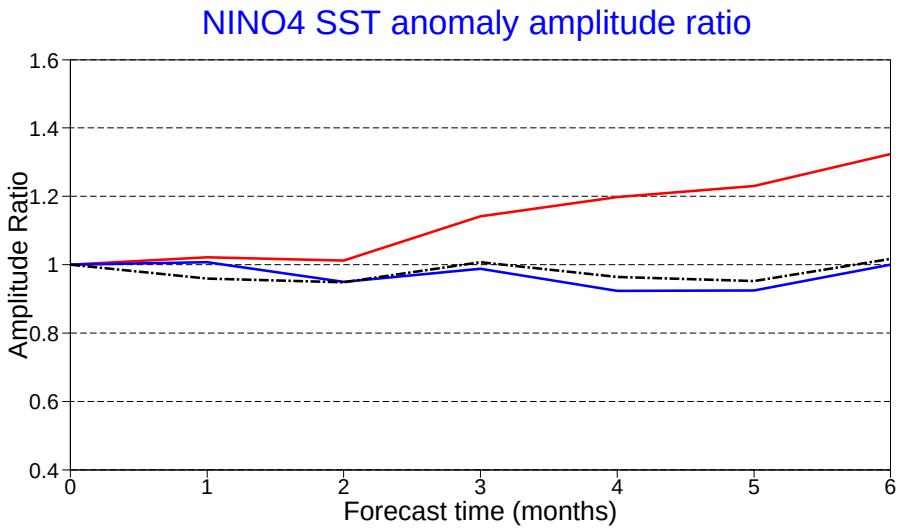


Figure 8: Similar to Fig 5 but comparing S1 and A2_O1. The ocean models are the same but the atmospheric models are A1 for S1 and A2 in experiment A2_O1. The effect of changing the atmospheric model leads to a small (not significant) degradation in Nino 3 but a bigger degradation in the Nino 4 region, which appears significant at the 95% level.



a)



b)

Figure 9: Similar to Fig 2 but comparing S1 (red) and A2_O1 (blue).

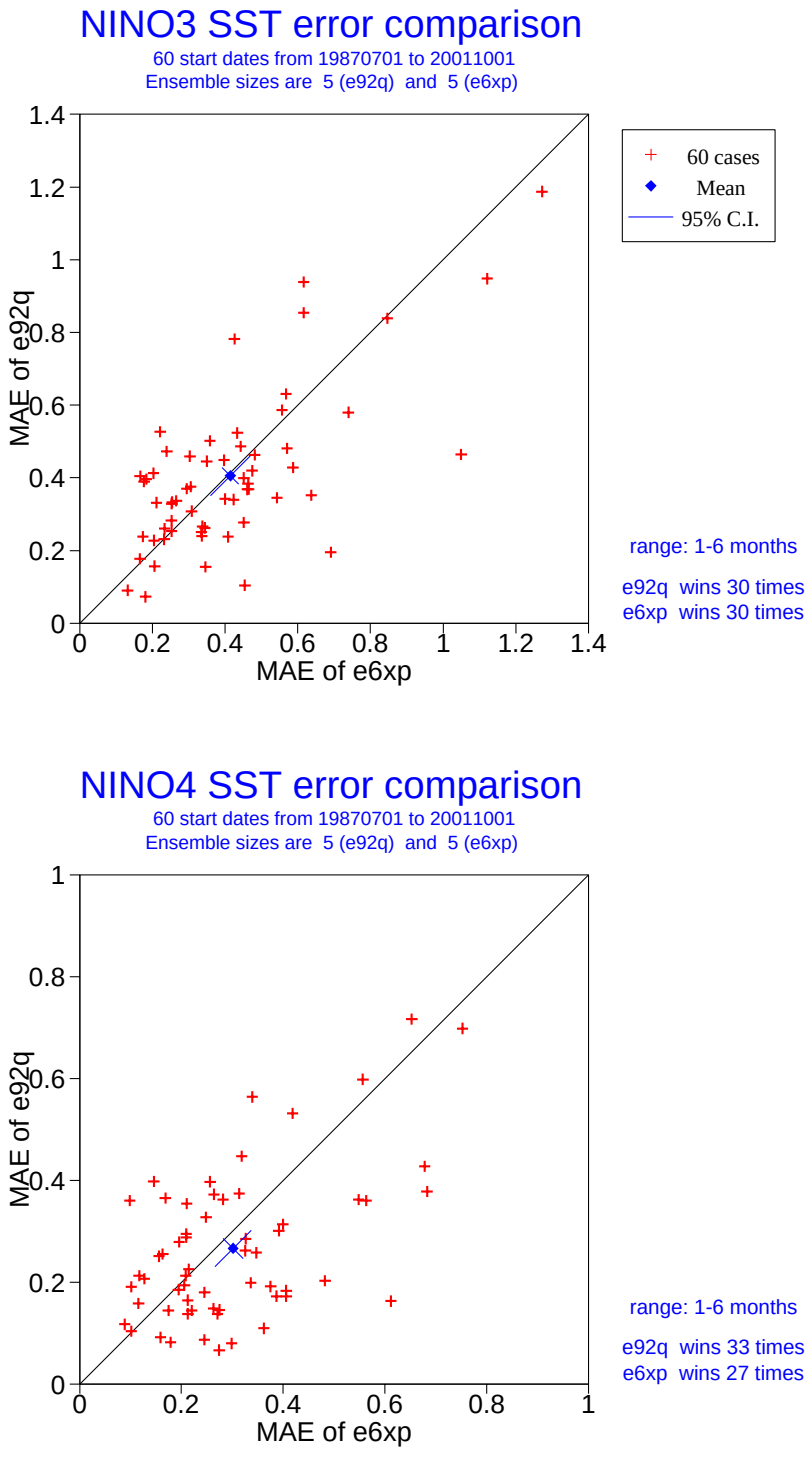


Figure 10: Similar to Fig 5 but comparing experiments A1_O2 (e92q) with A2_O2 (e6xp).

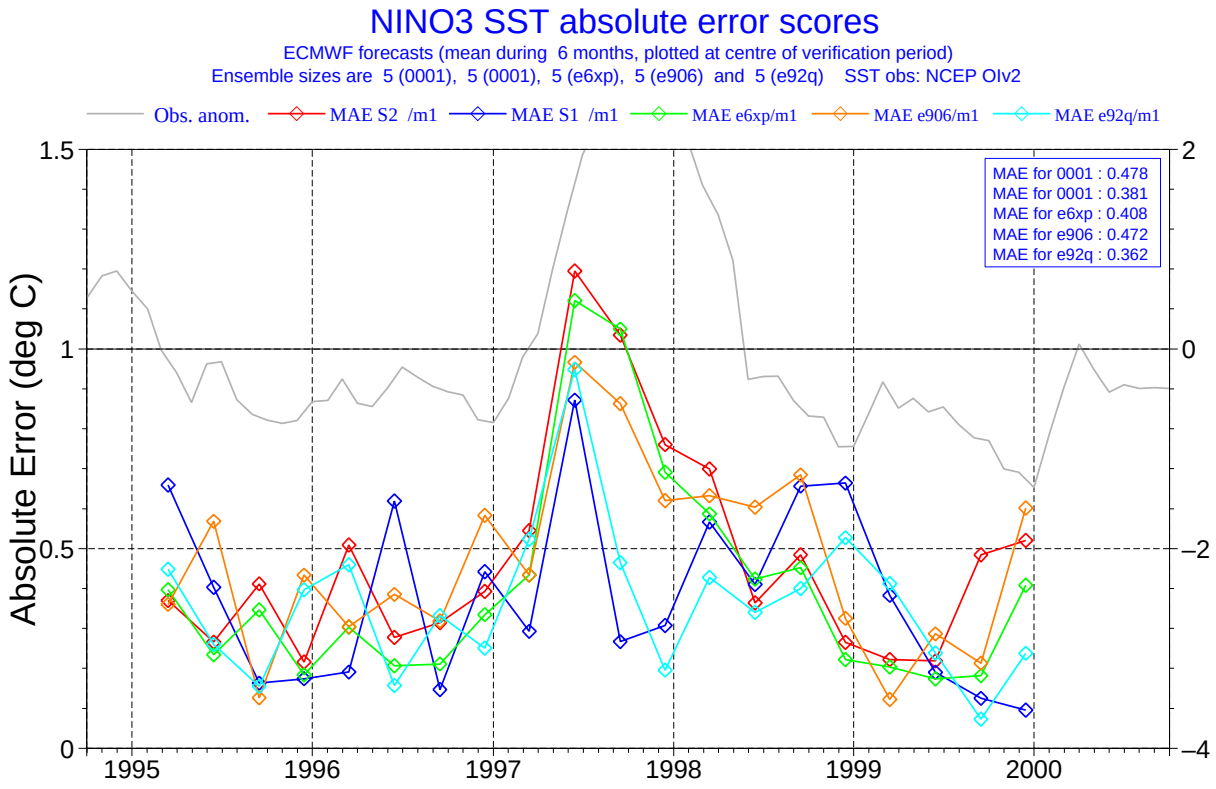


Figure 11: Plot of the mean absolute error averaged over the 6 month forecasts, for a series of integrations spanning the 1997/8 El Nino. Dark blue corresponds to S1, red to S2, green to A2_O2, light blue to the hybrid A1_O2, and orange to the hybrid A2_O1. The forecasts using the A1 atmosphere have lower errors than those using the A2 atmosphere. The scale is on the left. The grey curve indicates the observed SST, the scale for which is on the right hand: it is four times that for the error.

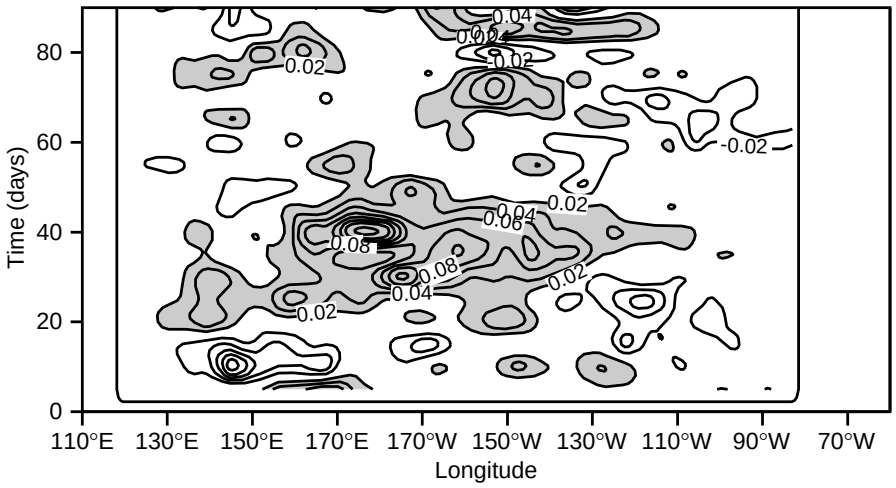


Figure 12: Zonal wind stress anomaly along the equator from 1 May to 31 July 1997. Units Nm^{-2}. From Balmaseda et al 2002.

3.3 The 1997/8 El Nino

In section 3.1 we considered the overall statistics of S1 and S2. Here we consider specific forecasts for the 1997/8 period. Fig 11 shows the mean absolute error for Nino3 for various experiments discussed above focussing on the 1997/8 period. For January starts all models forecast the onset and errors are moderate at around 0.5K. For April starts the rapid intensification was not well captured by any of the models including S1 and errors are large. Forecasts starting in the May-June period also performed poorly though not shown on this plot. In fact forecasts for May 1997, have been further studied by Balmaseda et al. 2002 in which various sensitivity studies have been performed. In not a single experiment, was the full extent of the observed warming from this period achieved. In May-June 97, the trade winds collapsed over much of the equatorial Pacific as fig 12 shows. Even when this wind anomaly was included in the coupled response, the coupled model failed to capture the full warming. The SST anomaly was increased as a result of the wind intervention but additional heat fluxes were generated which prevented its full amplitude being achieved. In ocean-only experiments using observed wind and heat fluxes, the amplitude was much better simulated, but even so some errors were present in the east Pacific arising either from errors in the forcing fields or errors in the ocean model. It was found that changes to the CAPE threshold used in the convection scheme could lead to improved coupled forecasts but none generated the full observed warming.

The character of the errors in the later stages of 1997 are somewhat different in that they show much greater sensitivity to different model combinations. The models with the lowest error are those using A1. There is not much to chose between A1_O1 and A1_O2 in the later stages of 1997 but in fact the model with the lowest overall error in fig 11 is A1_O2. The models with the largest errors are those using A2.

In Fig 13a we plot the plume diagrams for July 97 start times for all 5 experiments of table 1. The plume members which are closest to the observed SST are those using the A1 atmosphere. The three forecast systems using A2 are all rather similar and underpredict the growth of the SST. This separation is even clearer for forecasts started in October as shown in fig 13b. For completeness we also show, in Fig 14, results for a different region eg EQ2. In this case the forecasts using A1 are too active while those using A2 are underactive for July starts, but in all the cases the models underpredict months 1-2. For October starts, it is harder to classify the forecasts but again the A2 atmosphere gives weaker SST anomalies than the A1 atmosphere in the later stages of the forecasts: the SST anomalies decay faster with the A2 atmosphere. All the forecasts, except

S2, overshoot in the first 3 months of the integration. Detailed analysis of these results is given in Section 5.

Finally to put these results in context, we plot results from other forecasts made for the 97/8 event. There are two sets of forecasts: those made as part of the Met Office collaboration to develop a real-time seasonal forecasting system and those made as part of the DEMETER project. In fig 15 we show results from the real-time experiments. The purpose of these is to show the wide range of model behaviour for some start dates, and to show that a model which does well on one start date or in one region may do poorly on another start time or in another region. Fig 15a shows forecasts from May 1997 for S2 and the Met Office. The spread in the Met Office forecasts is much larger than in S2, and spans the observed SST. For forecasts started in July 97 (panel b) the spread is even larger. This is particularly striking since the forecasts of the onset from Jan, Feb, Mar, and April 1997 were remarkably consistent between the two systems and the spread in the two systems was broadly comparable (not shown).

In fig 16 we show forecasts for the Nino4 region for May and July 1997 starts. Here the two systems differ markedly in their predictions. While the met Office forecasts looked superior for the Nino3 region, they look much too active in the Nino4 region. Again, for start dates from Jan, Feb, Mar, and April 1997, the two systems looked quite comparable in spread and in the forecasts.

Results from DEMETER are shown in fig 17 a, b, c for start dates of May, August and November. The results are obtained using three models. Two are different to those in fig 16. The third model is a version of S2 (shown in blue) but it is initialised using ERA40 rather than products from the operational system. Despite these differences the ECMWF forecasts are very similar to those from S2. Comparing the ECMWF forecasts with the others in fig 17a, suggests that ECMWF is inferior. The forecasts which are best over the first couple of months come from the orange model. In fig 17b and c, however, it is clear that this model is far too active, with predicted anomalies in excess of 6K. The implication is that its success for forecasts starting in May was fortuitous and occurred because the model was too energetic. For the three start dates shown, the green model is better than the Demeter version of S2, but not better than the forecasts for S1. The conclusions are similar to those from fig 16: Those systems which caught the rapid warming for Spring starts did poorly for later starts and were far too energetic. No model did a good job throughout.

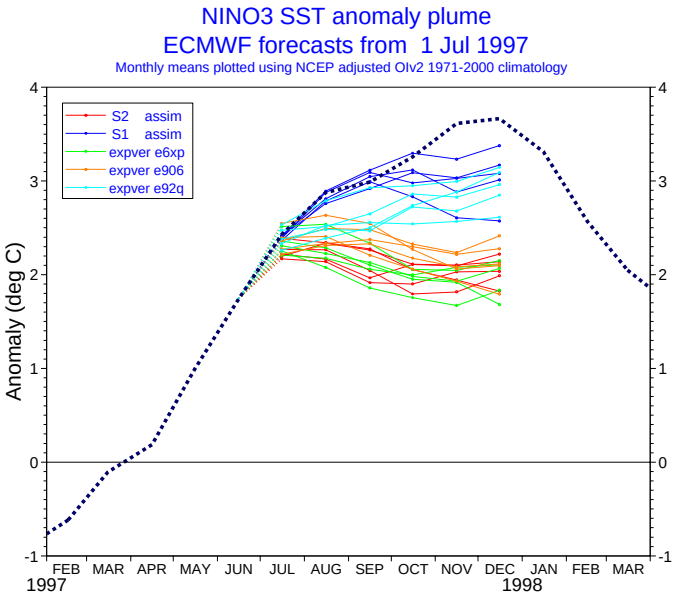
3.4 Summary of SST forecast behaviour.

In this section we have shown some overall scores for S1 and S2. In general S2 is better than S1 in Nino3 but not in Nino4. During the 1997 event the skill was reversed: S1 did better in Nino3 but S2 did better in Nino4. A series of hybrid experiments was performed in which the ocean and atmospheric components were interchanged and forecasts performed every season for the years 1987-2001. These showed that the switch from O1 to O2 was beneficial, especially in Nino3. Whether this resulted from changes in the ocean parameterisations or in the ocean data assimilation system was not examined. The switch from A1 to A2 was not beneficial. There was a noticeable degradation in the Nino4 region, with a smaller effect in Nino3. S1 was a relatively well-balanced system in terms of the amplitude of SST anomalies with a tendency to be slightly overactive. By contrast S2 is underactive. This seems to be both the effect of changing the atmosphere and of changing the ocean. With respect to the atmospheric change, it seems that the winds are damped. The winds are also damped in S1 but the damping starts earlier in S2, and it is more marked.

Forecasts for the 97/8 El Nino show that changing the atmosphere has a bigger effect than changing the ocean. Comparison with other model forecasts for 1997/8 shows a huge range in behaviour, strengthening the need for multi-model forecasts. No model seems capable of doing a satisfactory job throughout.

In fig 18 we show summary plots of the results presented from the various experiments. Panel a) shows the drift in Nino3 across the several experiments ranging from the very cold (S1) to the warm (S2). Pink and blue

a)



b)

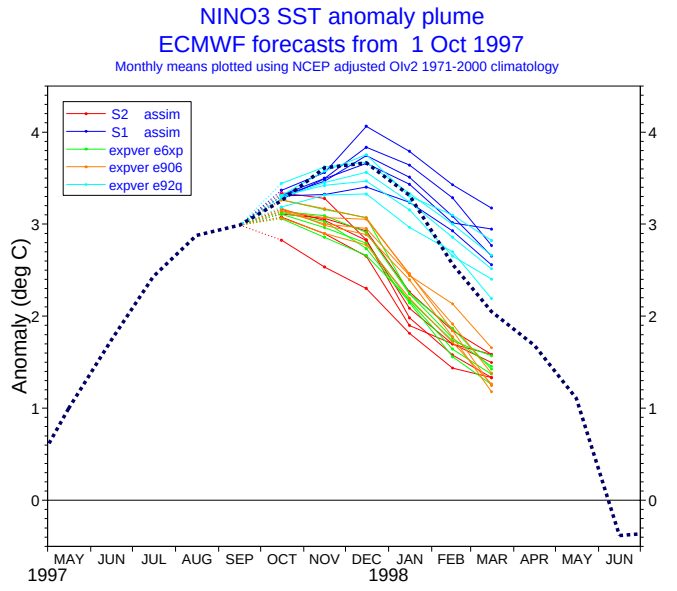
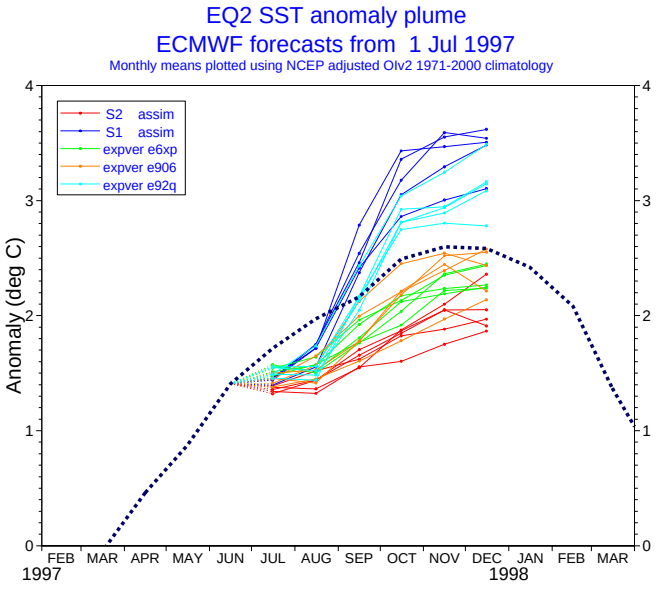


Figure 13: Plume diagrams for a) July 1997 and b) October 97 for the Nino3 region. All five experiments from the matrix of table 1 are plotted. The colour code is A1_O1 (S1) dark blue, A1_O2 light blue, A2_O1 gold, A2_O2 green and S2 red.

a)



b)

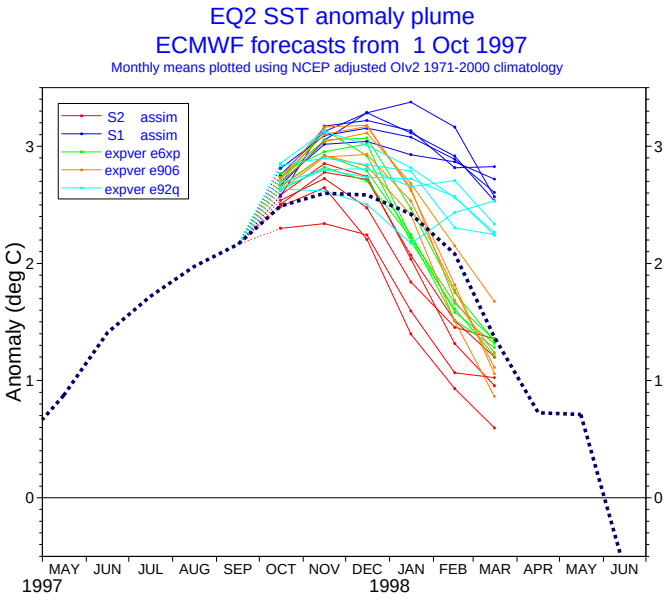
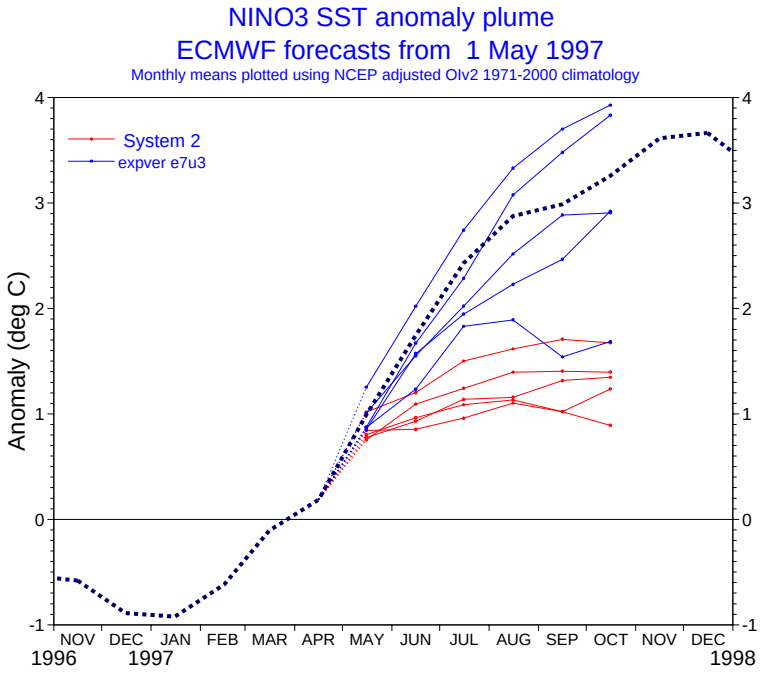


Figure 14: As for fig 13 but for EQ2.

a)



b)

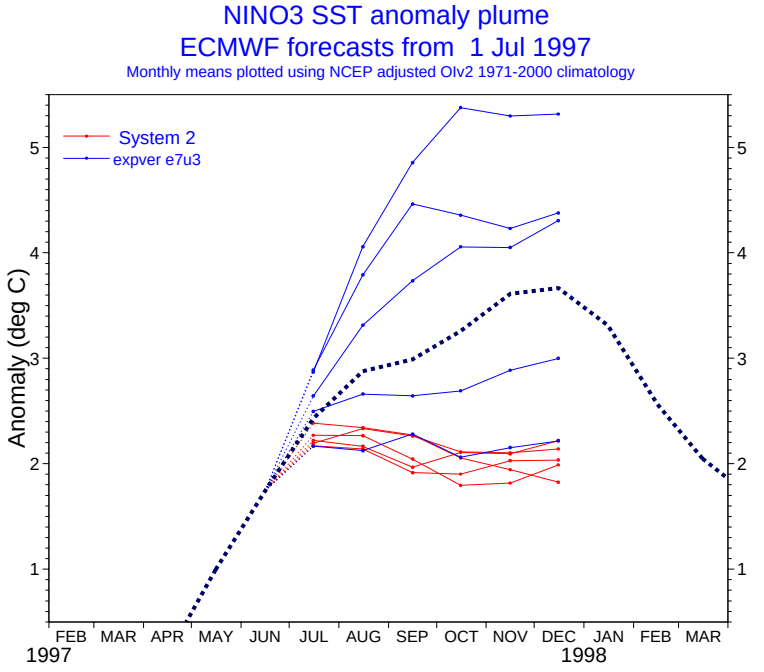
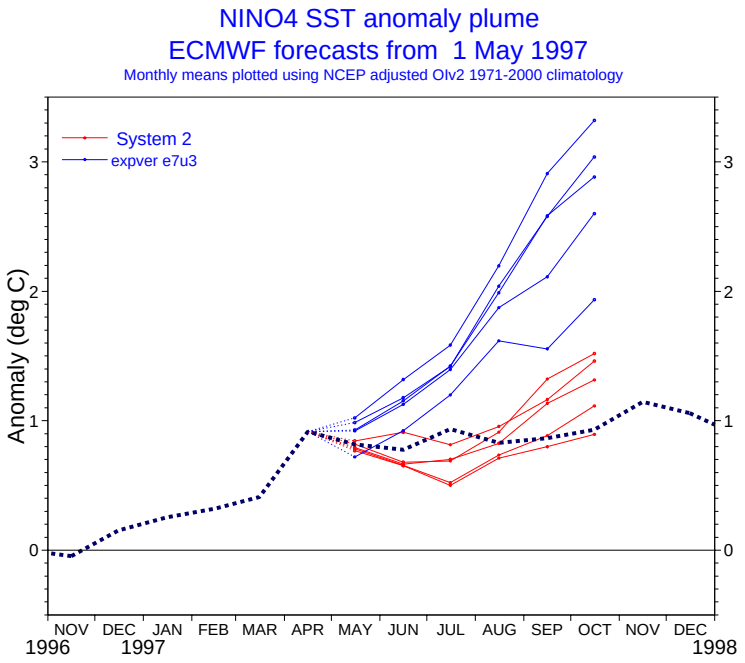


Figure 15: Plots of forecasts from S2 (red) and the Met Office (dark blue) for the Nino3 region for a) May and b) July starts in 1997.

a)



b)

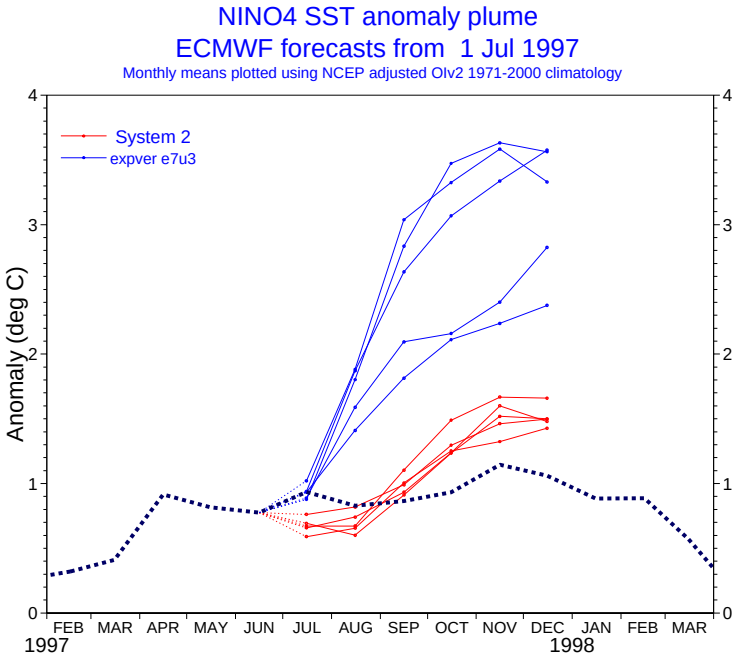
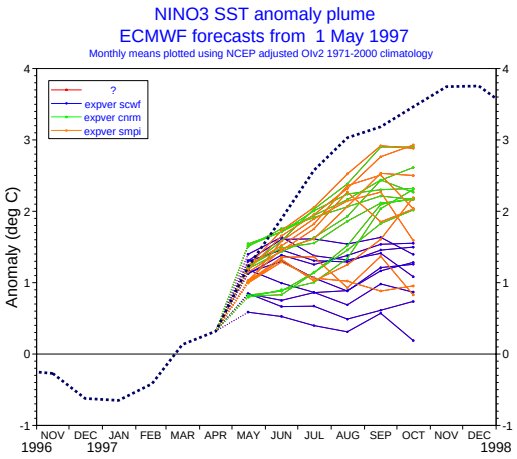
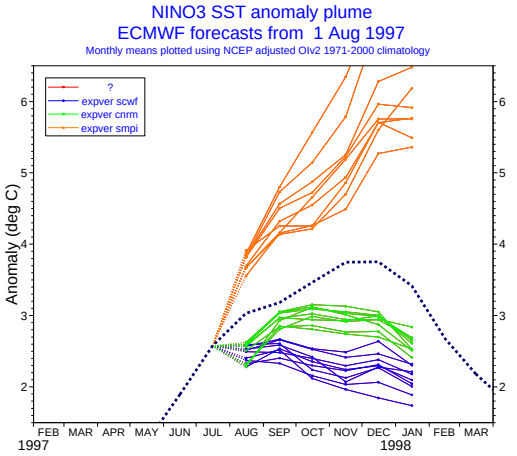


Figure 16: As for fig 15 but for Nino4.

a)



b)



c)

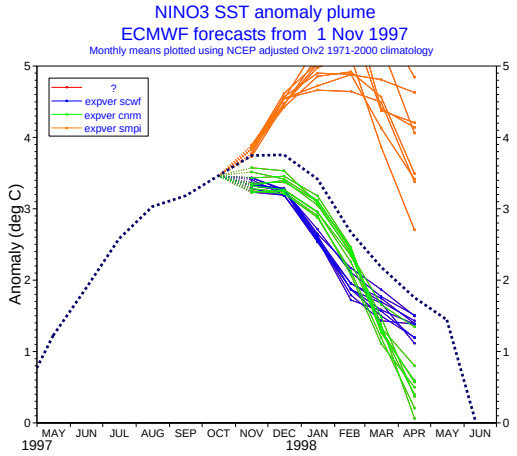


Figure 17: Plot of forecasts from the DEMETER integrations for a) May starts b) August and c) November starts. Anomalies relative to the 1987-96 period.

show experiments using A1 and A2 respectively. Changing the atmosphere has a greater effect on the drift than changing the ocean. In panel b) we show the mean absolute error as a function of model. Again, changing the atmosphere from A1 to A2 leads to a degradation though this is greater in the case of O1 than O2. Changing the ocean from O1 to O2 leads to an improvement regardless of whether using A1 or A2. If one is using the O2 ocean, then it appears that there is little difference in changing from A1 to A2. While this is true for Nino3, it is not true for Nino4: there is a degradation in using A2 with O2. The change in ocean resolution in O2 leads to a degradation. Fig 19 shows the 6 month mean absolute error for July and October starts for 1997. In this case the lowest errors are for S1. Changing the atmosphere or ocean leads to an increased error, but the change in atmosphere has a bigger effect than the change in ocean.

In the next section we will consider some aspects of the atmospheric model behaviour using coupled and uncoupled integrations.

4 Results from coupled and uncoupled integrations.

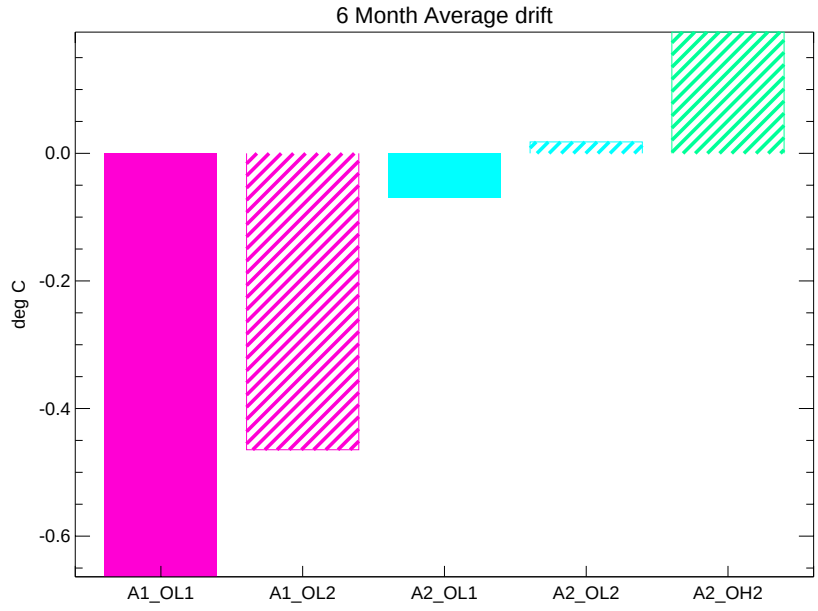
4.1 Bias in the coupled systems S1 and S2.

It is well known that any fully coupled model is likely to develop quite large biases in many fields. Our coupled systems are no exceptions. Fig 20 illustrates the bias in SST in the Nino 3 and Nino 4 regions as a function of the time of year. The full model SSTs are plotted (S1 in blue, S2 in red) together with the observed seasonal cycle (dashed line). The bias varies seasonally and is typically comparable in magnitude to the interannual signal one is trying to predict. For S1 the bias is always cold in both Nino 3 and Nino 4. S2 is always warmer than S1, and the bias is warm for part of the year in Nino 3, although for Nino 4 it still remains cold throughout the year. In both models there is a seasonal cycle in the bias, which in Nino 3 acts to damp the amplitude of the seasonal cycle. For example at 6 months lead, S1's seasonal cycle is 1 deg C peak-to-peak and S2 is only 0.6 deg C, but the observed amplitude is about 2.5 deg C. In Nino 4, the bias results in an incorrectly phased seasonal cycle in both models, with S1 having an exaggerated amplitude as well.

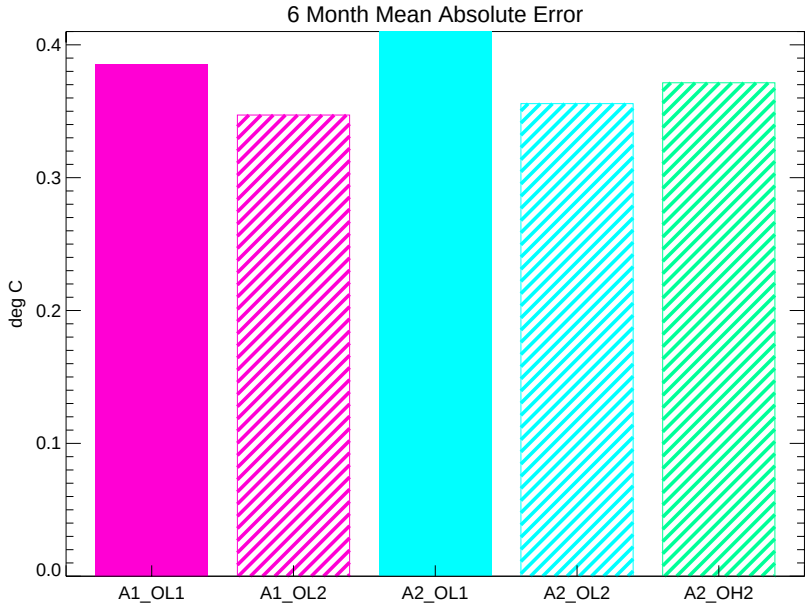
Fig 21 shows the bias in surface temperature for S1 after 1 month and after 6 months for October starts. (The reference climatology is ERA40). The bias in surface temperature over the ocean after 1 month is relatively small but after 6 months, the cold bias is present over much of the tropical Pacific. The bias is larger in a cold tongue along the equator, a feature that is seen in many coupled models. Of larger amplitude than the biases over the ocean, are those over land. They develop quickly, within the first month, and reach up to more than 6K in some regions as is very evident over Africa, South America, India and east Asia, and Australia.

The comparable drift in surface temperature for S2 is shown in fig 22. The ocean drift is less than in S1. Overall, S2 is warmer than S1. The difference in bias in land temperature is likely related to the revised land scheme used in S2. This scheme is also used in ERA40, the reference climatology. There are differences in the topography used in S1, S2 and ERA40 which will contribute to the differences ibn temperature over land.

In fig 23 and 24 we compare the drift in the surface winds in the two systems. In the first month the bias in the winds is quite noticeable in the Indian and in the Atlantic sectors. The westerly bias in the Atlantic is similar in both S1 and S2 but of opposite sign in the Indian ocean where S2 has an easterly bias in the tropics. These biases have a tendency to be present throughout the integrations though they do vary from month to month, either as a function of the season or of the forecast lead-time. A systematic easterly bias develops in the western equatorial Pacific in both coupled models. In S1 it is not present in the first month but develops later, whereas in S2 there is an easterly bias even in the first month. Easterly biases develop in the off-equatorial Pacific as well which are rather similar between the two coupled models, but as they are probably of less importance in influencing



a)



b)

Figure 18: Plot of the contributions to the SST of the various experiments in the Nino3 region to a) bias b) mean absolute error over 6 months. These statistics are for the whole period of the experiments (1987-2001).

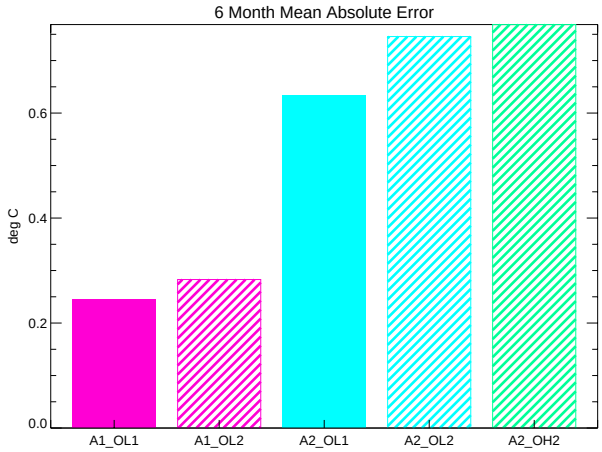


Figure 19: Similar to fig 18b but just for the July and October 1997 forecasts.

the equatorial SST evolution, we will not dwell on these here.

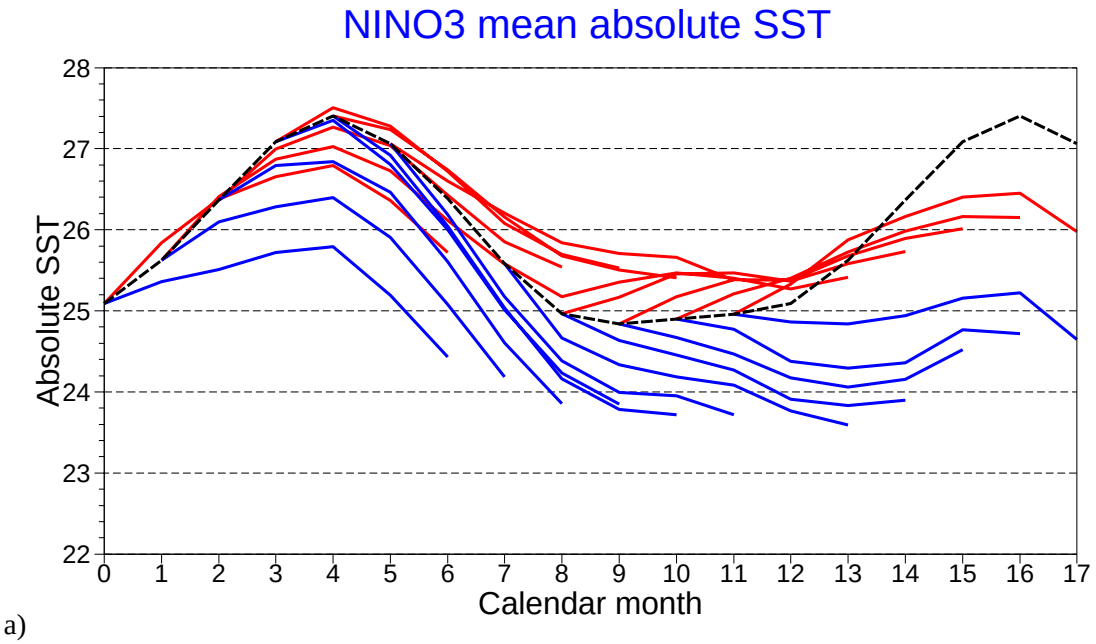
Differences are present in the precipitation in the two model integrations, but since the climatological precipitation is of somewhat debatable quality we will rather plot the difference between S1 and S2. The differences are quite large even from the first month. Fig 25 shows the differences in precipitation after 1 month and after 3 months. The main features of the tropical precipitation are the enhanced precipitation over the Indonesian region in S2, which stretches into the Indian ocean by month three, the enhanced precipitation west of the Andes and the suppressed precipitation to the east of the Andes. The banded dipolar structure that develops in the tropical Pacific suggests a shift of the ITCZ. For reference we show the Xie and Arkin climatology in fig 26 for October and December.

To avoid the criticism that there might be errors in the analysed surface temperature and 10m winds we will also plot the differences in these quantities between S2 and S1 rather than the differences with ERA40. We show in fig 27 the differences in the surface temperatures. These are substantial even in the first month. This could be due to the different treatment of land processes in the two atmospheric models. It is not a result of different initial conditions since both S1 and S2 were initialised using ERA15/OPS. There are differences in the orography used in S1 and S2.

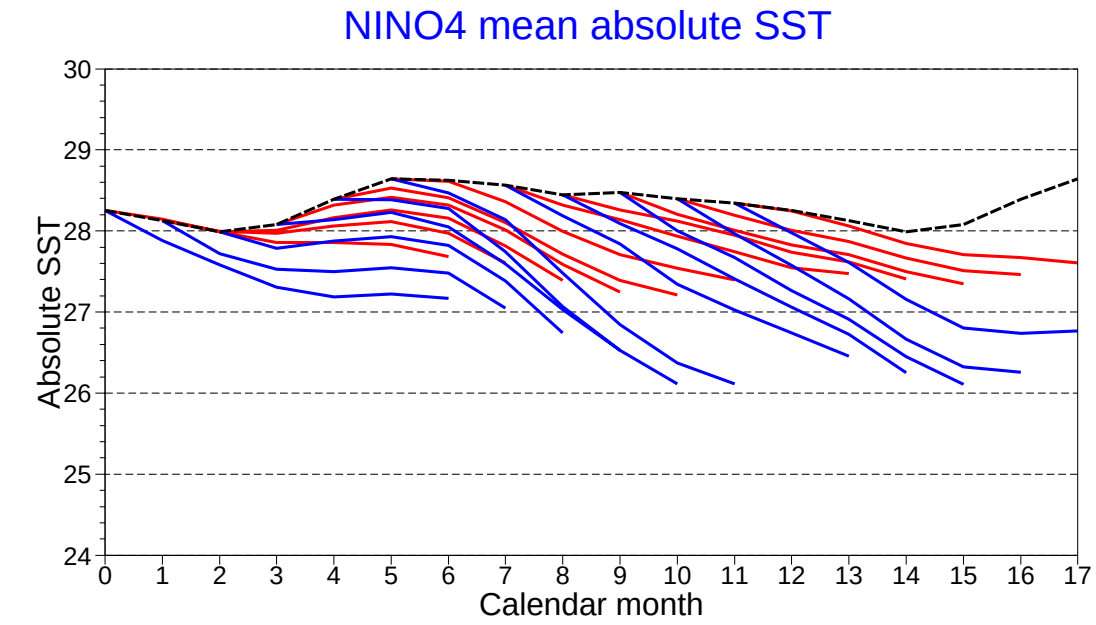
The wind differences shown in fig 28 are largest in the Indian ocean on and to the south of the equator, in the tropical Atlantic at three months and in the west Pacific, though this latter varies quite a bit throughout the 6 months of integration. There are reduced easterlies in the eastern Pacific in S2 relative to S1.

4.2 A matrix of experiments using coupled and uncoupled integrations.

The above figures highlight some of the differences between S1 and S2 in terms of the mean state. They do not indicate the cause of these biases, how much is attributable to errors in the ocean model and how much to errors in the atmospheric model, how much depends on forecast lead-time and how much on the seasonal cycle. To address the issue of how much is attributable to the atmosphere, we will consider an ensemble of integrations made using observed SSTs. We will denote these experiments as A1U and A2U. Together with the coupled forecasts from S1 and S2 a matrix of four experiments is available for comparison purposes. A 5-member ensemble of results is available for S2 for all months from 1987 with a start date of the 1st of the month. For S1 an 11-member ensemble starting on the 1st of the month is available but only for the period 1991-96. The two uncoupled integrations are for 1st October starts and comprise 11 member ensembles for the period 1991-96.



a)



b)

Figure 20: Plot of the mean absolute SST in S1 (blue) and S2 (red) as a function of time of year for the Nino3 and Nino4 regions.

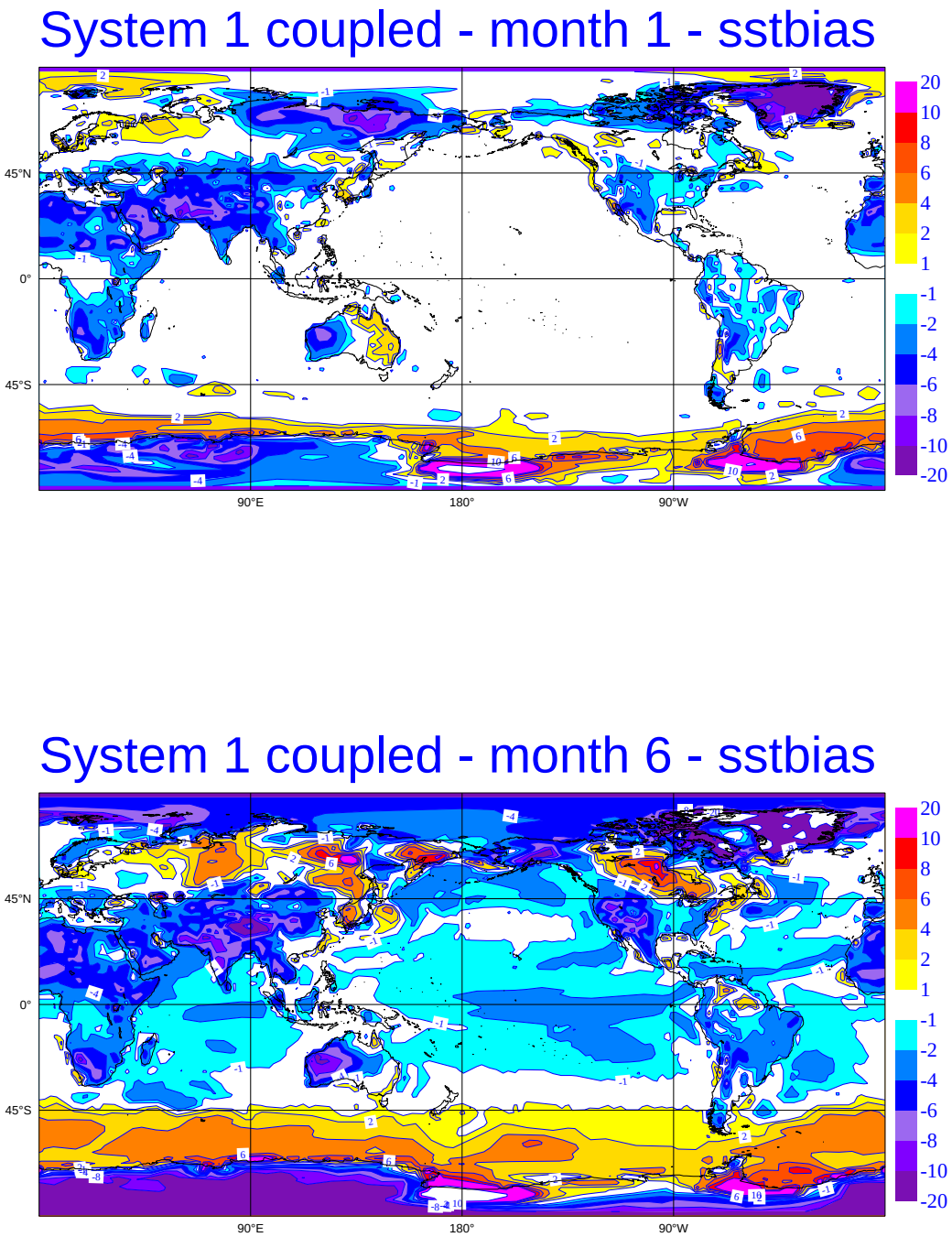
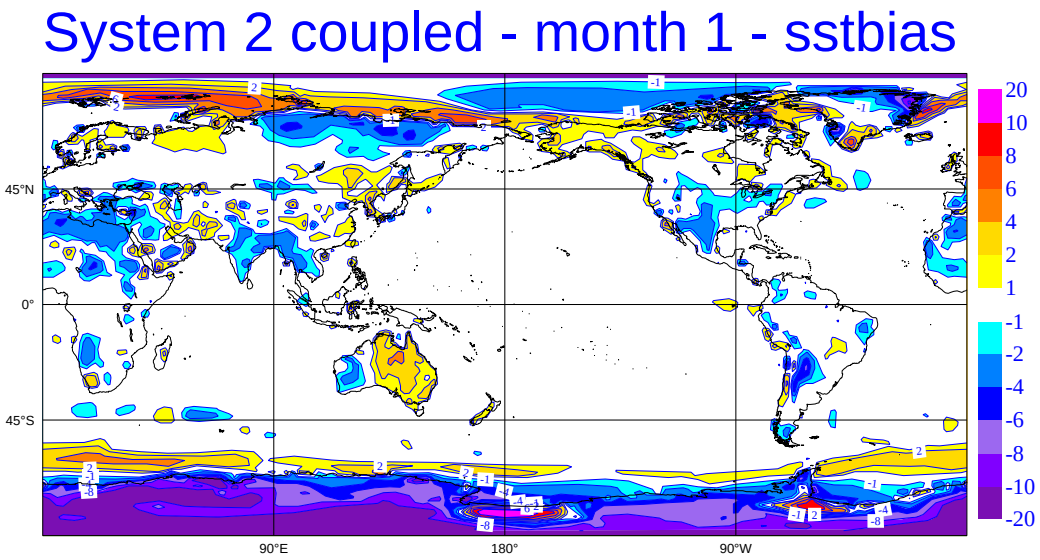


Figure 21: The bias in surface temperature over 1 and 6 months for S1. The reference climatology is ERA40.



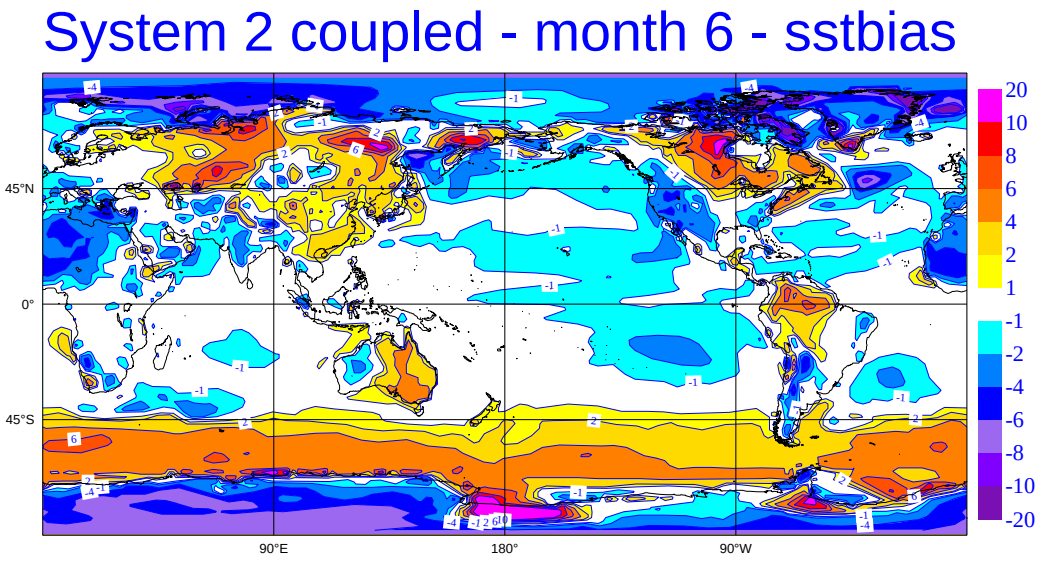


Figure 22: As for fig 21 but for S2.

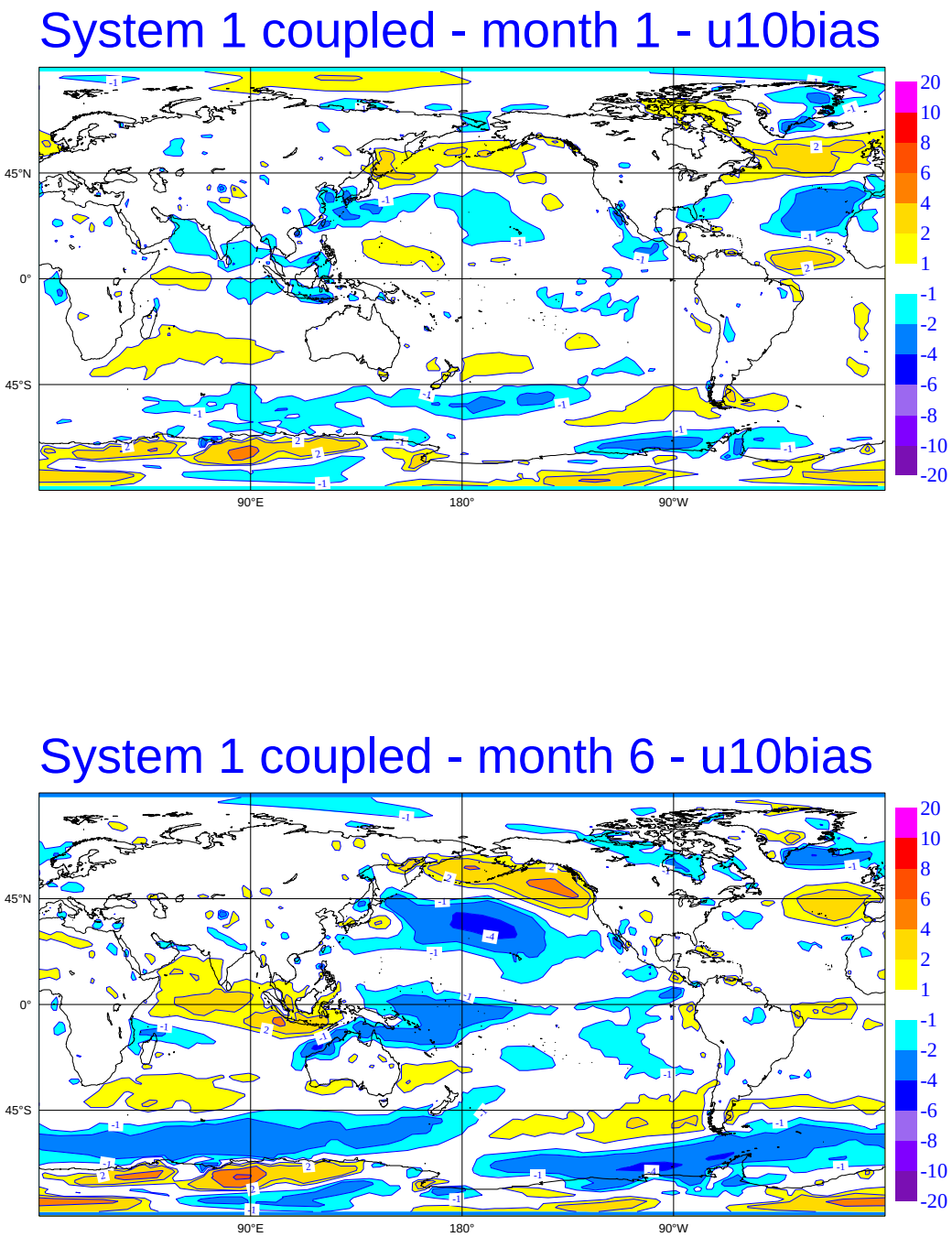
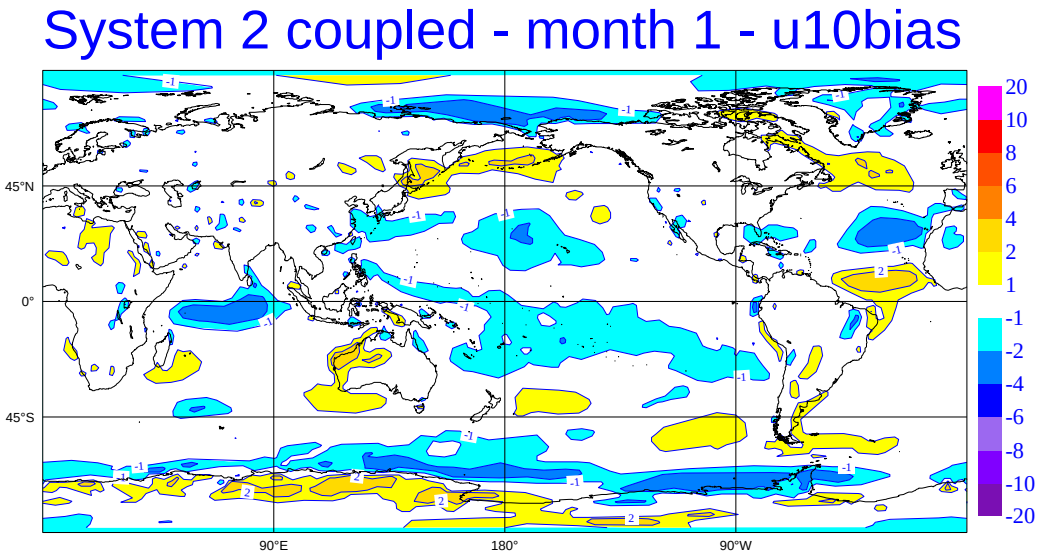


Figure 23: Plot of the bias in near-surface wind after 1 and 6 months for S1. The reference climatology is ERA40.



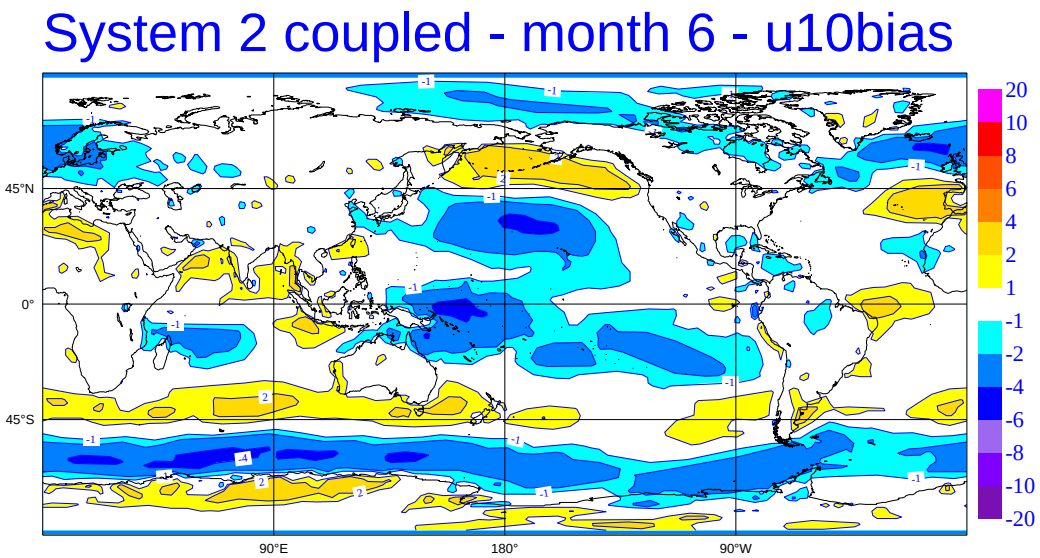


Figure 24: As for fig 23 but for S2.

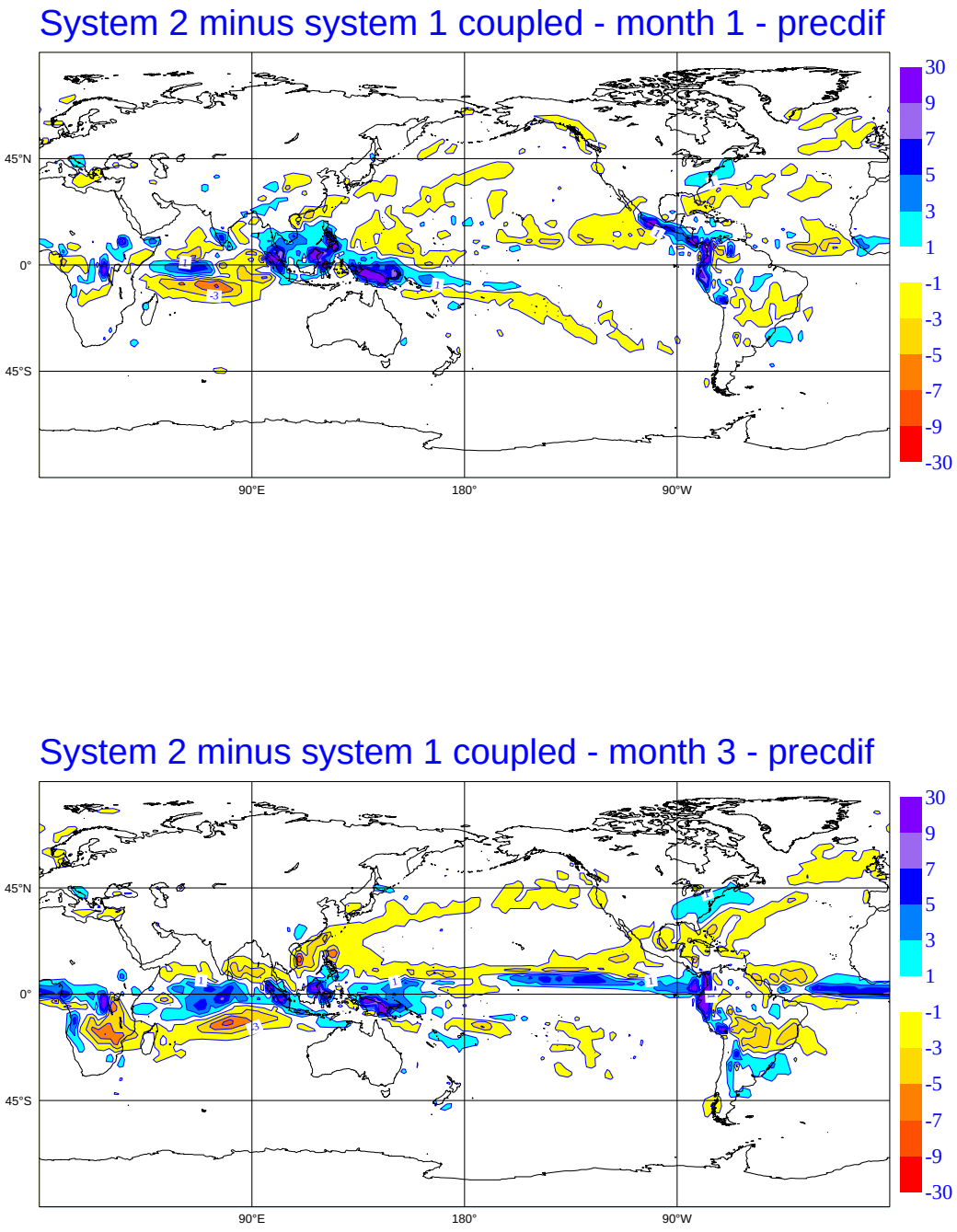
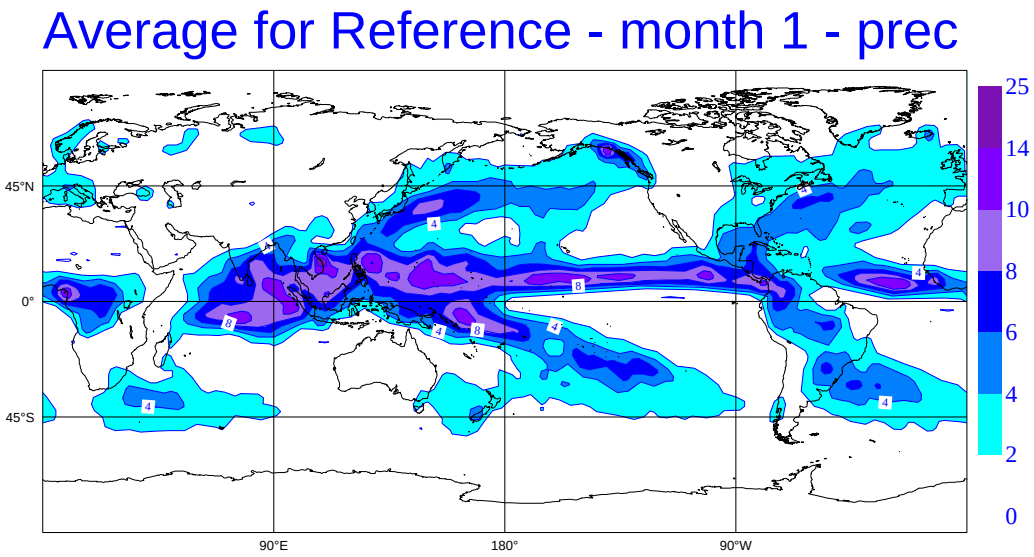


Figure 25: Plot of the difference in precipitation between S2 and S1 after 1 and 3 months.



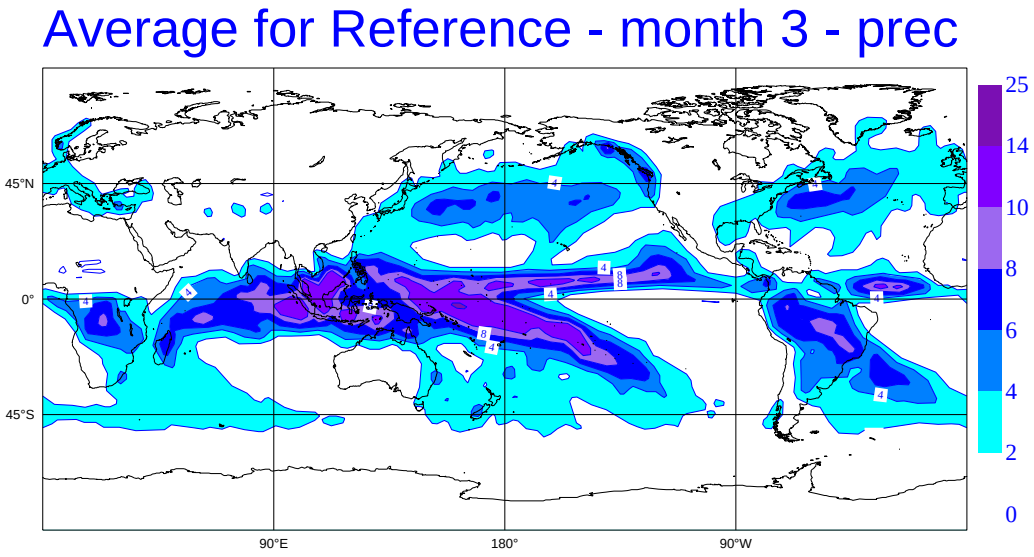


Figure 26: Precipitation from Xie and Arkin for October and December.

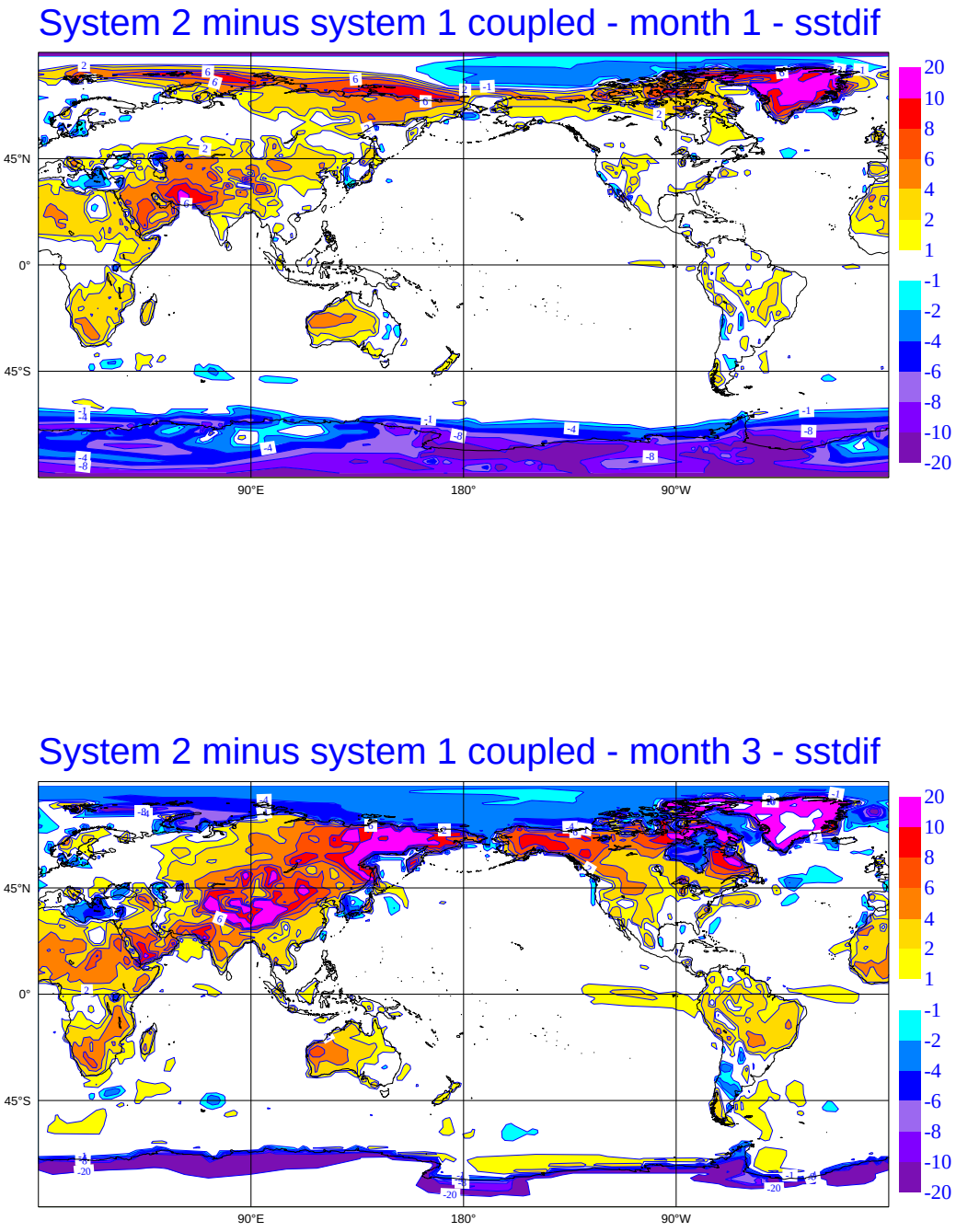


Figure 27: Plot of the differences in SST between S2 and S1 after 1 and 3 months

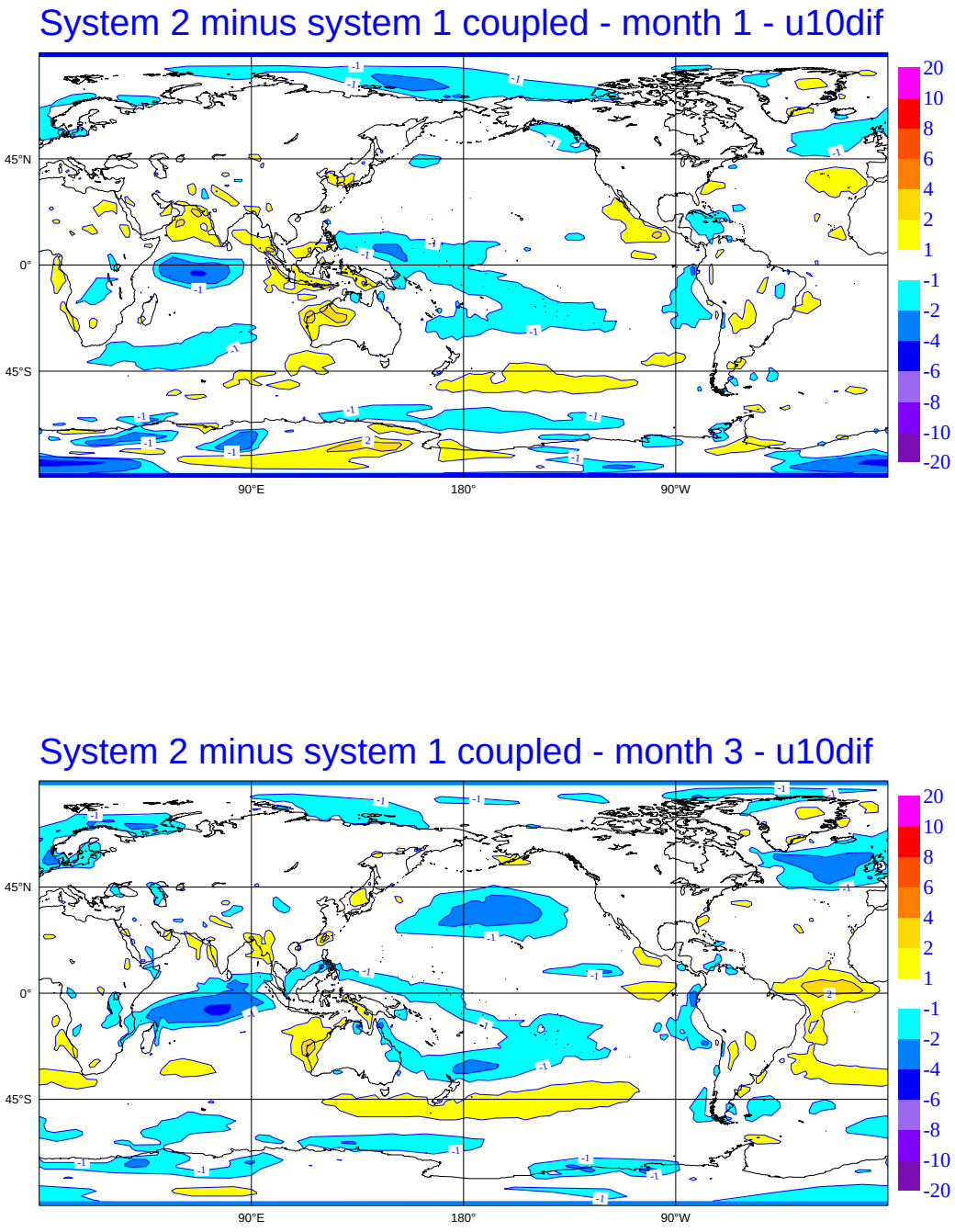


Figure 28: Plot of the differences in U10 between S2 and S1 after 1 and 3 months.

The ensemble is created by adding perturbations to the SST field. Thus the tightest comparison is for the period 1991-96 when all experiments start on the 1st of October. In order to allow a comparison between S1 and S2, we have sub-sampled the S1 ensemble to 5 members, since S2 only has 5 members. Comparison between the two uncoupled experiments A1U and A2U is based on 11 members.

In uncoupled mode, A1U is run at T63 as for S1, and A2U is run at T_L95 as for S2. Experiment A2U was run using exactly the same SST fields as experiment A1U and using ERA15/OPS for initial conditions, including soil moisture. Any differences between these experiments is then due to the atmosphere. The initial SST fields in S1 and S2 are different, but so are the ocean initial conditions since a different strategy was adopted for creating the ocean initial state. None-the-less many of the results are quite robust and are unlikely to depend strongly on these aspects.

Ocean atmosphere	A1	A2
coupled	S1(11)	S2(5)
uncoupled	A1U(11)	A2U(11)

Table 2: Summary of experiments showing the different coupled and uncoupled experiments. Numbers in brackets give the ensemble size.

4.2.1 Results from uncoupled integrations.

Fig 29 and 30 show the bias in the wind for October and December for A1U and A2U respectively. A2U has a stronger eastward bias in the Atlantic in all months and this bias develops rapidly. The eastern equatorial Pacific also has an eastward bias in S2 whereas in S1 the bias is westward. In the western Pacific the bias is also different in the two systems: A2U has a westward bias on the equator whereas A1U has an eastward bias.

Since we can not guarantee the absolute accuracy of the reference climatology, we will plot some difference maps. Fig 31 shows the same results for rainfall as fig 25 but for the difference in the uncoupled integrations. The patterns in rainfall after 1 month are very similar, indicating that the errors in precipitation seen in fig 25 are not caused by coupling to the ocean. After three months, the error patterns in precipitation are still quite similar. The dipolar structure of rainfall in the central eastern Pacific seen in fig 25 is much less evident but there is a greater reduction in rainfall in the western part of the ITCZ in S2. With respect to surface temperature, the patterns over land in figs 32 and 27 are very similar at both 1 and 3 months, again suggesting that the differences are largely of atmospheric origin and coupling plays a minor role.

It is with respect to the winds that we see considerable differences between A1U and A2U. The wind differences between A2U and A1U (fig 33) are quite similar to the differences between S2 and S1 (fig 28) at one month though perhaps larger in the western equatorial Pacific in the uncoupled case. By three months, however, there are noticeable differences, which are considerably bigger in the uncoupled integrations in the west Pacific. In the coupled integrations the largest differences are south of the equator in the Indian ocean and in the Atlantic. These same features appear in the uncoupled integrations.

4.2.2 Differences between coupled and uncoupled integrations

Some further checks can be made by comparing coupled and uncoupled integrations. In terms of even quite chaotic variables such as rainfall, the differences between coupled and uncoupled integrations are quite small in the first month in both S1 and S2 (figs 34 and 35). Differences are noticeable after 2 months and longer but the patterns are broadly similar between the two. In terms of surface temperature the differences are quite modest

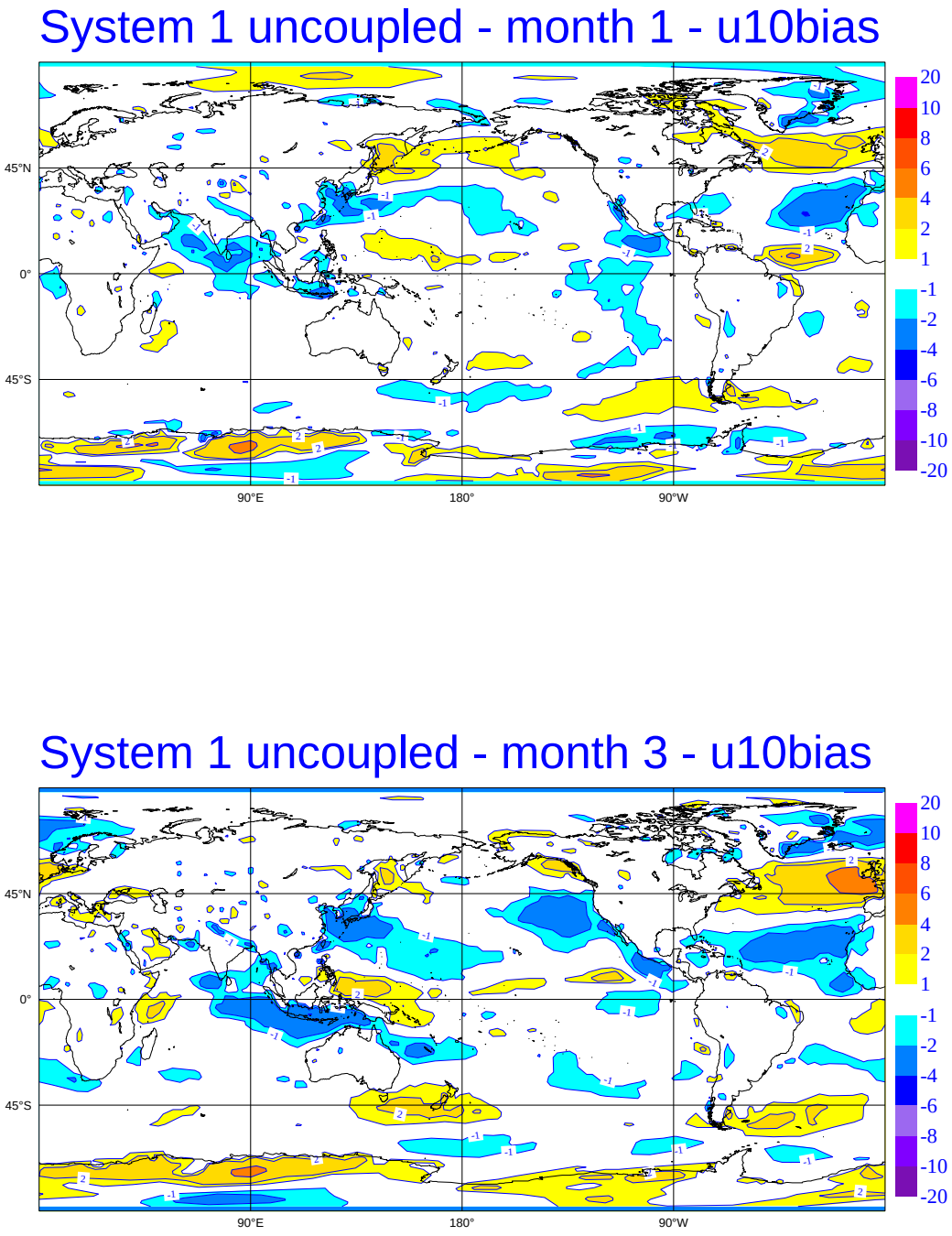
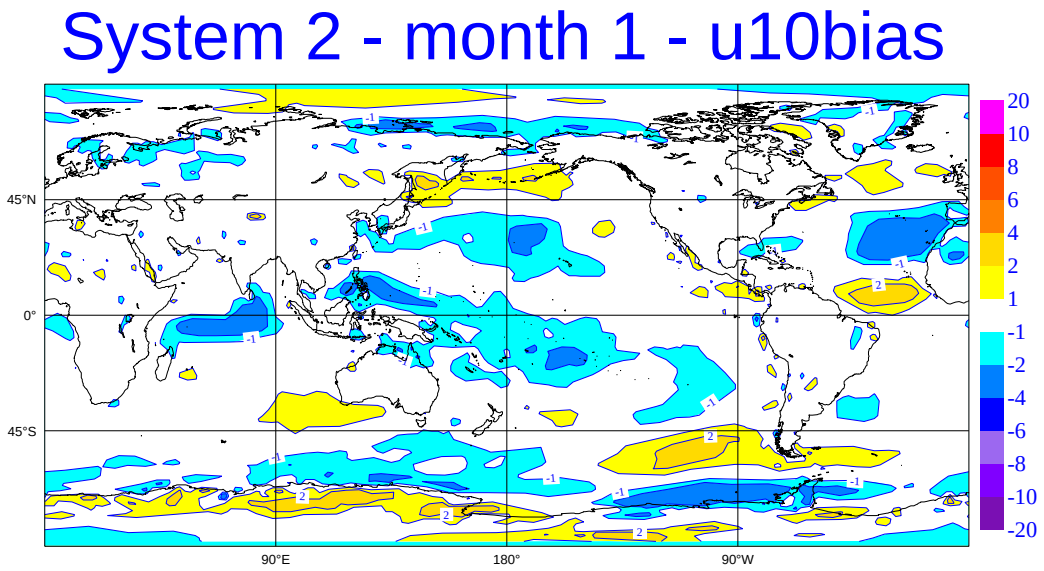


Figure 29: Plot of the bias in U10 for A1U after 1 and 3 months.



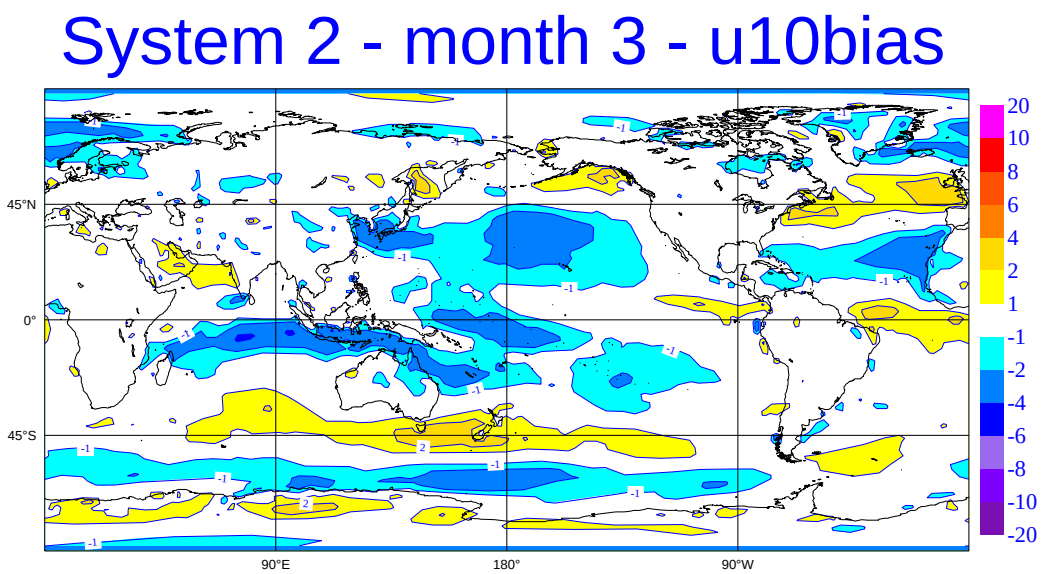


Figure 30: Plot of the bias in U10 for A2U after 1 and 3 months.

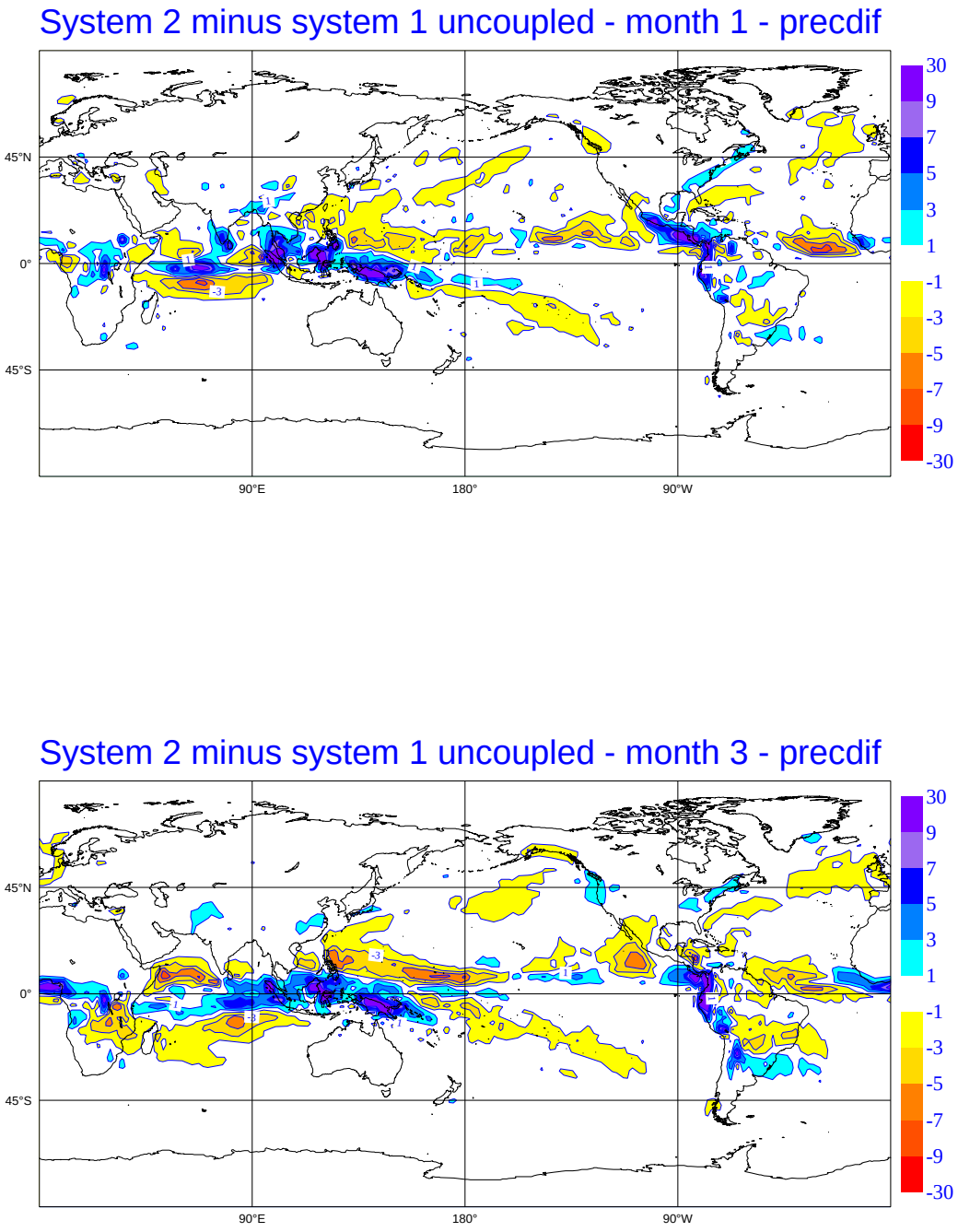


Figure 31: Plot of the differences in precipitation between A2U and A1U after 1 and 3 months.

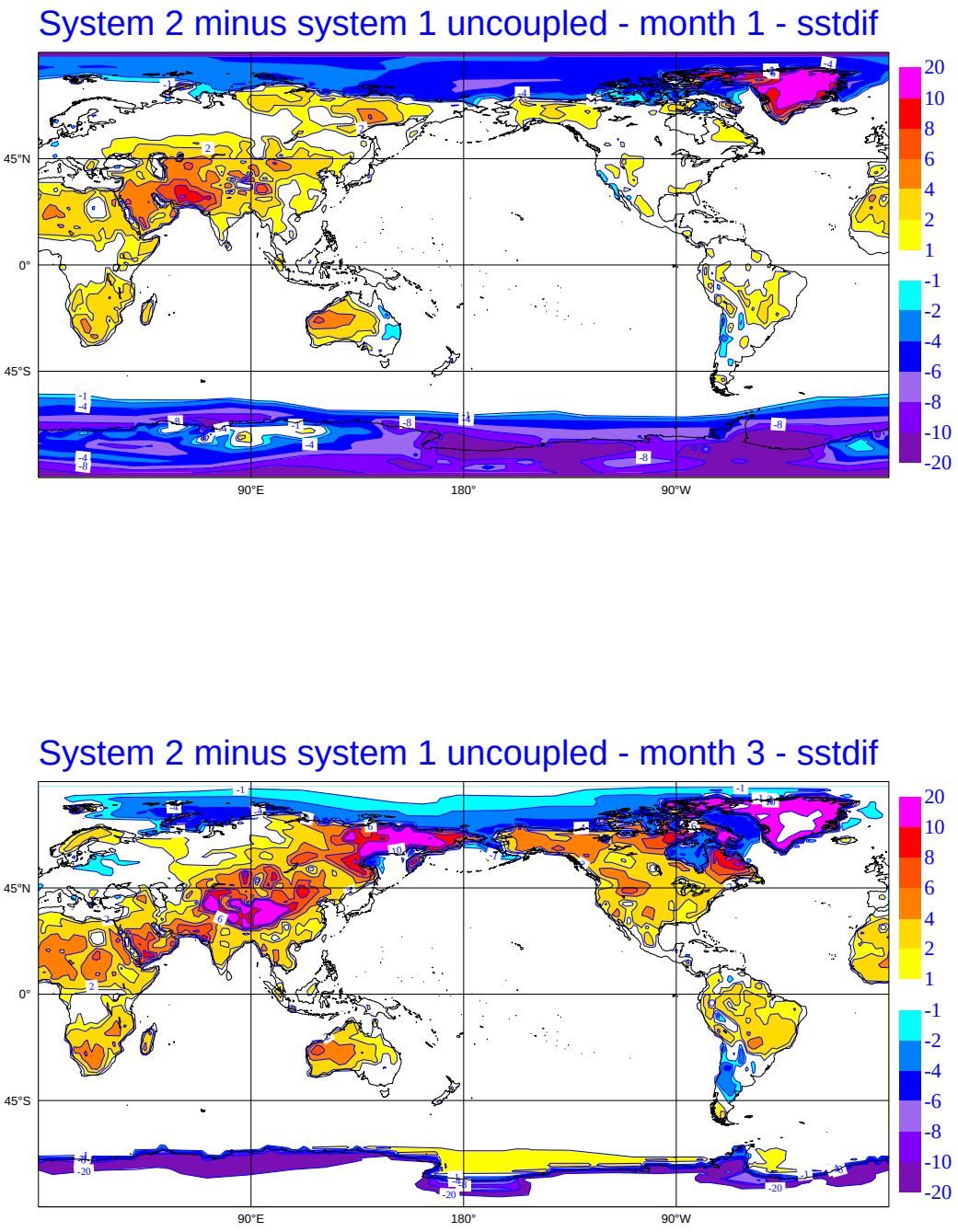


Figure 32: Plot of the differences in surface temperature between A2U and A1U after 1 and 3 months.

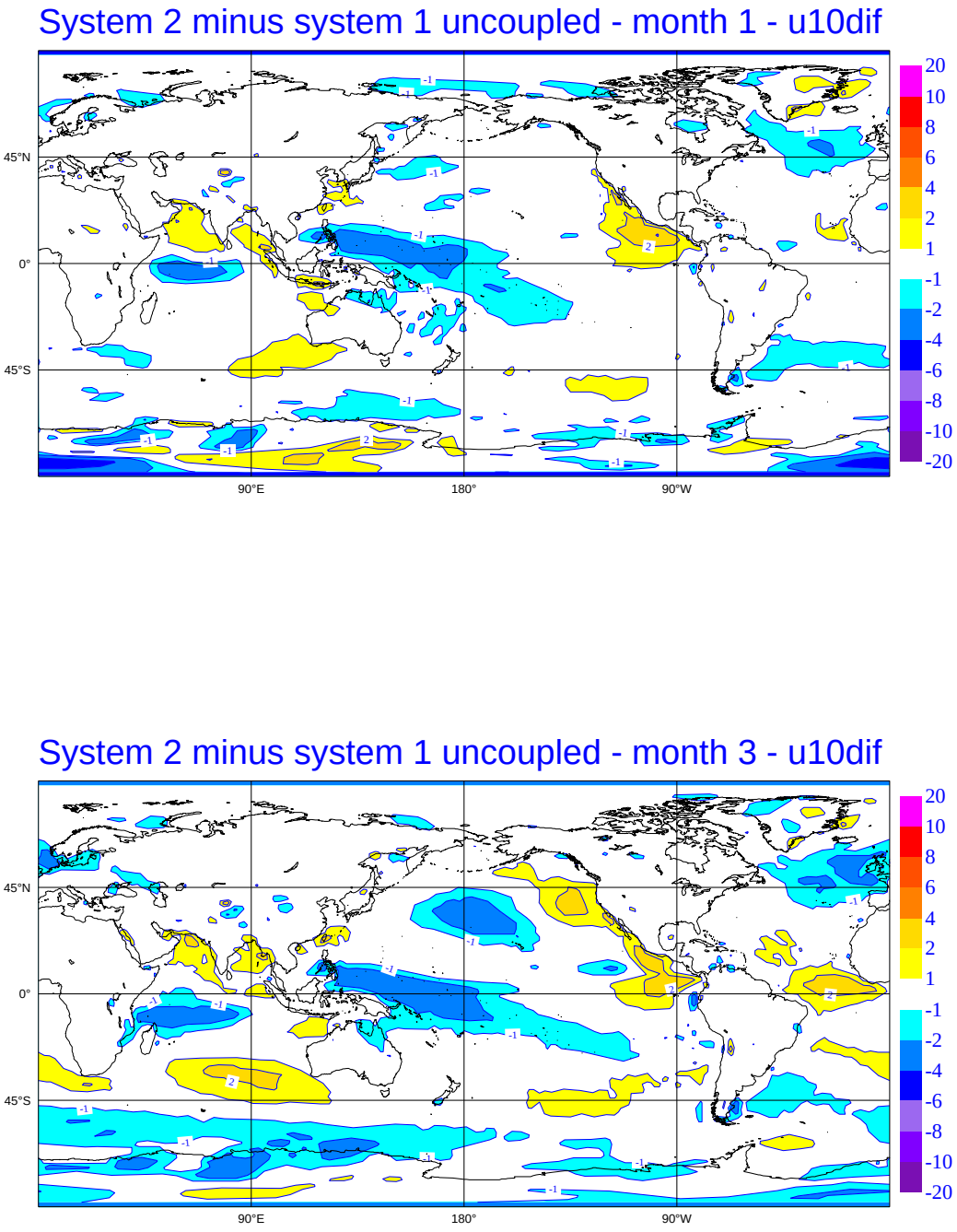


Figure 33: Plot of the differences in U10 between A2U and A1U after 1 and 3 months.

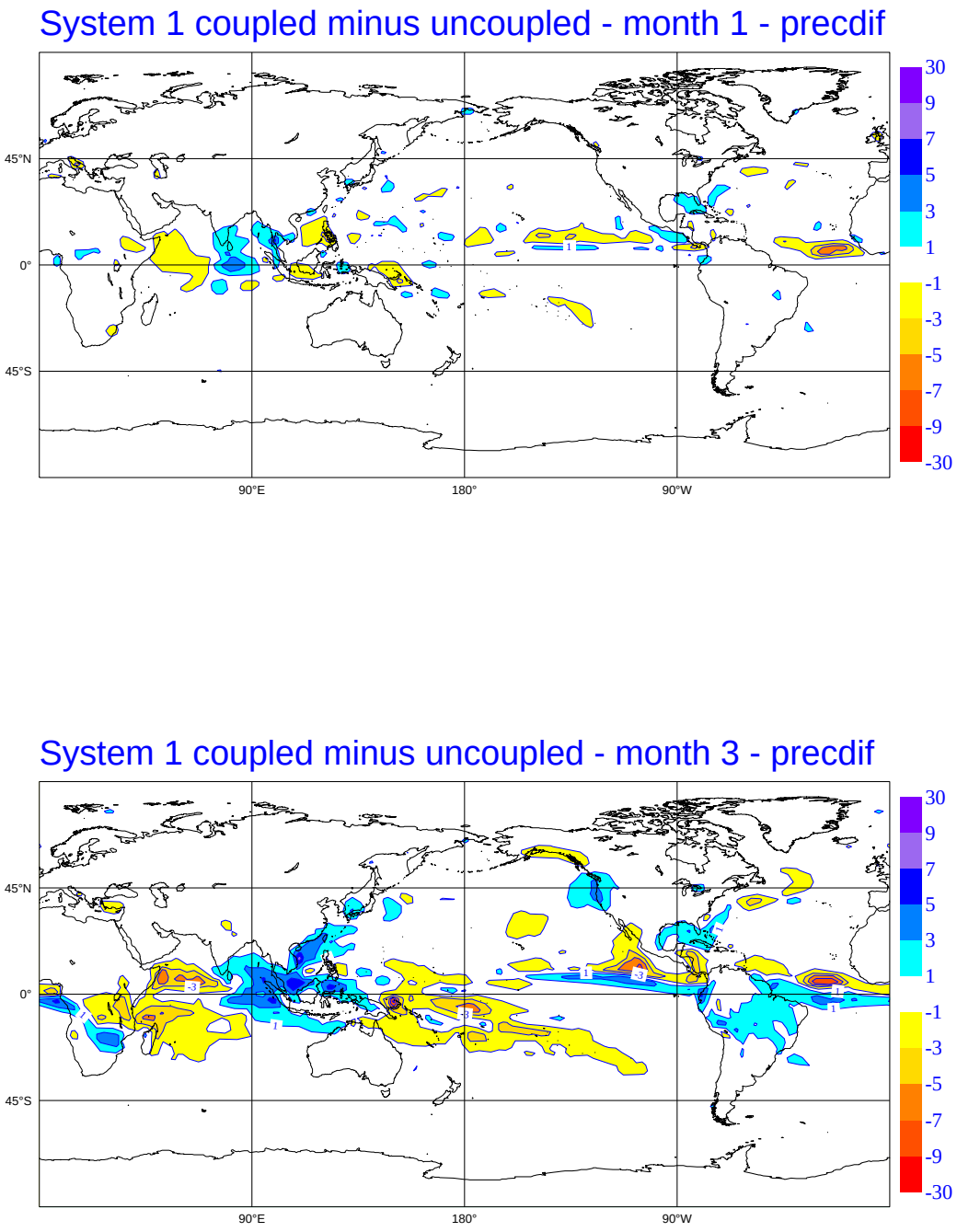


Figure 34: Plot of the differences in precipitation between coupled and uncoupled versions of A1 after 1 and 3 months.

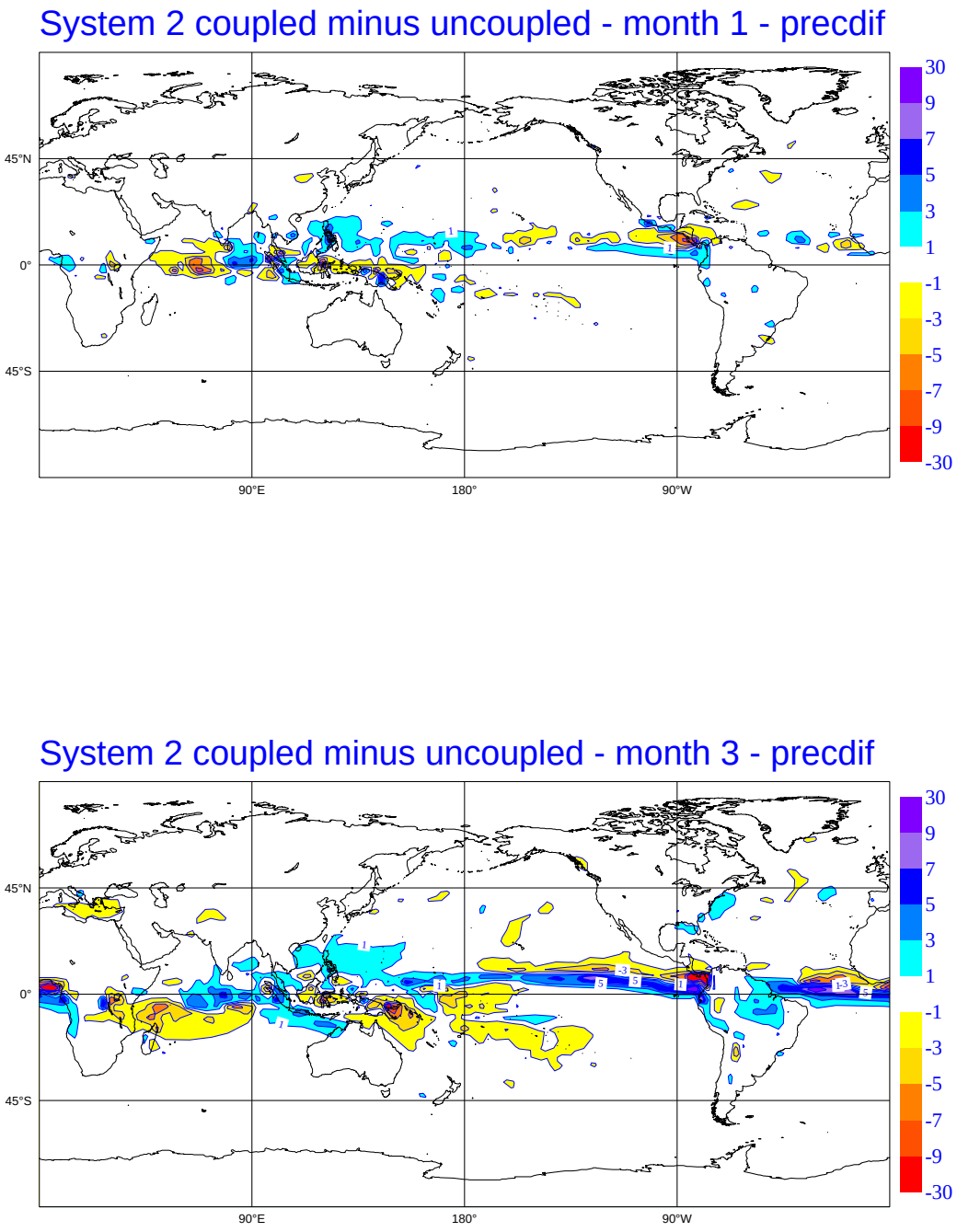


Figure 35: Plot of the differences in precipitation between coupled and uncoupled versions of A2 after 1 and 3 months.

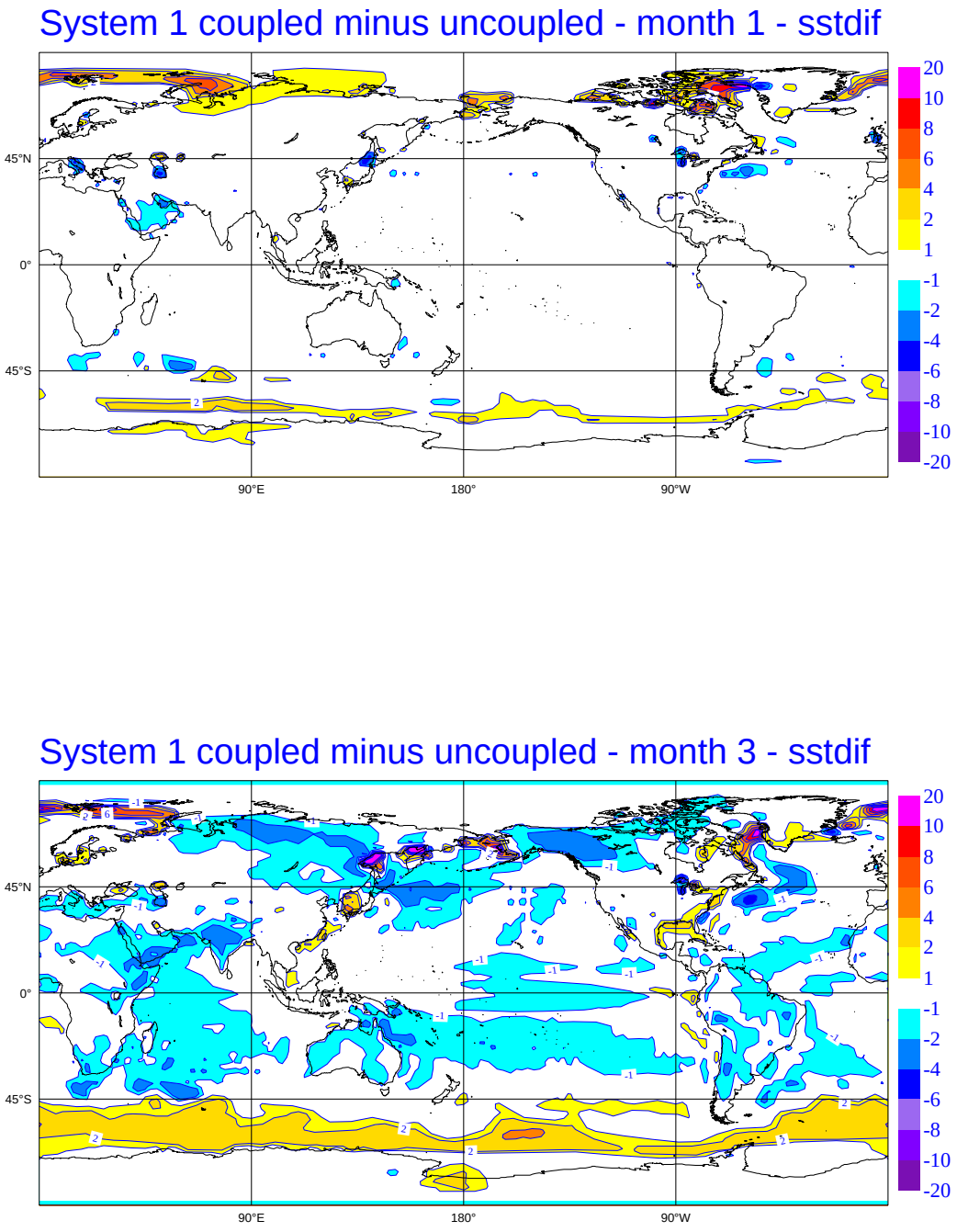


Figure 36: Plot of the differences in surface temperature between coupled and uncoupled versions of A1 after 1 and 3 months.

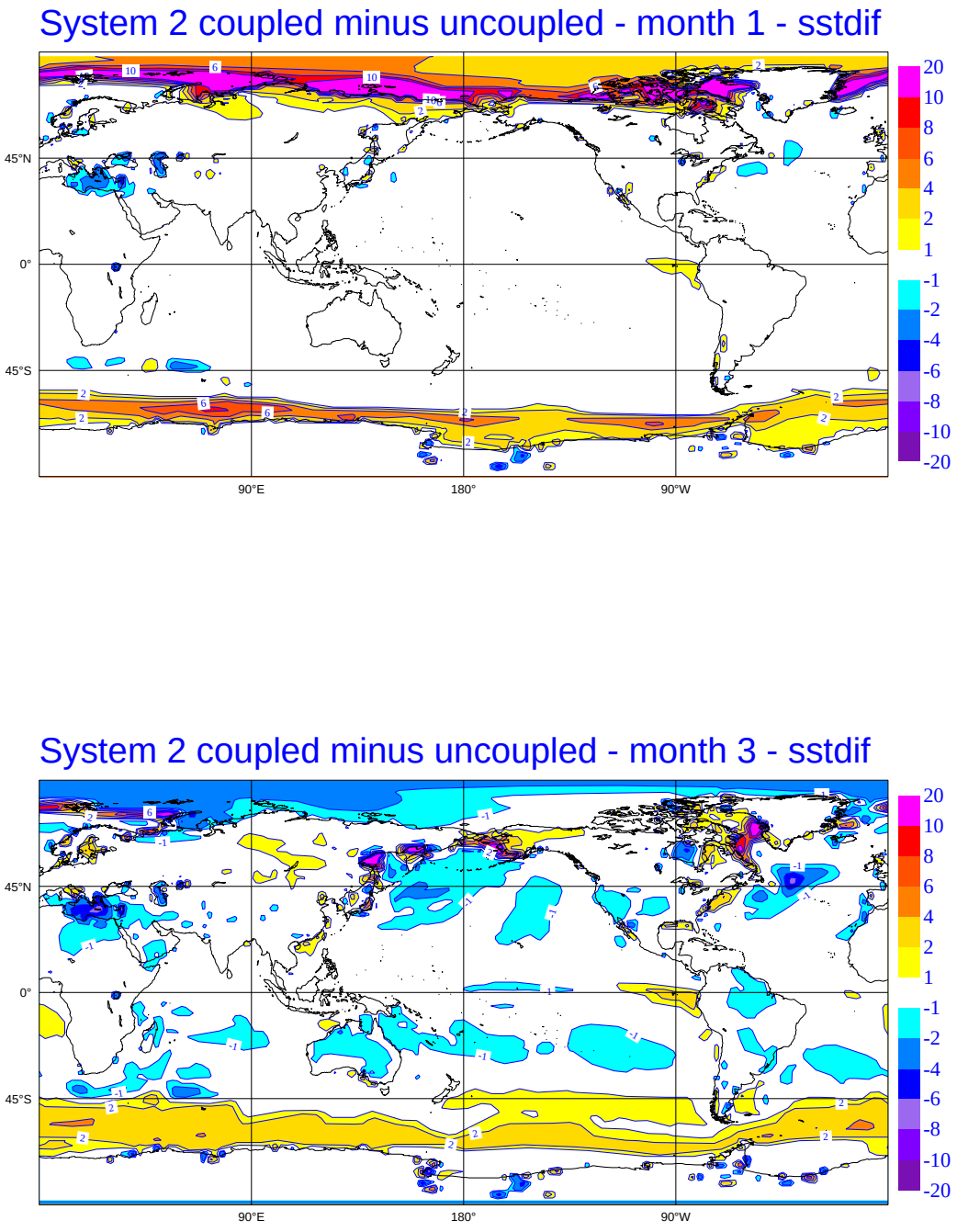


Figure 37: Plot of the differences in surface temperature between coupled and uncoupled versions of A2 after 1 and 3 months

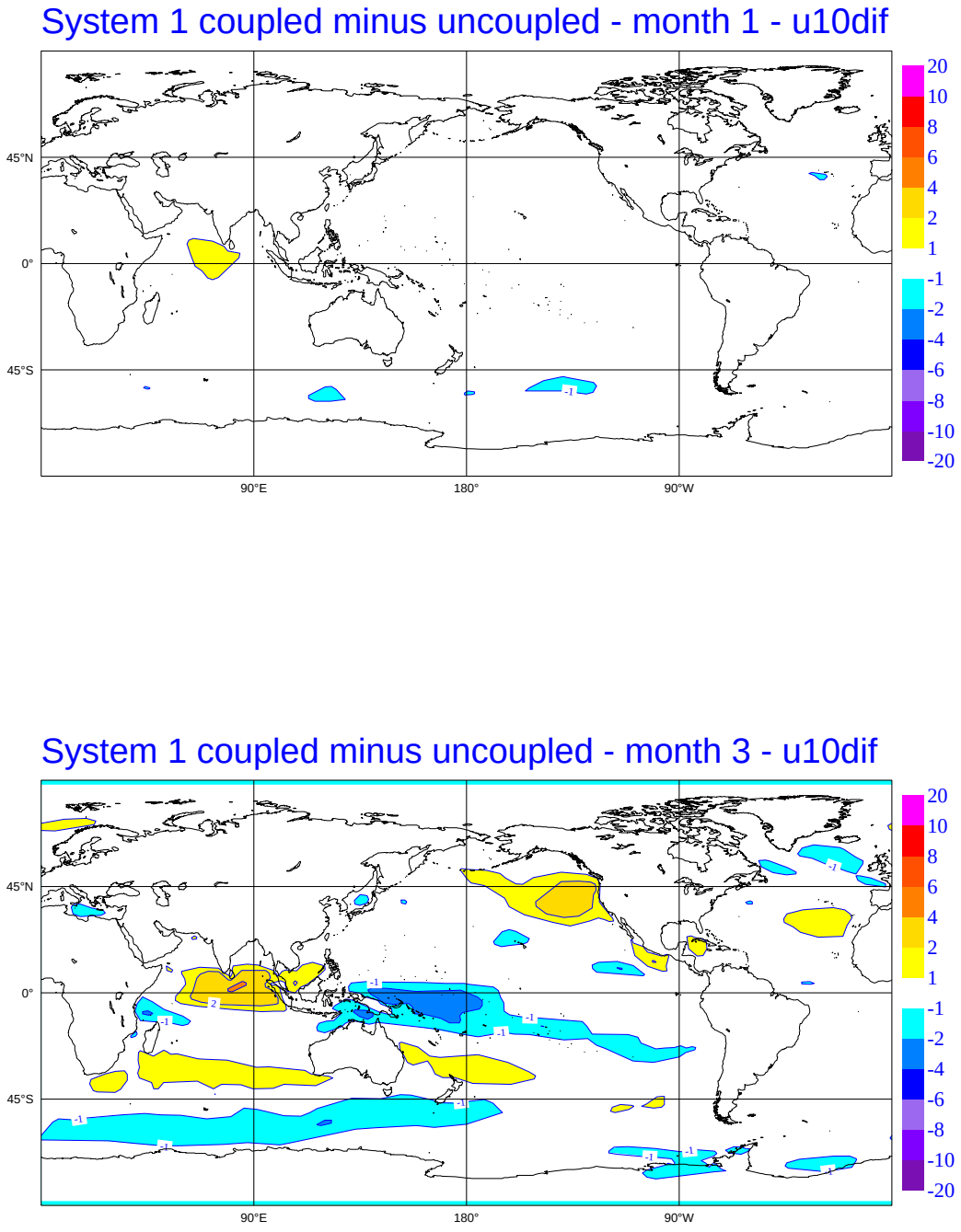


Figure 38: Plot of the differences in U10 between coupled and uncoupled versions of A1 after 1 and 3 months.

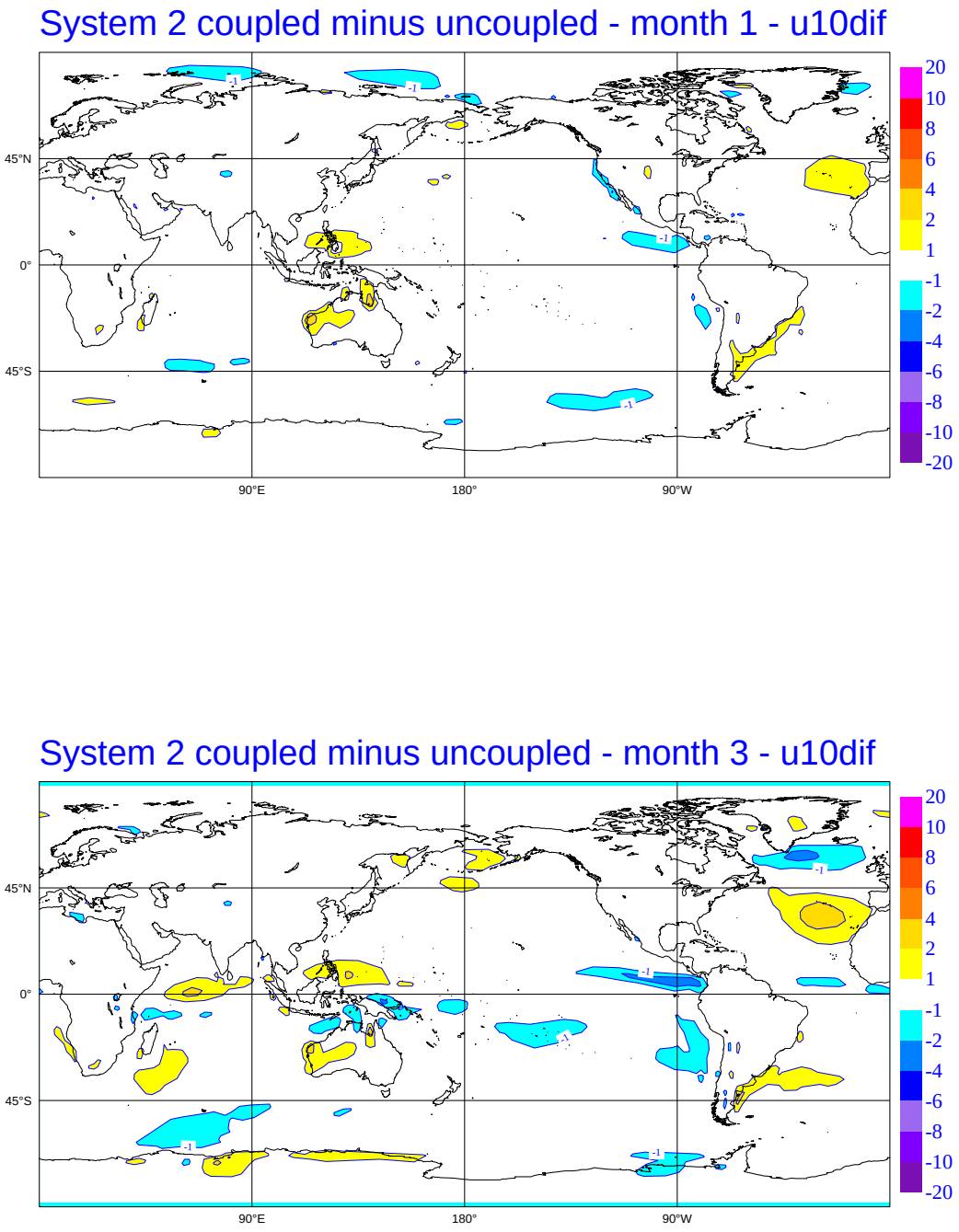


Figure 39: Plot of the differences in U10 between coupled and uncoupled versions of A2 after 1 and 3 months.

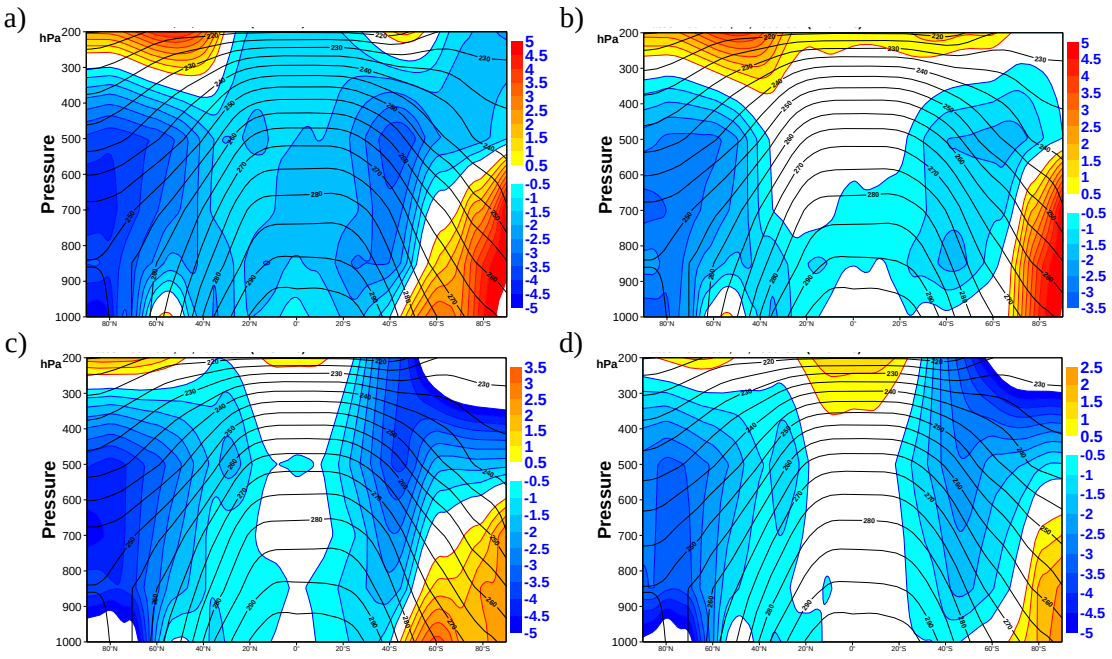


Figure 40: Error in the zonal mean temperature as a function of height for months DJF for a) S1, b) A1U, c) S2 and d) A2U. The errors are similar between coupled and uncoupled versions of the atmospheric model showing that they are largely of atmospheric origin. The errors are smaller near the ground in the uncoupled case since they are tied to the observed SST in this case.

for both S1 and S2 especially over land where the biases were quite large (figs 36 and 37). This confirms that the main cause of these errors is not coupling to the ocean.

If we look at the difference in the evolution of the wind error in S2 and A2U, we see that it takes 2 to 3 months before the drift is established. The differences between coupled and uncoupled are smaller for S2 than for S1, especially in the west Pacific (figs 38 and 39) .

4.3 Errors in the upper air

In the previous section we considered the errors in various surface fields. These are the most important in terms of interaction with the ocean but it may be of interest to know the vertical extent of model errors. We have considered vertical sections of zonal mean quantities as well as vertical sections along the equator. In the case of the coupled forecasts these are errors for months 3-4-5. The start date is 1 Oct. Fig 40 shows vertical sections of T for S1, A1U, S2 and A2U. In each case the averages are over 4 member ensembles for the period 1991-6 and for the months DJF. Comparing the coupled and uncoupled experiments, one can see a very strong similarity away from the tropics. The dominant error is clearly of atmospheric origin. In the tropics, the cooler SSTs in the coupled experiments result in a cooling of the whole troposphere – this is particularly noticeable in the case of S1. Near the surface the error might be expected to be smaller in the case of the uncoupled integrations. It will not be zero, however, since temperature over land can show a systematic error even if the temperature over the ocean has no error. Near the surface, the coupled integrations do show a slightly larger error than the uncoupled, as expected. This error can propagate vertically in the tropics, presumably aided by deep convection. The error near the surface is largest in the southern hemisphere near 60S, but this is probably not important for understanding errors in the equatorial Pacific SST such as Nino 3. It is also very clear that the errors in A1U and A2U have some broad similarities: too cold throughout much of the

extratropical troposphere. The degree of warming in the vicinity of Antarctica is different, as is the warming in the upper troposphere/lower stratosphere. The upper air errors develop first near 500hPa within the first month and spread.

With respect to winds, a vertical section along the equator showed that the differences between A1U and A2U are quite large in all three oceans. Differences between coupled and uncoupled are large when using A1 but more modest in the case of A2.

4.4 Wind variability

For the purposes of predicting interannual variability, it is not sufficient to know the systematic error: it is also important to know how well the atmospheric model generates wind variability. Errors in the mean and errors in the variability may be related but they do not have to be. In section 3 we showed that the wind variability in S2 was less than in S1 and the rapid collapse in wind variability suggested that this was an atmospheric rather than an oceanic problem. In this section we will consider spatial patterns of wind variability, filtering frequencies higher than 1 month. Monthly means of the wind data were produced and the variances were calculated. For the model results ensemble averages of the variance (not variances of the ensemble averages) were then produced. These are shown in fig 41 for December. The variance of the wind for A1U and A2U and the same quantity calculated from ERA40 are shown in fig 41.

In the analysis, the peak variance in the equatorial Pacific in December is close to the date line, though it does move latitudinally and longtitudinally with the seasons. There is considerable variance stretching south eastwards along the South Pacific convergence zone. The variance for A1U is close to the peak value of the analysis and does extend eastwards along the equator though the extension down the south Pacific convergence zone is too weak. The variance in the ITCZ in the northern hemisphere is too strong. The pattern in the Indian ocean north of the equator is a poor match for the observed. When we examine the variance for A2U we see that it is much weaker than for A1U and much weaker than ERA40. The maxima are 8 m^2s^{-2} for ERA40, 10 m^2s^{-2} for A1U but only $5m^2s^{-2}$ for A2U. The maximum also lies a little too far west in S2. It has similar deficiencies to S1: viz. too little variance in the SPCZ and northern Indian ocean, but too much in the northern Pacific ITCZ.

4.5 Extratropical differences

An essential requisite for a seasonal forecasting system is its ability to simulate the observed statistics of the atmospheric circulation. This section briefly documents the performance of S1 and S2 with respect to atmospheric variability during Northern Hemisphere winter. Additionally we use the ensembles of atmospheric simulations driven by observed Sea Surface Temperatures to analyse the effect of the coupling on the extratropical circulation and to document the performance of the atmospheric component. These experiments have already been introduced as A1U and A2U.

Since no flux corrections are used in the coupling, a substantial climate drift occurs during the seasonal integrations. Currently, the mean error is subtracted a posteriori by comparing the results of a particular ensemble with statistics from the model climate. This procedure is implicitly based on the assumption of a quasi-linear behaviour of the atmospheric and oceanic anomalies. For the extra-tropical atmosphere, the importance of non-linear effects on the dynamics of large-scale anomalies is well established. It is therefore important to assess the ability of the coupled system to represent non-linear dynamical aspects affecting large-scale atmospheric systems. By including in our analysis an ensemble of atmospheric simulations it is possible to address the climate drift effects on the Northern hemisphere winter variability.

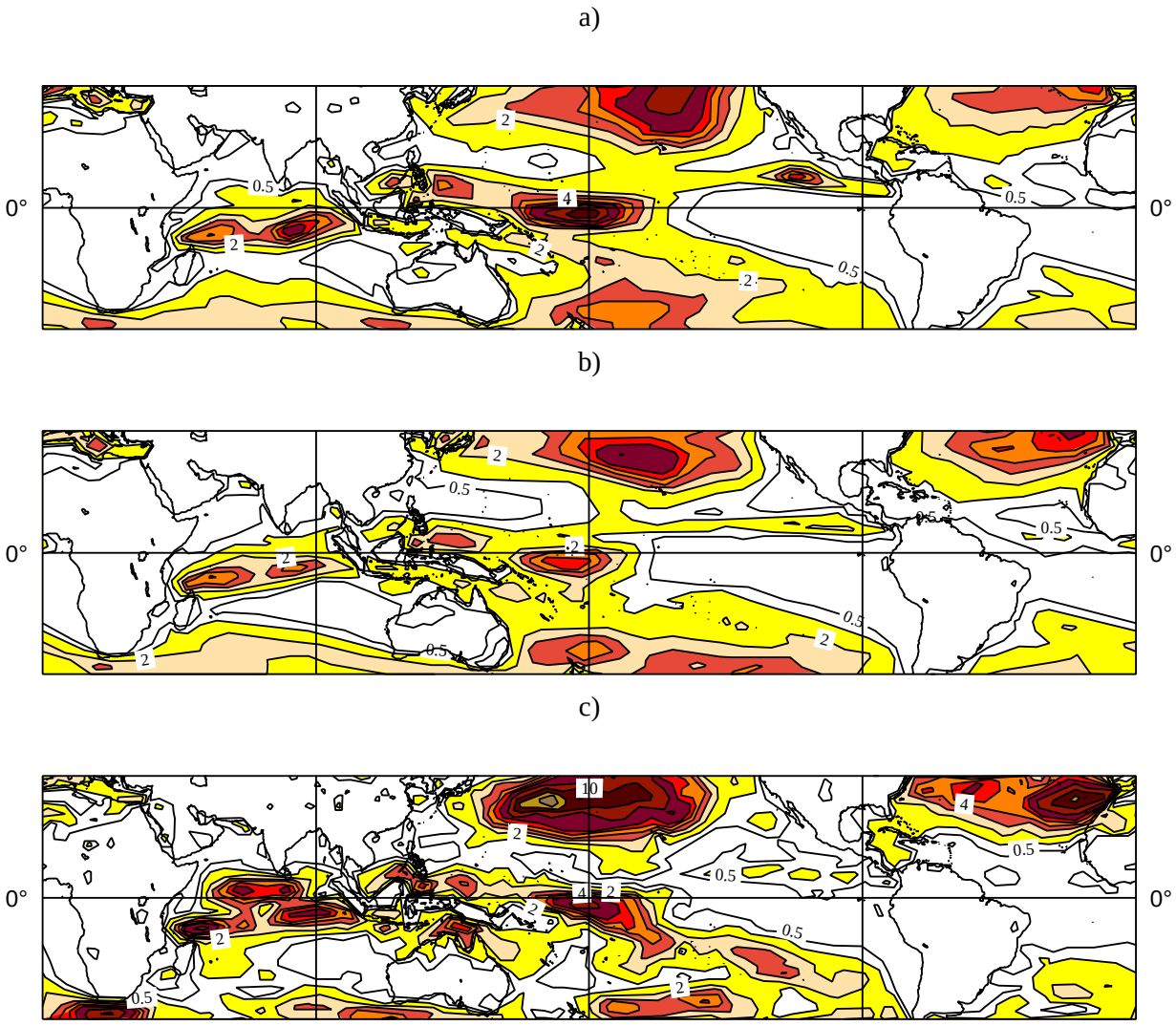


Figure 41: Plot of the variance of the 10m winds in experiments A1U, A2U and ERA40 for December. The contours are 0.5, 1, 2, 3, 4, 5, 6, 8, 10, 12, 14, 100 $m^2 s^{-2}$.

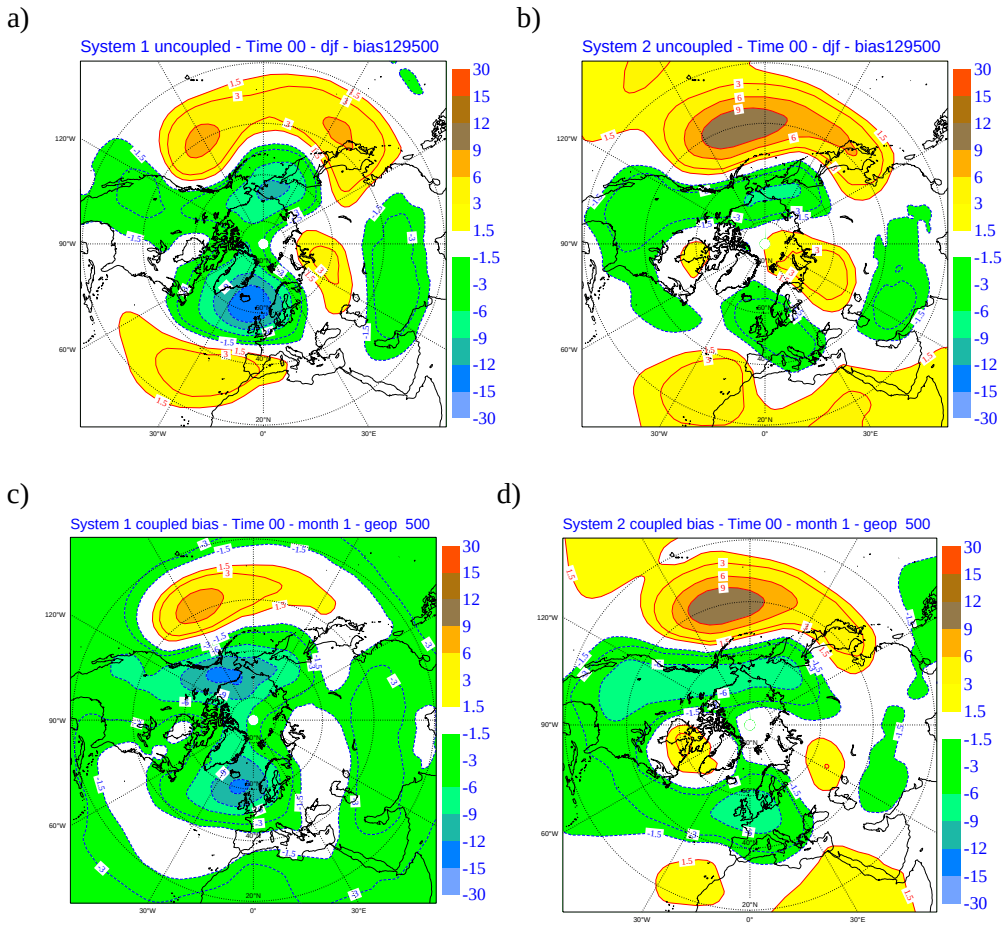


Figure 42: 500 hpa geopotential height mean error for DJF (months 3-5 of integration) based on simulations for winter periods 1991-1996 based on: a) atmospheric simulations with cy15r8 b) atmospheric simulation with cycle 23r4; c) coupled system 1 d) coupled system 2.

4.5.1 Mean error

Fig 42 shows the mean bias for Z500 for DJF 91 to 96 for the atmospheric simulations A1U and A2U (panels a and b) and for the two coupled integrations S1 and S2 (panels c and d). The bias from the atmospheric simulations A2 is smaller than A1 over the European sector. Over the Pacific sector, however, the excessively zonal flow is present in both versions. Comparison of A1 with S1 and A2 with S2 shows that the major features of the mean error in the Northern Hemisphere are similar between the coupled and uncoupled simulations indicating that the most of the error is of atmospheric origin. The negative error of S1 associated with the global cooling of Sea Surface Temperature (SST) is improved in S2.

4.5.2 Hemispheric Variability

The standard deviations of 500 hPa geopotential height monthly means for 1991 to 1996 (Fig43) from the seasonal forecast and for the uncoupled simulations are used as a measure of the inter-annual variability. During this period the extent of variability of Equatorial Pacific SST was rather limited and so Fig43 gives only an

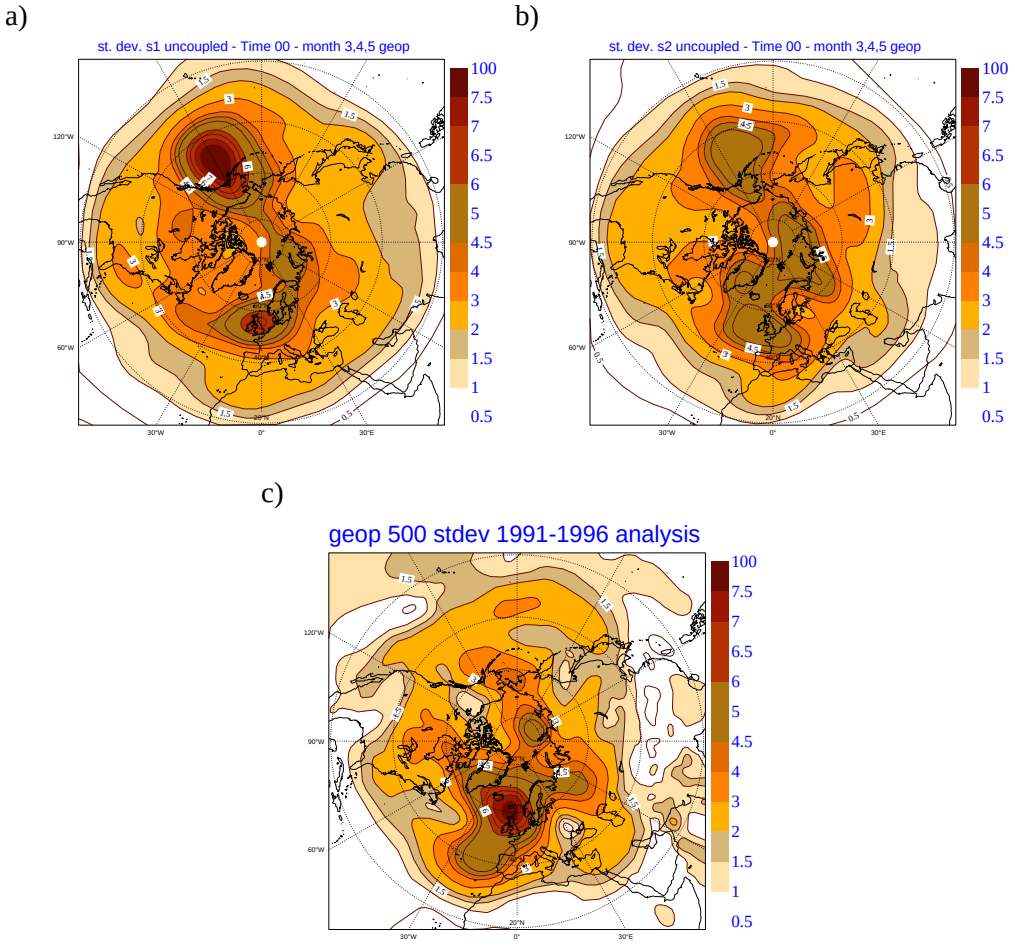


Figure 43: Standard deviations of interannual variability of the seasonal mean 500 hPa geopotential height for DJF 1991-96. a) uncoupled runs with cy15r8 b) uncoupled runs with cy23r4 c) analysed values. For the model runs, the standard deviation is calculated for each ensemble member separately, and then the ensemble mean is taken.

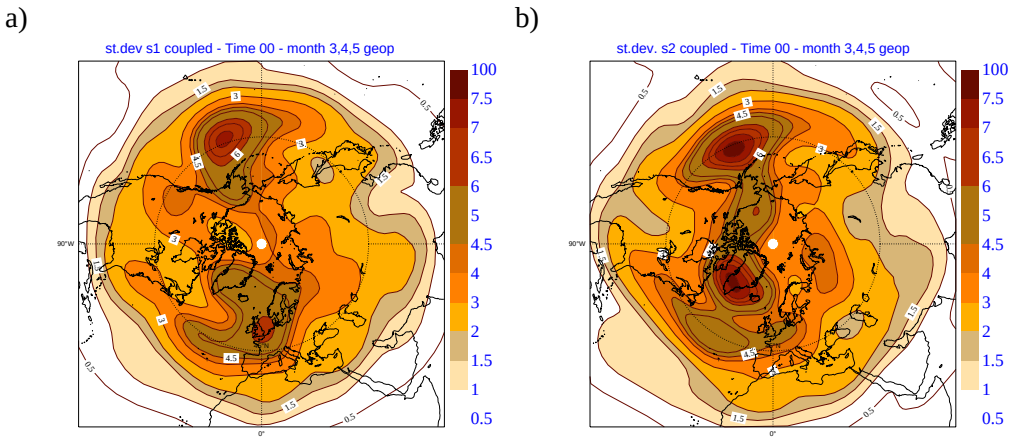


Figure 44: Standard deviations of 500 hPa geopotential height over winter periods.a) coupled system 1 b) coupled system 2.

approximate indication of the typical location of maximum variance. Although the inter-annual variability is estimated as the ensemble mean of variances computed for each realization of the sequence of 6 years as simulated by each individual ensemble member, the value, given the short period, might be subject to considerable sampling error. Unfortunately we do not have results available for a longer period, which would allow a more reliable picture to be given.

Comparison between the standard deviations of uncoupled simulations indicates that the atmospheric cycle 23r4 has a spatial distribution of variability over the Atlantic Ocean, Greenland and Iceland closer to the analysis and generally has reduced maxima of inter-annual variance with respect to cycle 15r8. The connection between the realism of the variance distribution over the European sector and the representation of the NAO mode is discussed in the next sub-section. Both cycles have a maximum of variance over the North Pacific, which is well known observationally but is not visible in the analysis for the particular years 1991-96. The atypical nature of Fig 43c reminds us that this six year sample is not adequate to represent climatology, and makes interpretation of the model/observation comparison difficult. That is, to what extent was the unusually low variance in the analysis a feature which the models should have reproduced, and to what extent was it a chance event which should not exist in the ensemble mean of the model variances?

Fig 44 shows that, in general, the effect of the coupling is to modulate the amplitudes of the variability with no major impact on the spatial distribution of the inter-annual variance. System 1 presents a reduction of variability over the Pacific sector compared with the uncoupled simulations (43). In the case of system 2 the coupling seems to enhance the variability, with increased maxima in both the Pacific and Atlantic sectors (compare 44b with 43b). How stable these amplitude changes would be if we looked at results from a longer time period is unclear. The atmosphere behaviour can be different between the coupled and uncoupled integrations because of changes in the mean state associated with the SST drift, and/or differences in the SST anomalies due to the imperfect nature of the coupled SST forecasts. Again, from these experiments it is not clear which of these factors may dominate. What *can* be said is that the coupling has not had any dramatic effect on the atmosphere variability, and that we should not be afraid to use the atmosphere model output from the coupled integrations.

Since the mean drift associated with the coupling does not have a substantial effect on the spatial structure of the winter variability, the extra-tropical winter variability modes simulated by the coupled system are similar to those from the atmospheric model and are quite realistic (e.g. http://www.ecmwf.int/products/forecasts/seasonal/verification/hemispheric.html).

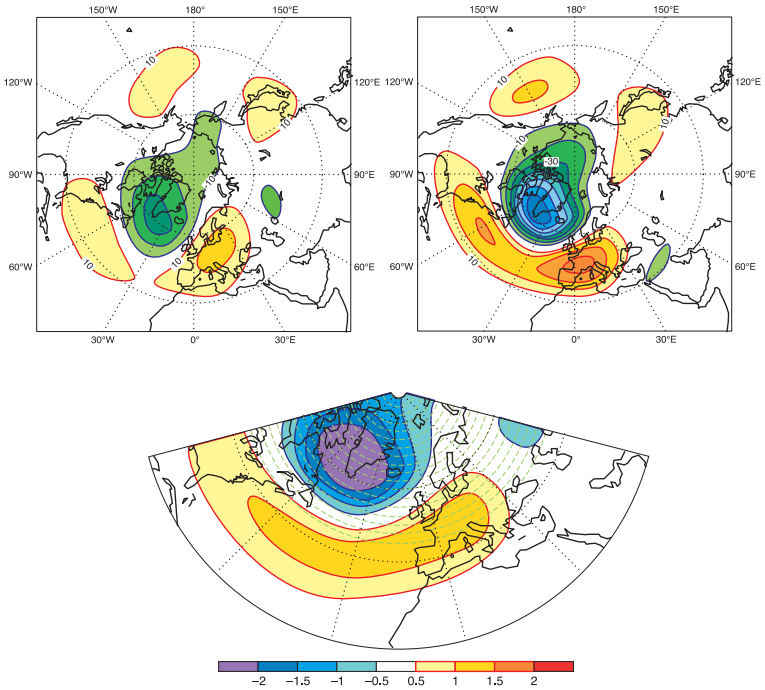


Figure 45: Pattern of the EOF of 500 hPa height in winter (DJFM) for the period DJFM 91/92 -DJFM 96/97 and from an ensemble of seasonal forecasts. Left) coupled system 1. This mode explains 15% of the total variance. Right) coupled system 2. This NAO-like mode explains 25% of the total variance. Contour interval 10 m. Lower) NAO reference mode described by an EOF pattern computed from the NCEP re-analysis data for the period 1948-2000.

The most prominent modes of low-frequency variability in the northern hemisphere extra-tropics are the Pacific North American (PNA) pattern and the North Atlantic (NAO) pattern. Seasonal predictions of those modes are discussed below.

4.5.3 The North Atlantic Oscillation (NAO)

The North Atlantic Oscillation is a coherent vacillation in surface temperature and sea-level pressure showing substantial variability on the year-to-year time scale and on longer periods. Since it plays a major role in controlling European climate and the Northern Atlantic ocean (Bjerknes 1964), it is of great interest to assess how well it is represented by the forecast. In both S1 and S2 the NAO mode is described by one of the leading EOF patterns (see Fig 45). It is interesting to see that the NAO pattern simulated from S2 is more realistic and explains a larger portion of variance than that from S1. This result is consistent with the improvement of A2 in representing the inter-annual variability over the Atlantic sector.

Fig 46 shows projections of the observed NAO mode (Fig 45c) onto seasonal-mean anomalies from S2. The skill of NAO seasonal predictions is rather limited but is comparable to that of the atmospheric simulations driven by observed SSTs and to the skill of S1 (not shown). It is important to note that as the NAO has a marked decadal and biennial variability, a period of 15 years might not be sufficient to properly assess the skill of the NAO forecasts.

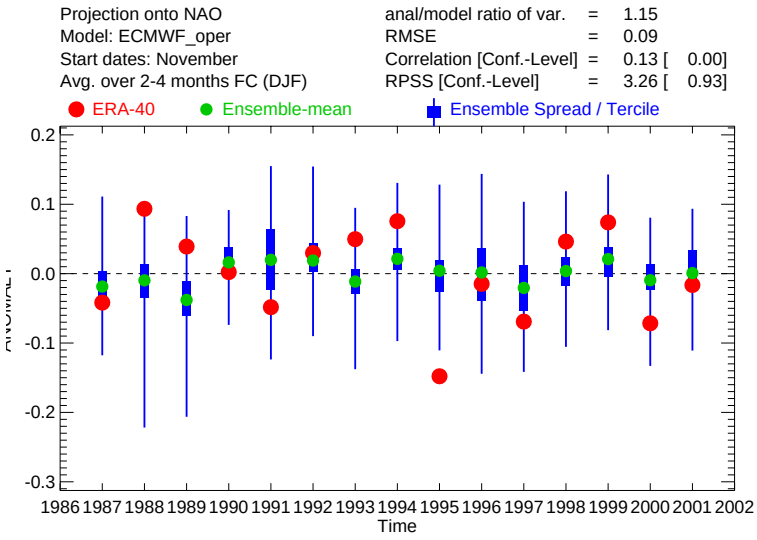


Figure 46: Seasonal forecast projections from system 2 onto the NAO reference mode. Seasonal forecasts are initiated in November and the ensemble consists of 40 members.

4.5.4 The Pacific North American (PNA) pattern

The relationship between PNA and SST anomalies in the tropical Pacific is well documented in many observational studies (e.g. Namias (1969), Horel and Wallace (1981), and Van Loon and Madden (1981)). ENSO warm events have an influence on the PNA pattern leading to above-normal geopotential heights in western North America and below-normal heights in the eastern half of the continent. Inversely, the ENSO cold event is commonly associated with lower geopotential heights over western North America and a higher geopotential height over the eastern half of the continent, called the negative PNA pattern. Although both of these long-wave patterns can develop without the influence of an ENSO event, they commonly occur in association with specific ENSO events. Results from an EOF analysis (not shown) indicate that both seasonal forecast systems simulate a realistic PNA pattern. Fig47 shows that PNA seasonal predictions from S2 have a reasonable degree of skill. Similar skill is found in the uncoupled simulations and in S1 predictions (not shown).

4.5.5 Intraseasonal variability over the northern Hemisphere winter

Intraseasonal variability is represented by the standard deviation of 500 hPa geopotential height covering the winter period 1987-2000. It is computed for two frequency ranges: low- frequency fluctuations with period longer than 10 days have been extracted using a low-pass filter, while high frequency variability is defined as the residual between the total signal and the low frequency part. The standard deviation has been calculated with respect to each seasonal mean, for each run separately, thus excluding inter-annual variations. The seasonal cycle is removed. Fig 48 shows the high-frequency variability in the Northern Hemisphere for DJF simulated by S2 (panel a) for comparison with the corresponding observed high-frequency variability (panel b). The high-frequency transience defines the locations of the major storm tracks. The position of the Atlantic storm track is represented well as is its intensity which is only slightly weaker than that analysed. The simulated Pacific storm track compares well with the analysis. Low frequency transience is illustrated in Fig 49. The simulated maximum over the Atlantic, associated with the position of Euro/Atlantic blocks, shows a southeastward shift

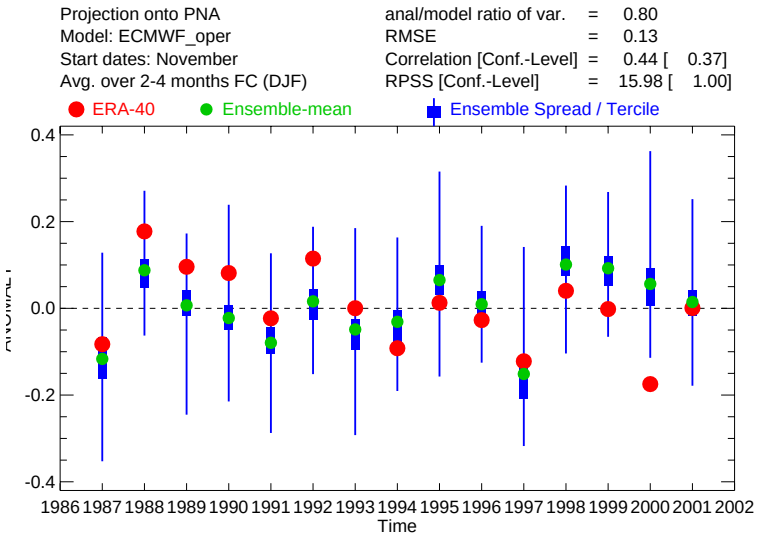


Figure 47: Projections as in previous figure but using as a reference mode the PNA pattern computed from the NCEP re-analysis data.

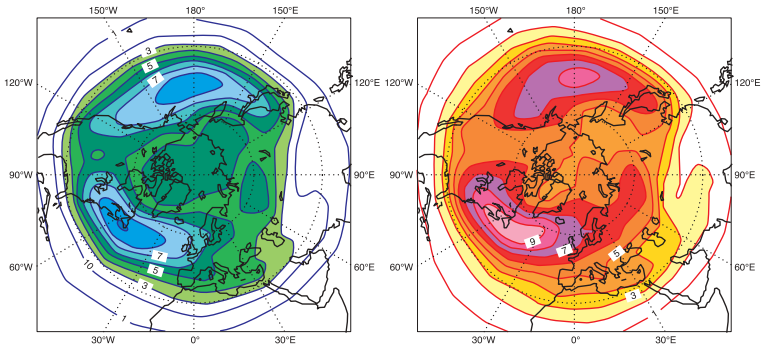


Figure 48: High frequency component (periods shorter than 10 days) of the standard deviation of 500 hPa geopotential height. Left) system2, right) analysis.

compared with the analysis. Over the northeast Pacific, the low frequency transience is well represented.

4.6 Summary of section 4

In this section we considered systematic errors in various quantities and showed that there were serious biases in the coupled system. Much of the bias was of atmospheric origin and was present in uncoupled experiments. Even the wind errors were initially largely of atmospheric origin although coupling did make some difference from month two onward. The different drift in the coupled system itself could be a consequence of the systematic error in the atmosphere. Thus the drift could result from errors in the fluxes passed to the ocean. If the ocean model were perfect but the fluxes were not, then the ocean would drift and this drift could modify the atmospheric drift. It could work to increase the error but it could also work to reduce it, and we sometimes see cases where the error in the coupled system is less than in the uncoupled. Of course we are not arguing that the ocean model is perfect. We will return to this point later.

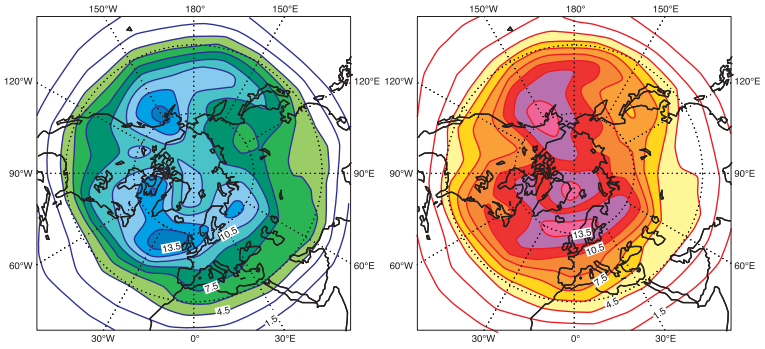


Figure 49: Same as previous figure but for the low frequency part (periods longer than 10 days).

Spatial plots of the variability in the near-surface wind for S1 and S2 showed that both systems under-represented the variability in the SPCZ but had too much variability in the northern ITCZ. S2 was much too inactive near the date-line. Analyses of upper air thermal errors showed that these were mainly of atmospheric origin and developed in the first month.

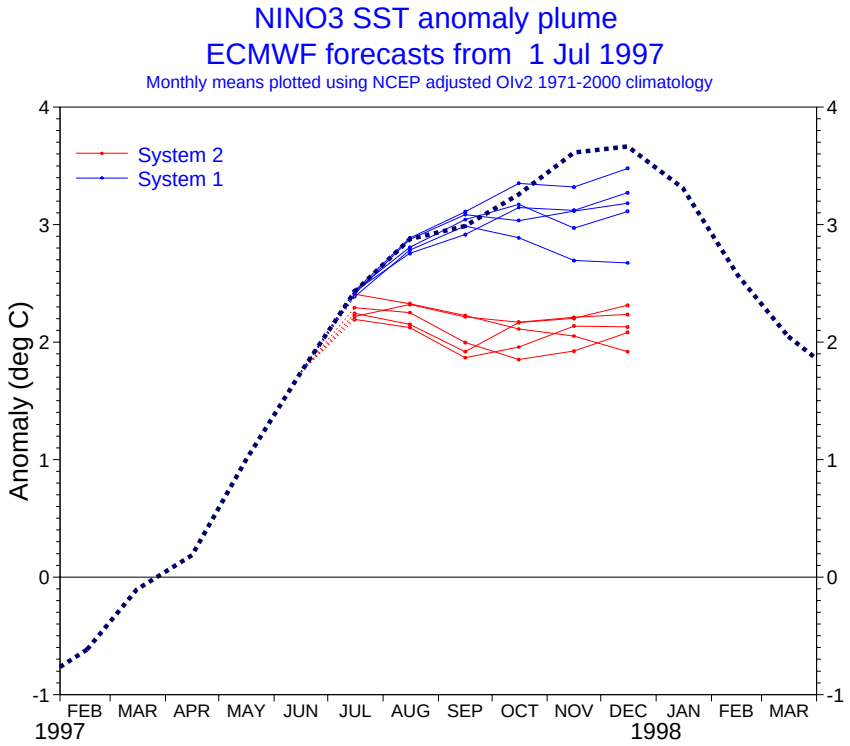
The bias from the atmospheric simulations A2 shows a substantial reduction of mean error over the European sector relative to A1, although over the Pacific sector the excessively zonal flow is present in both versions. The major features of the mean error in the Northern Hemisphere are similar between the coupled and uncoupled simulations indicating that the most of the error is of atmospheric origin. The standard deviations of A2 have a spatial distribution of variability over the Atlantic Ocean, Greenland and Iceland closer to the analysis and generally have reduced maxima of inter-annual variance with respect to A1. The effect of coupling is to modulate the amplitudes of the variability with no major impact on the spatial distribution of the inter-annual variance. The position of the Atlantic and Pacific storm tracks are represented well in S2. The simulated maximum of low frequency transcience over the Atlantic, associated with the position of Euro/Atlantic blocks, shows a southeastward shift compared with the analysis. Over the northeast Pacific, the low frequency transience is well represented.

5 Further diagnosis of SST variability in the 1997 El Nino.

We have considered some of the general differences between S1 and S2 in terms of the overall amplitude of SST variability, and how this is related to differences in the ocean and atmosphere models. We have also seen how the evolution of SST in the 1997 El Nino is affected by changing the individual ocean and atmosphere model components. What remains is to consider why the model SST evolves as it does in the various forecasts of the 1997 event. Is the consistent under- prediction of the event primarily due to the general tendency of Cy23r4 to under-predict interannual wind variability, or is the story in this particular case more complex?

Figure 50 shows again the behaviour of the S1 and S2 Nino3 forecasts, where we have chosen the July and October starts as representative of the general failure of S2 to generate or maintain the correct amplitude of SST anomaly. The very different SST development between the systems could be due to any combination of different forcing on the ocean from the atmosphere, differences in the ocean initial conditions, and differences in the ocean model. Since we are looking at the SST anomaly after it has been corrected for mean drift, the different forecast behaviour can also be considered in terms of the anomalies of these quantities and differences in the mean state.

a)



b)

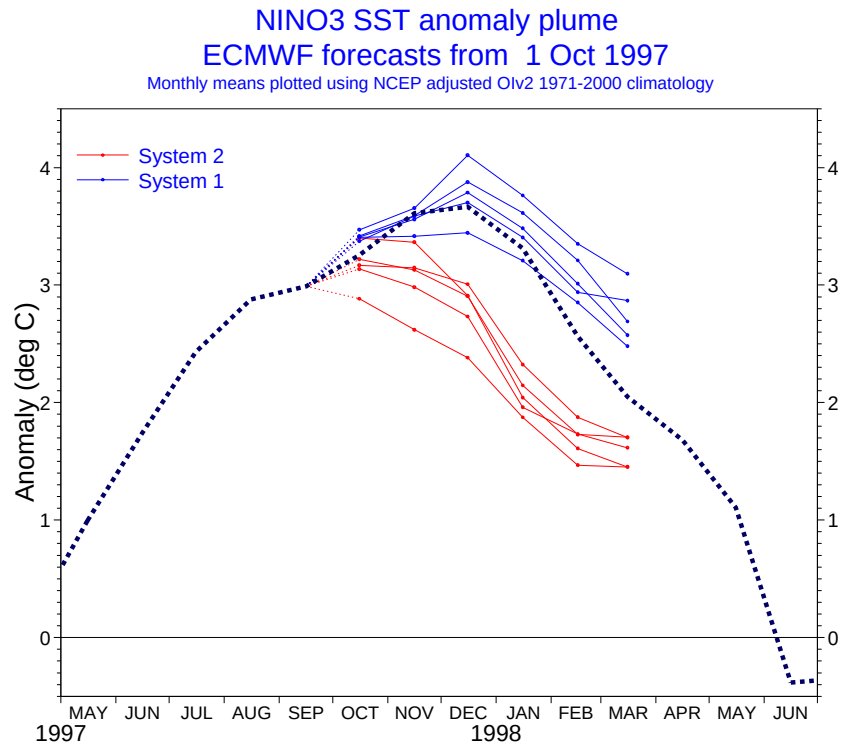


Figure 50: Forecast plumes for Nino 3 SST from S1 (blue) and S2 (red) for forecasts started in (a) 1 July 1997 and (b) 1 October 1997. The first 5 members are shown.

5.1 Net surface heat flux

Let us first consider the forcing of the ocean by the atmosphere model. A very direct forcing on the ocean is the surface heat flux. Figure 51 shows the net surface heat flux anomaly in the Nino 3 region from the S1 and S2 forecasts. It is immediately clear that the anomalous heat flux is acting to cool the warm SSTs, and that the cooling is stronger in S1 than in S2. This is the expected physical response to the higher SSTs in S1. The SST anomaly differences cannot be caused by the difference in heat flux anomalies due to formulation differences between the two atmosphere models. We note also that the heat flux anomalies from the forecasts match the observed estimate relatively well. The goodness of fit is a little surprising, since we do not have a high level of confidence in these 'observed' values, which are taken from a combination of ERA15 and operational short range forecasts. The mean heat flux is also generally a good match to observational estimates (not shown). Thus problems with the net surface heat flux, averaged over the Nino 3 region, do not seem to be a substantial cause either of the difference between S1 and S2, or of the errors within them. The situation in the far eastern Pacific, which may have some influence on the Nino 3 region, is considered later.

5.2 Zonal wind anomalies

5.2.1 Coupled wind anomalies

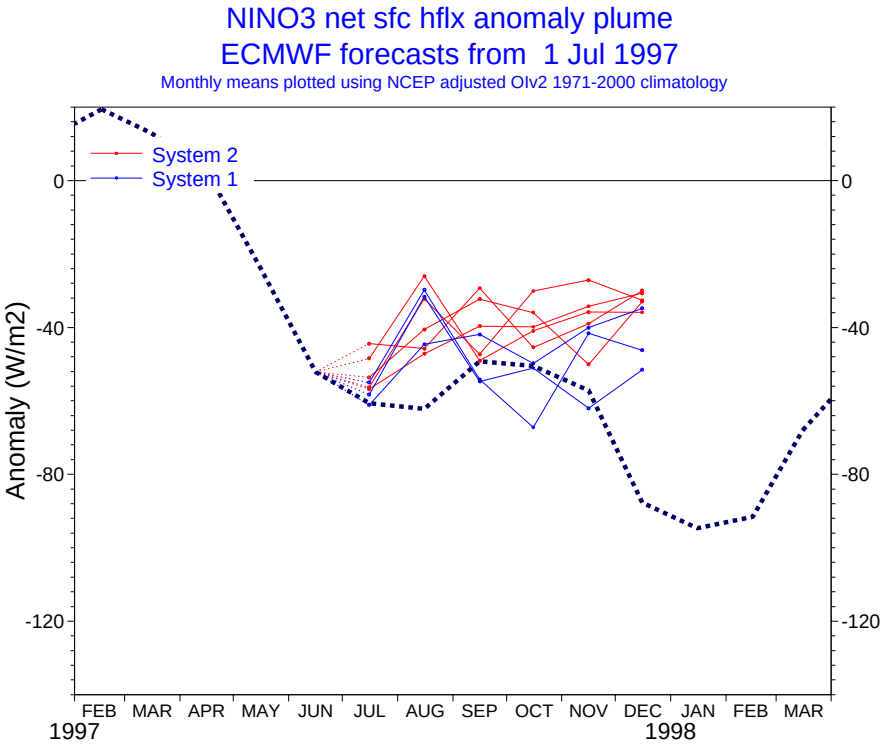
The zonal winds near the equator are the other major forcing of SST variability in the Nino 3 region. Wind anomalies might be either local (where there is generally little variability) or at some distance to the west (where the variability is generally larger, and where a time lag of 1 to 2 months for the Nino 3 SST response is expected). The wind field is noisier than the SST, and plume plots of wind anomalies show greater variation than plume plots of SST, both within each plot and in terms of differences between plots from successive forecast start dates. A single plot may not be representative, and so all start dates from June to October have been examined, and values for the Nino 3, Nino 3.4, Nino 4, EQ2 and EQ3 areas have all been plotted for each of these dates. Although the full set of plots is not presented here, the discussion is informed by an examination of all of the plots.

Figure 52 shows the zonal wind in the Nino 4 region for the July starts, and the Nino 3.4 region for October starts. These index regions are chosen to emphasize the differences in the winds between S1 and S2.

For the July starts, the wind anomalies are actually very similar for S1 and S2 in the central and east-central Pacific (Nino 3.4 and EQ2), although both systems underestimate the wind anomalies in these regions in the later stages of the forecast. Only in the western Pacific (Nino 4 and EQ3) are the wind anomalies stronger in S1. As can be seen in Fig 52a, the stronger values are present in the first month, and then persist through all subsequent months, although there is not a complete separation from the values in S2. For this particular start date, we can deduce (i) that the wind difference in the west Pacific may well be inherent in the atmosphere models, rather than a response to the diverging SST; (ii) that the very different evolution of Nino 3 SST in the first 1 to 2 months is not likely to have as its primary cause a difference in the wind anomalies; (iii) that over the course of the 6 months, it might be expected that the integrated effect of the substantially different wind field would have a big impact on the Nino 3 SSTs; and (iv) that S1 got the correct amplitude of Nino 3 SST, and a too strong amplitude of Nino 3.4 and Nino 4, despite having too weak zonal wind anomalies in an overall sense.

For the August and September starts, the overall situation is similar, but there are some differences. The difference in wind in Nino4 and EQ3 is not evident in the first month, but is more strongly evident than before in subsequent months. S1 does not now underestimate the wind anomalies in the west Pacific, although S2 still does. S1 also now has stronger anomalies in the central Pacific (Nino 3.4 and EQ2) than S2, so we would

a)



b)

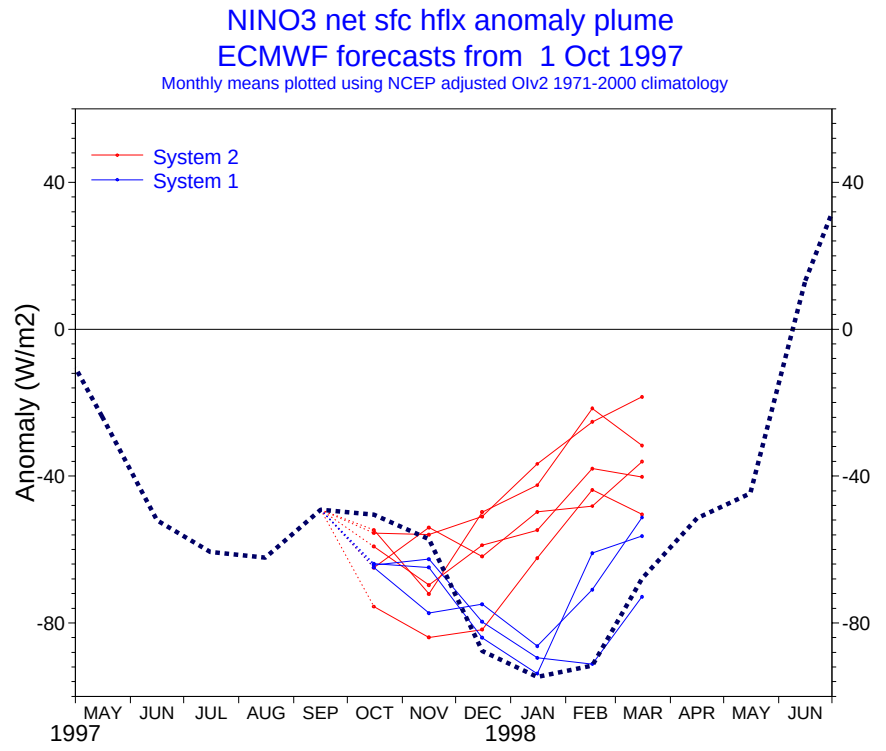
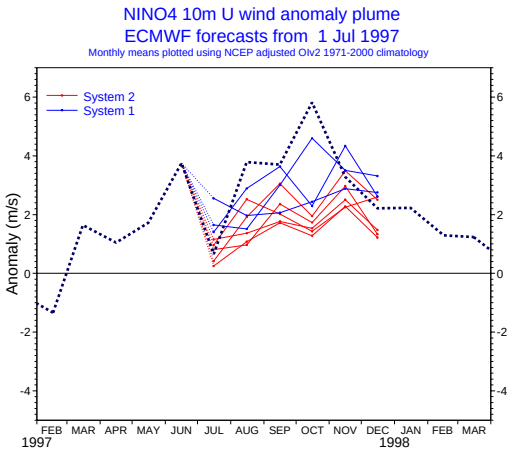
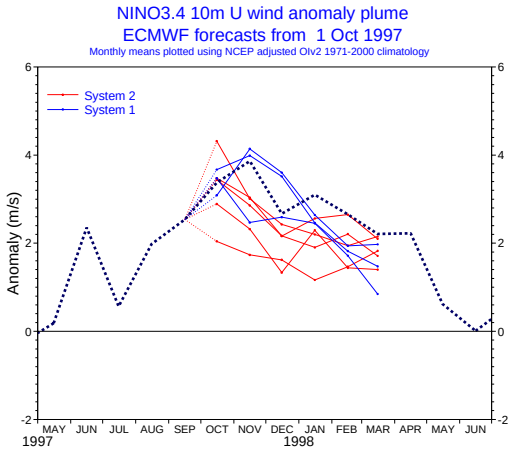


Figure 51: Net surface heat flux anomaly in Nino 3 from forecasts from S1 (blue) and S2 (red) for forecasts started in (a) July 1997 and (b) October 1997. For S1, data are available from only 3 of the 5 forecasts. Differences in heat flux cannot be the cause of the differences in SST. Dashed line shows EC analysis, which is of uncertain quality.

a)



b)



c)

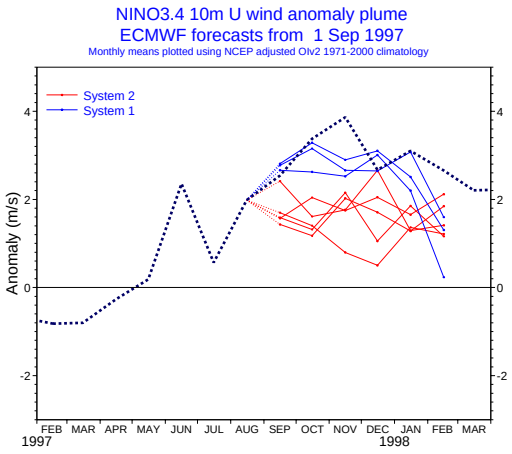


Figure 52: Zonal 10m winds from S1 (blue) and S2 (red) forecasts: (a) Nino 4 region, starting 1 July 1997; (b) Nino 3.4, starting 1 October 1997; (c) Nino3.4, starting 1 September 1997.

expect the wind differences to be a faster acting and stronger influence on Nino 3 SST. Figure 52(c) illustrates this for the winds in Nino 3.4 for September starts.

For the October starts, there is again a tendency for S1 to have stronger wind anomalies, illustrated for Nino 3.4 by Fig52(b). For central regions such as Nino 3.4 and EQ2, the S2 anomalies are too weak, and the S1 winds are about right, but in the west (Nino 4, EQ3), the S2 winds are actually very good while the S1 anomalies are too large. For all regions, it is only in the second month that the differences in wind anomalies between S1 and S2 become apparent. Again it seems that the difference in SST forecast in the first month, while not as large as in the July starts, is still bigger than can be accounted for by the wind differences.

In summary, differences in wind anomalies between S1 and S2 are definitely an important part of the story, especially in the longer range of the forecasts. It does seem, though, that the wind differences cannot be the only explanation of what is happening, particularly in the first month or so of the forecast.

5.2.2 Uncoupled wind anomalies

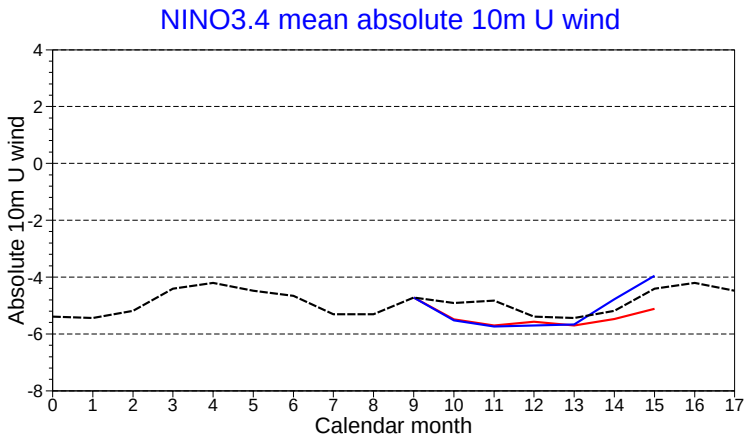
In the above diagnostics we cannot cleanly separate the atmosphere response from the coupled response – wind differences may in part be due to the response to SST differences. So, we will use results from uncoupled experiments. Fig 53 compares the zonal winds from S2 with those from an uncoupled experiment for October starts made with the S2 atmosphere and observed SSTs (A2U). The mean bias in the winds over all Octobers is almost identical in the first 4 months, with a divergence occurring only in the later stages of the forecast. When we look specifically at the anomalies in 1997 (Fig 53b), then we see that again in the first 2 months there is no clear distinction between the coupled and uncoupled experiments (red(S2) and green(A2U)), but that by month 3 the anomalies in the uncoupled experiment are stronger and closer to the observations. Given the substantial differences in SST by this time, the result is as might be expected.

For comparison we would like to plot the wind anomalies from the uncoupled S1 experiment A1U, but unfortunately a plot in this form cannot be produced with our present software. Also, no uncoupled experiments have been made for forecasts starting in July. Coupled/uncoupled experiments were discussed at length in Section 4, but anomalies for 1997 were not considered. Given the limited information available, here we simply state that in the first few months the coupling has only limited impact on the wind anomalies, and is not likely to explain the difference between the S1 and S2 wind anomalies. Ideally more experimentation should be carried out to confirm this.

For the October starts, at least, it does seem as if differences in the wind anomalies bewtween S1 and S2 in the first few months, are inherently due to the different atmosphere models (Cy15r8 and Cy23r4). It is the difference in the winds in the first 3 months which seems crucial in explaining the SST. Thus, the difference in the atmosphere models appears to be a major contributor to the SST difference.

A further aspect of the October starts can be noted at this point. The difference in wind anomalies between S1 and S2 is moderate in size, and may or may not be sufficient to explain the difference in Nino 3 SST anomalies. The difference between S2 winds and observations is smaller, however. This is not evident in the plots for Nino 3.4 but in Nino 4 the S2 wind anomalies are actually an almost perfect match of the observed values (while the S1 winds are substantially too strong - figure not shown). The integrated effect of the underestimation of the wind by S2 appears to be modest, and there is considerable uncertainty as to whether it can be the dominant cause of the underestimation of Nino 3 SST. The explanations of the underprediction of S2, and of the difference between S1 and S2, may not be the same.

a)



b)

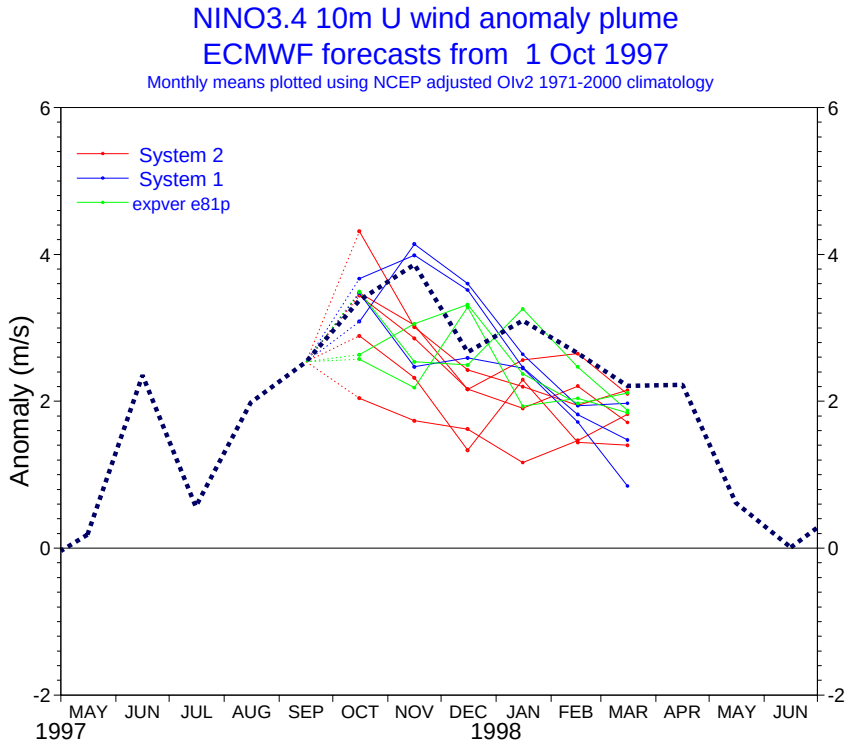


Figure 53: Zonal 10m wind in Nino3.4. (a) Mean wind from 10 years of forecasts for S2 coupled (red) and uncoupled (blue) (b) Forecast anomaly from 1 Oct 1997 for S2 coupled (red) and uncoupled (green), and S1 coupled (blue).

5.2.3 Wind anomalies in cross coupling experiments

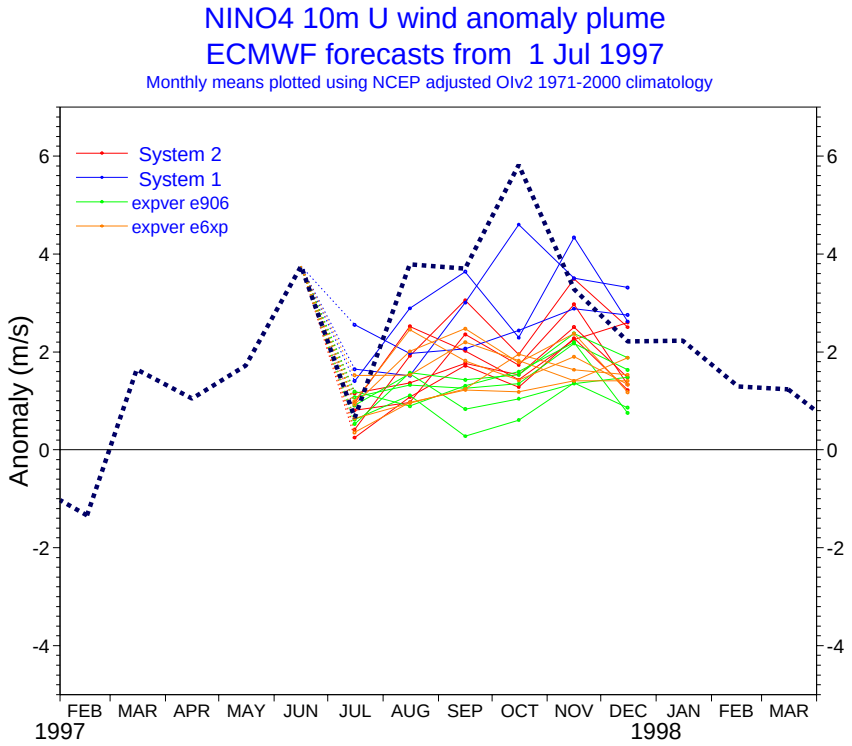
Another way of confirming the origin of the difference in wind anomalies is to look at the results from the cross-coupling experiments. Note that for technical reasons we are not yet able to provide atmospheric diagnostics for the S1 atmosphere coupled to the S2 ocean (A1_O2) – the data exists but needs substantial software development for us to be able to process it equivalently to the A2 data.

Figure 54 shows the wind in Nino 4 from the July forecasts, and in Nino 3.4 for the October forecasts. These regions are chosen to highlight the maximum difference between S1 and S2. It is immediately clear that all of the experiments with the S2 atmosphere have similarly weak wind anomalies, regardless of which ocean model and ocean initial conditions are used. Fig 54 can be compared with Fig 55, which shows the SST anomalies in Nino 3 for the same set of experiments. We see that although the choice of atmosphere model dominates the difference in SST, there is a small contribution from the choice of ocean model/ocean initial conditions. For example, in the second month of the July start A2_O1 (green) becomes systematically warmer than A2_O2 (orange), and this additional warming persists for the rest of the forecast. Curiously, this warming in the east Pacific is associated with weaker positive anomalies in the winds in the west central Pacific. In fact, looking at the whole of the equatorial SST gradient, the wind difference appears to be counter gradient (*i.e.*, a weaker SST gradient is associated with a stronger zonal wind). A possible explanation for the wind differences is that off-equatorial or remote SST differences are responsible, although random sampling effects may also be involved - the statistical significance appears to be only moderate. The reason for the SST anomalies remaining warm despite an unfavourable wind could be a difference in ocean initial state anomaly. Also, the warmer SST anomaly may be helped by differences in the mean drift between A2_O1 and A2_O2. The O2 ocean is systematically warmer by around 0.2 deg C in its climate, which may help keep the anomaly in 1997 that little bit smaller.

If we look at the difference between SST in A2_O1 and S1 (green and blue), we see that this is negligible in the first month, moderate in the second month, and then becomes large in the third month and beyond. This is what we would expect given the space-time structure of the difference in the winds between the two atmospheric components. S2 has a high resolution ocean. Comparing S2 and A2_O2, the difference in resolution may be responsible for a small difference in the ensemble mean SST in Nino 3 in the first month, but afterwards has little effect in this region. These results are all consistent with the difference in Nino 3 SST between S1 and S2 for the July forecasts being an approximately linear combination of: (i) the difference in ocean giving a modest difference of 0.2-0.3 deg C, approximately constant throughout the 6 months, and (ii) the difference in atmospheric winds causing a difference which appears in the second month, and grows to a substantial size by the third and fourth months. (The results are also consistent with other factors and non-linear interactions being important contributors to the size of the difference, even if there is no positive evidence for this at this stage).

In passing, note the July SST forecasts for the Nino 4 region (shown in Fig 56a). Here the change in atmospheric model causes a reduction in the large overestimate of SST anomalies given by S1, even if this is for fundamentally incorrect reasons (the wind anomalies in Cy23r4 are far too weak). However, the ocean resolution change is particularly beneficial, with the high resolution model (red) giving much more realistic anomalies than the low resolution models. The same effect is seen even more clearly in the EQ3 region, in the far west Pacific (not shown). The better performance in this case seems associated with the much reduced cold bias seen in the high resolution model (Fig 56b). The link between the cold bias and forecast performance in the west Pacific is discussed later. Note that the overall statistics for 15 years of Nino 4 SST forecasts are made slightly worse by the high resolution in the context of S2. This may be because the over-sensitivity of the low resolution ocean may help to compensate for the under- activity of the Cy23r4 atmosphere, and we might hope that with a more realistic atmosphere model, the benefit of the high resolution ocean would be more generally apparent. We do not present a study of the impact of ocean resolution here, but simply note that it can have a

a)



b)

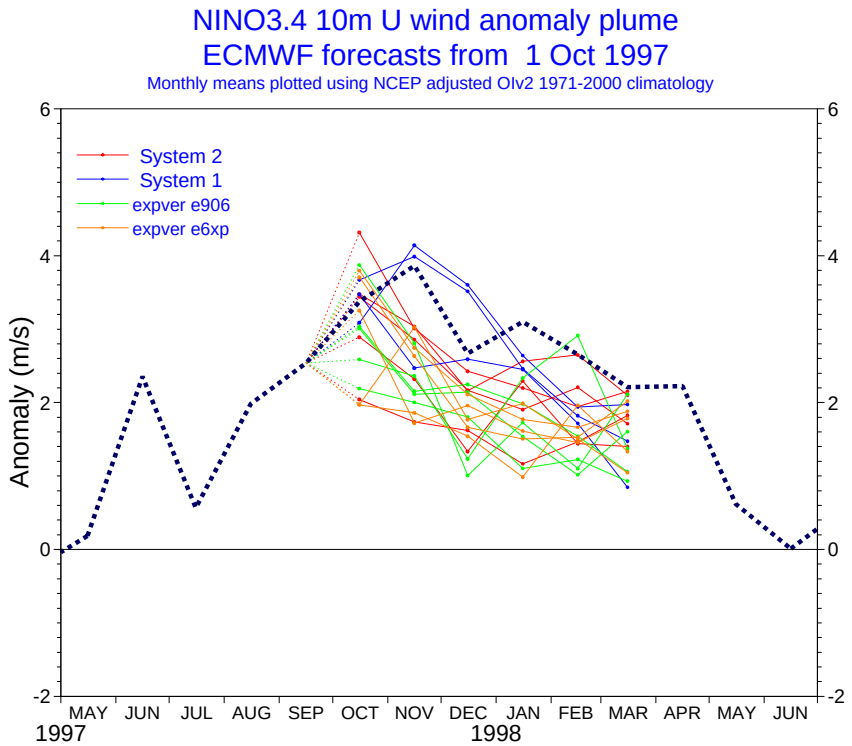
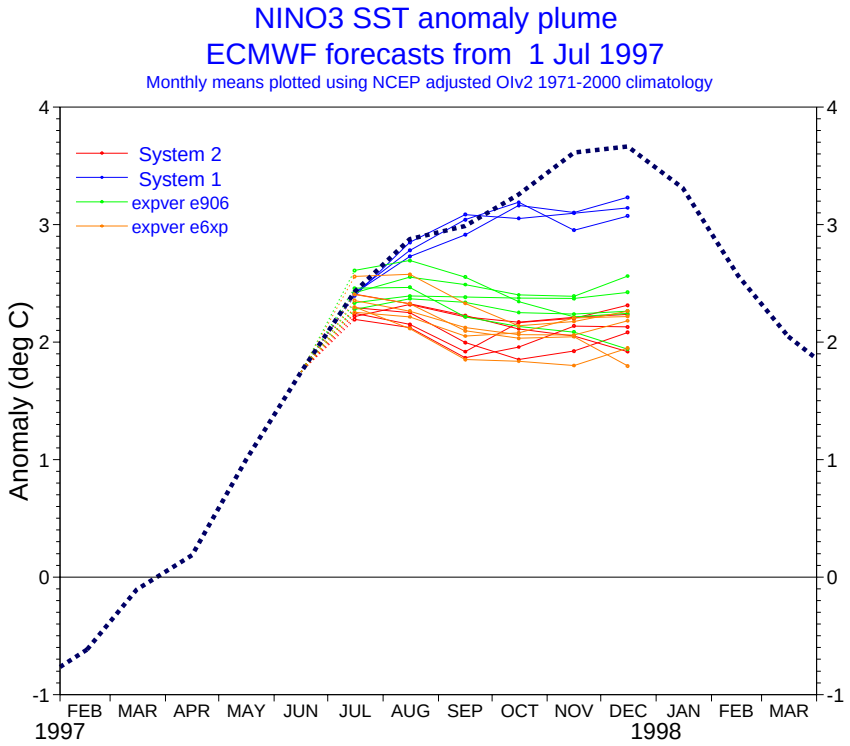


Figure 54: Zonal 10m wind anomalies from S1, S2 and hybrid experiments: S1 (blue), S2 (red), A2_O1 (green) and A2_O2 (orange). (a) Nino 4 region, starting 1 July 1997; (b) Nino 3.4, starting 1 October 1997.

a)



b)

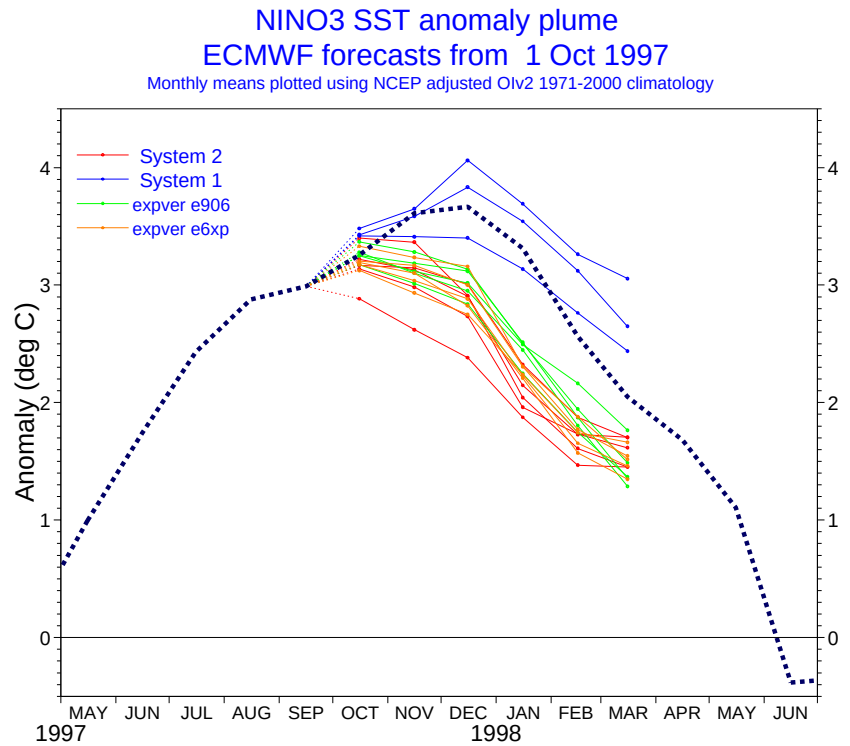
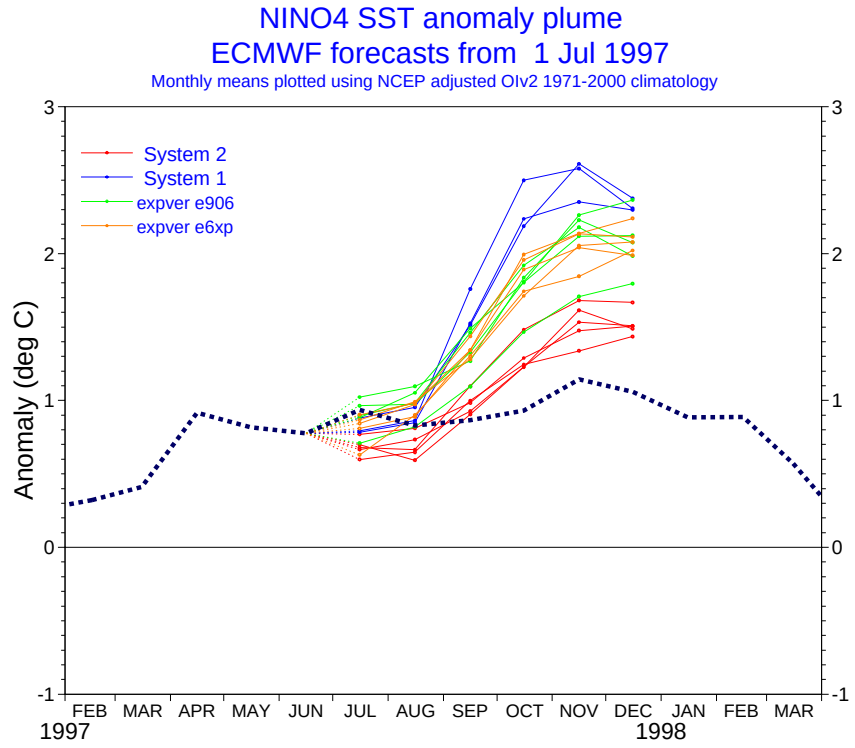


Figure 55: Nino 3 SST anomalies from S1, S2 and hybrid experiments: S1 (blue), S2 (red), A2_O1 (green) and A2_O2 (orange). (a) starting 1 July 1997; (b) starting 1 October 1997.

a)



b)

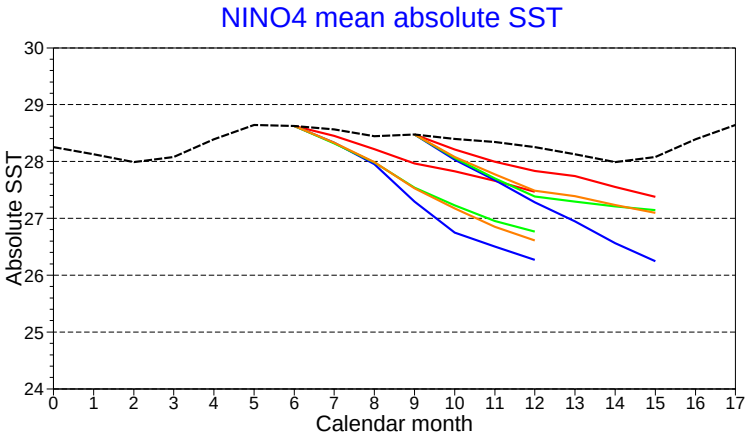


Figure 56: SST in Nino 4 for S1 (blue), S2 (red), A2_O1 (green) and A2_O2 (orange). (a) Forecast anomaly from 1 July 1997. The high resolution ocean in S2 has a real impact. (b) Mean SST from 10 years of forecasts. Again S2 is very different to A2_O2.

significant impact in certain cases.

For the October forecasts, the S1 Nino 3 SST anomalies are well separated from the model combinations which use the A2 atmosphere, and the choice of ocean model/ocean initial conditions seems to make very little difference. The winds in the coupled forecasts in Nino 3.4 are not separated in the first month because of the large spread, but in the second and third months a fairly consistent difference can be seen in the wind between experiments using the A1 atmosphere and those using the A2 atmosphere. A similar picture occurs in Nino 4, and there is no difference in the winds in Nino 3 itself. In summary, the difference between S1 and S2 is all down to the atmosphere model, in terms of both SST and wind anomalies.

In the previous section we noted that the errors in S2 wind anomalies may not be sufficient to entirely account for the errors in S2 SST anomalies for the October starts. Despite this, the cross coupling experiments demonstrate that it is the atmosphere model that makes the difference between S1 and S2, even for the October starts. To make progress, we need to consider other ways in which the atmosphere model can influence the SST anomalies. Important possibilities are the effects of the mean winds and heat fluxes on the ocean mean state.

5.3 Effect of the mean state on SST anomalies

5.3.1 Mean zonal winds

It may well be that errors in the mean wind are also playing a role in degrading the forecast. Fig 57a shows the time evolution of the mean zonal wind in the Nino 4 region. All of the coupled models have zonal winds which are generally too strong. For S1, the winds are about right for the first month (or sometimes even too weak), and develop a strong bias later in the forecast. This is probably associated with the strong cold bias in SST which develops in this region in S1. The atmospheric model itself does not have this bias (cf Fig 29). For S2, however, the too-strong winds are present immediately in the first month. The SST drift is not strong in S2, especially at the start of the integrations, and the bias is therefore likely to be inherent in the atmospheric model. (This is confirmed by the uncoupled integrations, see Fig 30). Other plots (not shown) confirm that the S2 winds in the west Pacific have a seasonally varying bias which has little sensitivity to the lead time. Note also that the bias in the winds from A2_O2 and A2_O1 is very similar to S2 in the first few months, despite the appreciable difference in SST drift shown in Fig 56b.

For both S1 and S2, the bias in the winds in Nino 4 will result in changes in the ocean mean state, which are likely to affect the sensitivity of ocean model SST to wind anomalies. The resulting impact on SST anomalies in Nino 4 itself is discussed in a later section. Wind errors in Nino 4 should affect Nino 3 remotely, and result in a slight raising of the thermocline mean state in all of the models.

Figure 57b shows that in the Nino 3 region, there is almost no bias in the zonal winds from S2, whereas S1 has a modest bias to stronger easterlies. The slightly weaker mean winds of S2 compared to S1 could result in weaker upwelling and reduced sensitivity of SST to local and remote forcing. The zonal wind bias is strongest in months 4-6 of the October starts, which fits with a period where wind anomalies alone may be insufficient to explain the Nino 3 SST difference between S1 and S2. Note that although the difference in mean wind might explain the SST difference between S1 and S2, it cannot explain the error in S2 SST, since S2 has the mean winds in Nino 3 about right. The source of the error in S2 must thus be something else. For example, it could be that the ocean model has errors such that the right sub-surface and mixed layer structure in the eastern Pacific is not obtained, even with the correct forcing.

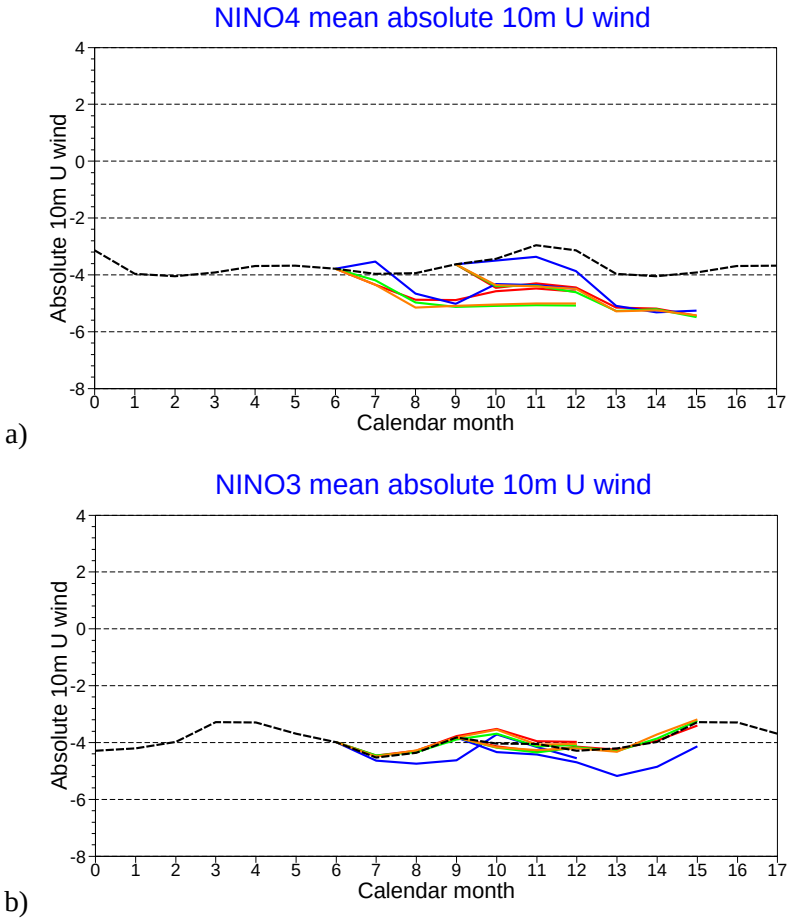


Figure 57: Mean 10m zonal wind for S1 (blue), S2 (red), A2_O1 (green) and A2_O2 (orange): (a) Nino 4 and (b) Nino 3.

5.3.2 Mean heat fluxes and the far eastern equatorial Pacific

An interesting feature is seen in the behaviour of mean heat flux and mean SST in the far east Pacific, for example Nino 1+2 (Fig 58). Here, the mean SST in both models rapidly becomes too warm, with S2 being worse than S1. Together with this rapid warming, the net heat flux into the ocean is substantially less than observational estimates. It seems that the heat flux is about right when the SSTs are correct at the start of the coupled integrations, and that the heat flux then reacts to oppose the large SST warming which has some other cause. This diagnosis assumes that the estimates of observed heat flux, derived from ERA15 and operations, are not seriously in error. In fact, the ECMWF estimates are likely to be biased high, because of an underestimation of stratus in this region. For the July starts, it is evident that the heat flux would need to be reduced by more than 100 W/m2 to obtain the correct SSTs, and we do not believe that the estimated heat flux is wrong by this amount at this time of year. For the later stages of the October starts it is less clear, because the SST bias is no longer large. Note also that the difference between S1 and S2 cannot be explained by the heat fluxes - the warmer system has less heat going into the ocean.

Thus although the heat flux does not seem to be the major source of error in the mean state, it is apparent that something else is, and that the mean state of SST in the far eastern Pacific is unrealistic in the models. Examination of plots of the local winds reveals nothing obvious that might produce a large warming. The zonal winds (not shown) in Nino 1+2 are substantially too weak in S1, but are the right strength in S2. The meridional winds, which are also important for ocean dynamics in the eastern equatorial ocean, are too strong in both models but this time with S1 the stronger of the two (Fig 58c). For both models, these winds would be expected to give too cold SSTs, both because of increased upwelling and increased mixing. The net result of the wind differences on the SST seems to be negligible, since the cross coupling experiments show that the modest difference in SST in Nino 1+2 is due to the ocean model, not the atmosphere. For both systems, the warm bias in Nino 1+2 appears not to be due to local atmosphere forcing errors, and must therefore be related to the state of the ocean. Errors in the ocean state may be due to errors in the ocean model itself (either local or non-local), and remote forcing errors for example in the equatorial waveguide. The detailed structure of the winds along the coastline might also be an issue.

If we look on the equator, we again see signs of the warm bias problem in the east Pacific. Comparing regions such as EQ1 and Nino 3, the impression is of a combination of a warming problem in the far east, and a cooling problem further away from the coast (spatial plots confirm this, cf Fig 22). The net result for an extended region such as EQ1 (90W-130W) is variable: S2 has strong warming for July starts, for S1 the net result is closer to neutral. Unfortunately we do not have an index region defined more locally for just the extreme east Pacific on the equator, so we must work with EQ1 for now. For this region, the meridional winds are too strong in both cases, by about the same amount. The zonal winds are somewhat stronger in S1, as discussed earlier for Nino 3. The cross-coupling experiments show that unlike Nino 1+2, the SST differences in EQ1 are dominated by the change in atmosphere model, with the ocean model being almost irrelevant and the ocean resolution making only a small difference.

It is worth noting that this sort of incorrect structure of mean SST in the far eastern Pacific, with too warm water near the coast, is a generic problem with coupled models. It is also a general problem that the SST variability in the far east Pacific is too weak (the maximum amplitude of variability is always shifted away from the coast and too far to the west). Since these two problems are generally seen together, it is likely they are linked. Certainly, if the problem in the far east is a lack of ocean upwelling in the mean state, then one would expect variations in sub-surface temperature and upwelling to have a reduced impact on SST. Although the problems are typically worst close to the coast, the Nino 3 area is affected at its eastern edge. The fact that S2 seems to have slightly bigger and more westward extending errors than S1 may contribute to the differences in Nino 3 SST anomalies. The warmer SSTs in the east Pacific may be responsible for the reduced zonal wind seen in S2 shown in Fig

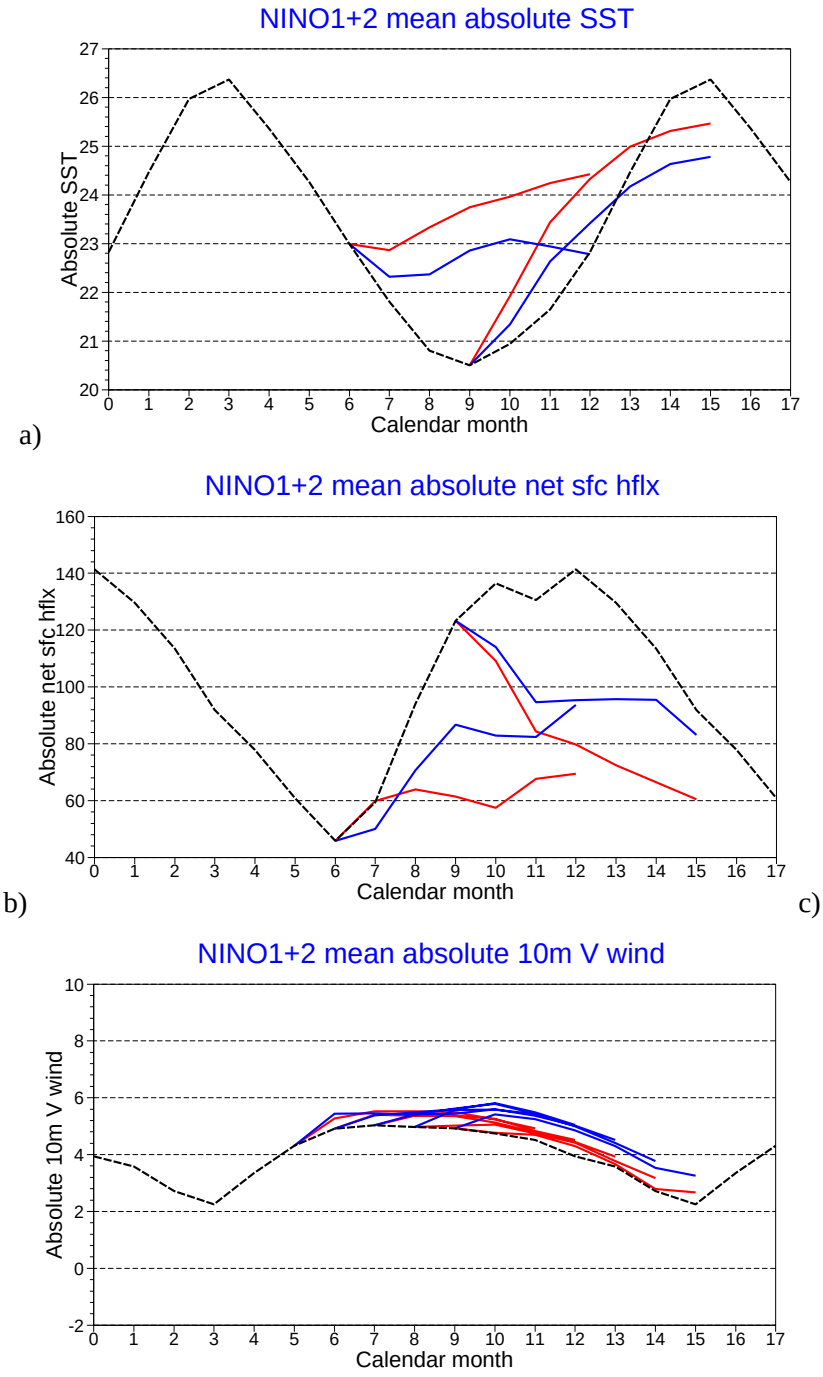


Figure 58: Mean coupled model fields for Nino 1+2 region, showing S1 (blue), S2 (red): (a) SST, (b) Net surface heat flux and (c) meridional 10m wind.

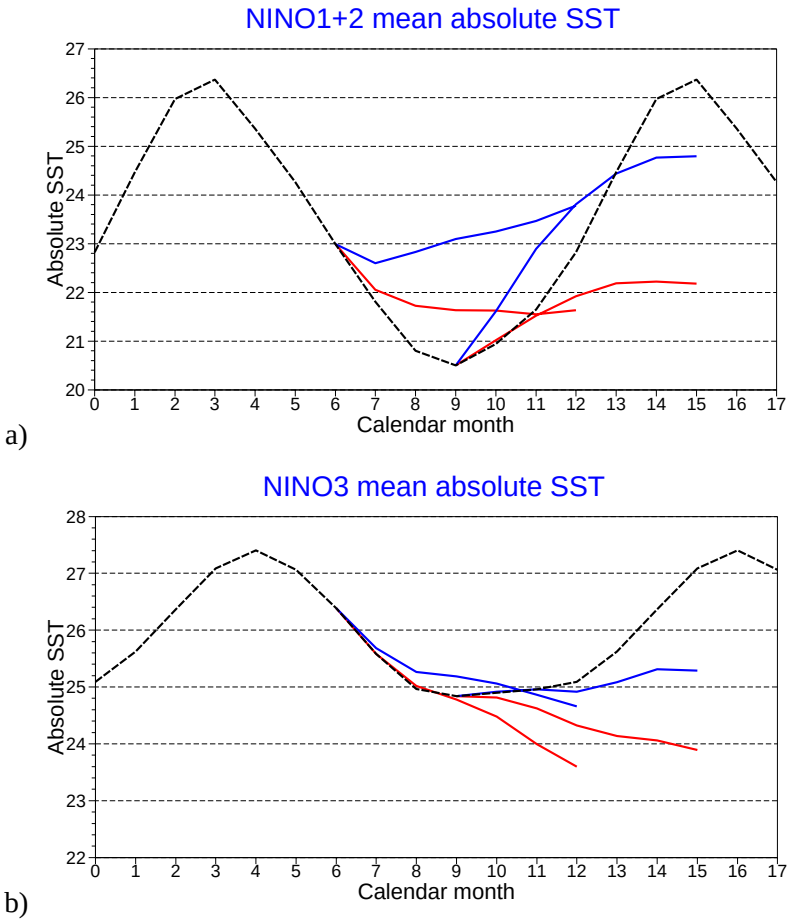
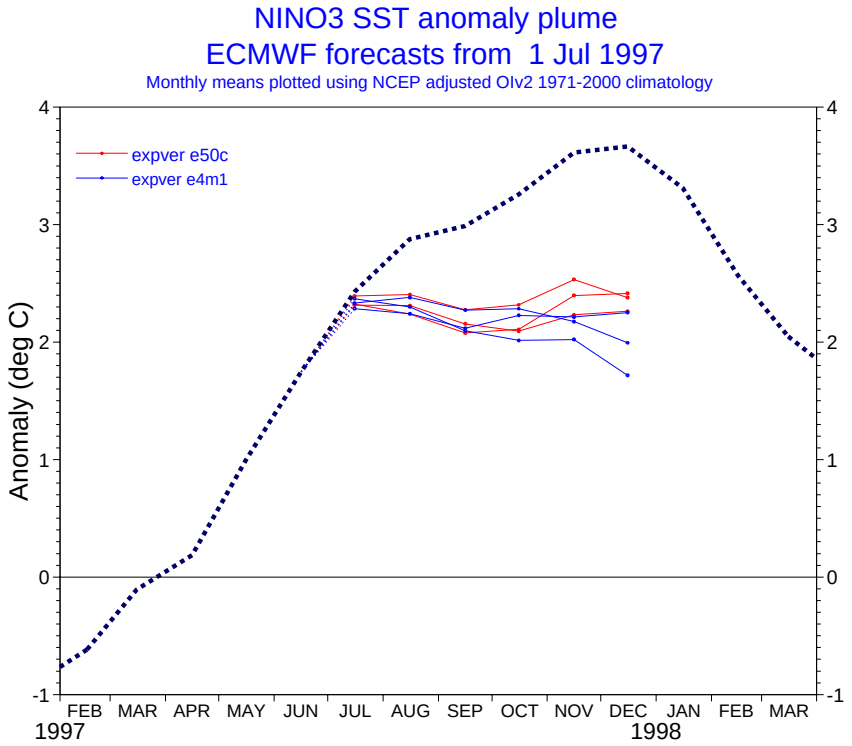


Figure 59: Mean SST from coupled model experiments, showing an early version of S2 with (red) and without (blue) heat flux correction in the eastern Pacific. (a) Nino 1+2 and (b) Nino 3.

57b. Note that although the winds look more realistic in S2, they are being forced by a quite unrealistic SST gradient.

Experiments have been carried out which apply a substantial heat flux correction of 60 W/m2 cooling to the far eastern Pacific. The experiments were made with a prototype version of S2, and the heat flux perturbation was applied to an extended version of the Nino 1+2 region only. Even though we do not believe that heat flux errors are the original cause of the excess SST, the heat flux cooling has the expected effect of substantially reducing SST (Fig 59), both in the region of the perturbation and along the equator towards the west. Typically the net result on forecast anomalies is barely perceptible, meaning that the system is showing perhaps surprisingly strong linearity against this type of perturbation. This can be true even during strong SST anomalies such as those of the 1997 warming; the July 1997 forecast is an example (Fig 60a). One notable exception is the October forecast (Fig 60b). Here, the colder mean state allows warmer anomalies, making S2 look more like S1. A difference in mean state of Nino 3 SST typically 1 deg C in the last 3 months of the forecast corresponds to a difference in anomaly of just over 0.5 deg C. Although this shows that there can be an effect on the anomalies from what is essentially an SST perturbation, it might be noticed that the effect is noticeable only at the time when in fact there is not a problem with the far east Pacific being too warm. In fact, we only obtain larger anomalies by making the east Pacific much colder than observed.

a)



b)

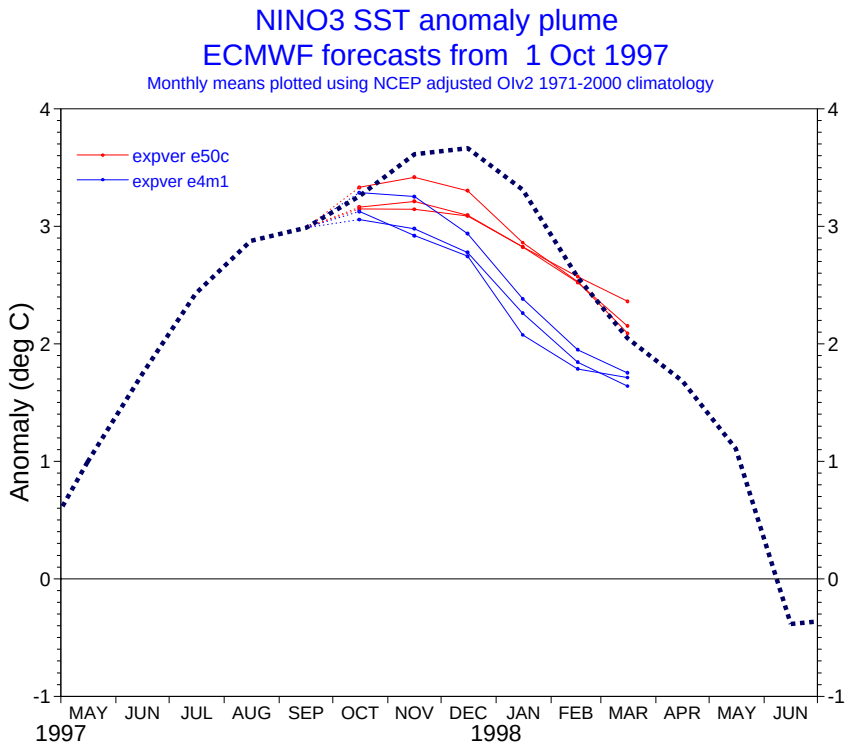


Figure 60: Nino 3 SST anomalies from experiments with (red) and without (blue) heat flux correction in the eastern Pacific. (a) July 1997, (b) October 1997 starts.

In summary, there are significant errors in the mean state of the far eastern Pacific. The primary cause seems not to be surface heat flux, and must be due to errors in the ocean state caused by ocean model errors and/or remote wind forcing errors.

Changing the mean SST directly has almost no impact on the forecast anomalies in most cases, but October 1997 is an exception: the later stages of the anomaly can be made larger if the mean SST is made colder. This may explain part of the difference between S1 and S2, but it seems that correct mean SSTs will not result in correct anomaly forecasts. It might still be true that a correct ocean mean state (meaning a realistic mixed layer and upper ocean structure) would result in better SST anomaly forecasts - simply perturbing the SST has not explored this question. On physical grounds, we would expect that significant mean state errors would impact the behaviour of SST anomalies, and our diagnosis of the model mean state makes it clear that significant errors do exist.

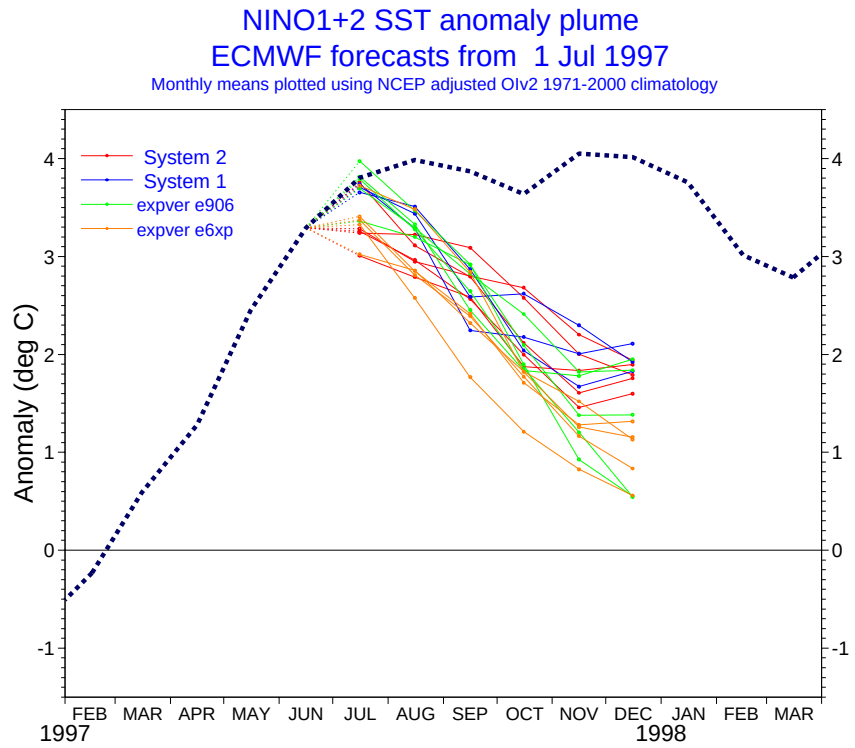
As an aside, the behaviour of the Nino 1+2 SST anomaly is interesting (Figure 61). In July, all model combinations show an identical and erroneous decline of the anomaly. In October, S1 is an exception: it remains fairly warm, while the other experiments cool. It is the atmosphere model which makes the difference. When we look at the local forcing in Nino 1+2 (heat flux and winds), there is no difference between S1 and S2 that can explain the different SST behaviour. Obviously there are differences between S1 and S2 in terms of the remote wind forcing, which drive warmer anomalies in S1 along much of the equator. The additional wind forcing of S1 is similar in both the July and October starts, however, and not too much weaker than observed. One hypothesis to explain this is that errors in the sensitivity of Nino 1+2 SST to the additional remote wind are very different for the two seasons. Looking at the mean state this is certainly plausible, since in the July starts the mean SST for S1 is much warmer than it should be, and this is probably associated with a too-warm sub-surface temperature structure; by contrast, in the October starts the S1 mean state is apparently not much warmer than it should be, and the remote effect of wind anomalies has a reasonable chance of producing the right magnitude of impact on the SST.

The fact that the mean state in the July starts is badly erroneous, and that all experiments fail to produce the right SST anomalies in the July 1997 starts, strengthens the suggestion that mean state errors may be causing trouble with the anomalies. In the flux perturbation experiment described above, the substantial improvement in the mean state SST in the July starts *does* improve the Nino 1+2 anomaly forecast for 1997, with the anomalies being warmer by about 0.25 deg C (not shown).

Careful analysis of the drift and anomaly performance in Nino 1+2 of the cross-coupling experiments further supports a relationship between the mean state and the anomaly. A plot of the drift (not shown) demonstrates that the change in mean bias in Nino 1+2 between S1 and S2 is almost entirely due to the change from the old to the new (low res) ocean model, and that this change in mean bias is associated with a change in the July 1997 anomaly of the expected sign (orange is below green in Fig 61a). The reason that the S2 anomaly is not particularly weaker than that of S1, despite the unfavourable changes to the mean state and the reduced remote wind forcing, is that the increase in ocean resolution produces a noticeable increase in the anomaly (red is higher than orange, and returns to a similar level to blue). This occurs without any change in the mean SST bias, although of course the sub-surface temperature structure may have changed. Note that this applies only to the July forecast: in October the resolution change does not affect the anomaly.

This examination of the Nino 1+2 SST anomalies demonstrates significant interactions between many different aspects of model and forecast failures, sometimes demonstrating striking non-linearities, such as that illustrated by Fig 61. It is clear that much work will be needed to reduce all of the errors in our forecasting system to a satisfactorily small level.

a)



b)

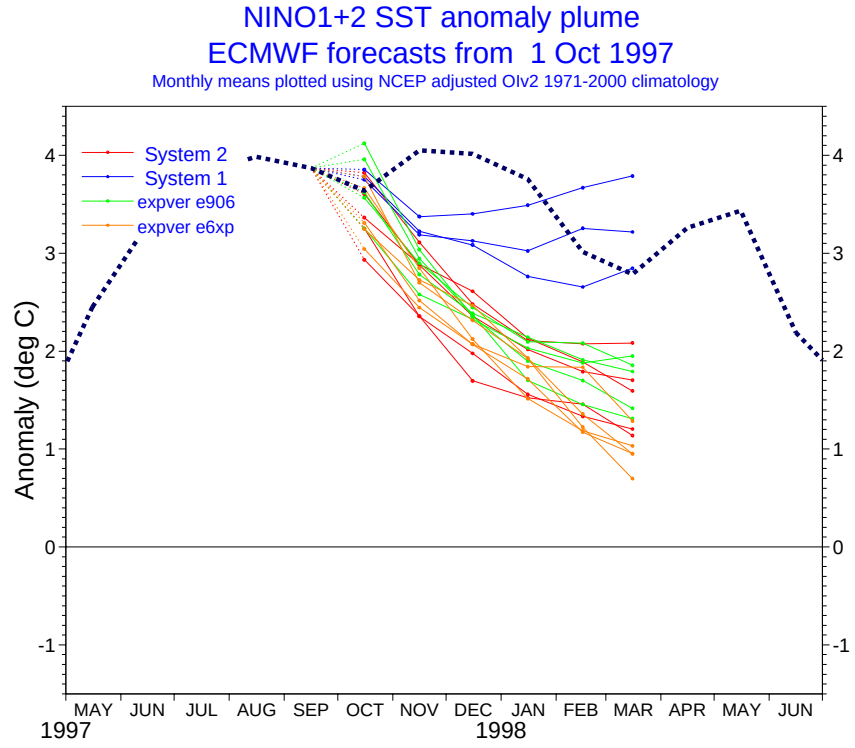


Figure 61: Nino 1+2 SST anomalies from S1 (blue), S2 (red), A2_O1 (green) and A2_O2 (orange): (a) July 1997, (b) October 1997 starts.

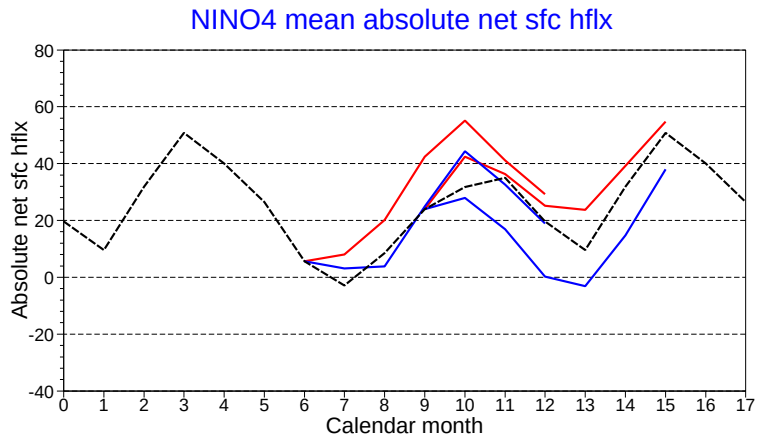


Figure 62: Nino 4 mean heat flux from S1 (blue) and S2 (red).

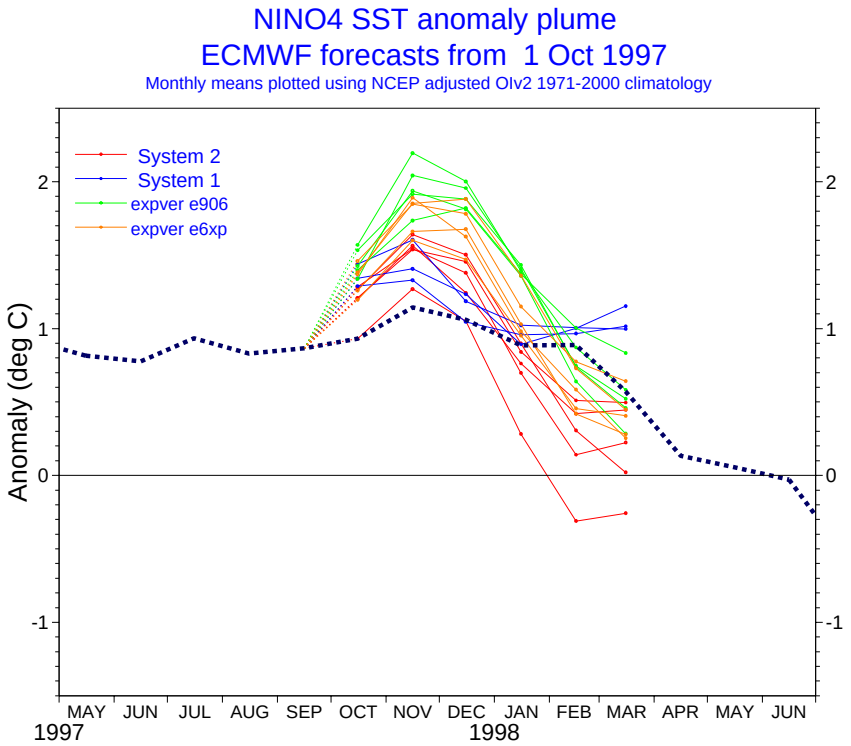


Figure 63: Nino 4 SST anomalies for October 1997 starts from S1 (blue), S2 (red), A2_O1 (green) and A2_O2 (orange).

5.4 The west central Pacific

We have concentrated on the factors controlling the Nino 3 SST, because this is the index which received most attention with S1 and about which most concern has been expressed with S2. We have already noted the impact of ocean resolution on the SST anomalies in the Nino 4 region from the July forecasts, as seen in the cross coupling experiments (refer back to Fig 56). What do more general diagnostics show for this region?

For the mean state, S2 has a higher net surface heat flux than S1 of 10-20 W/m2, despite its higher temperatures (Fig 62). S2 also has stronger zonal winds in the early stages of the forecast, which will tend to cool SSTs by producing stronger upwelling. In later months, the S1 winds strengthen to similar values seen in S2, apparently as a result of SST drift in the coupled system. The combination of wind and heat flux differences seems to result in very similar SST drift in low-resolution experiments for the first 2 to 3 months, followed by stronger cooling in S1 in the later stages when the winds are similar but the heat flux difference remains (cf Fig 56b). If S2 did not have the wind errors, the evolution of mean SST in Nino 4 might be quite realistic with only a small drift.

For the July 97 starts, S1 SST anomalies become unrealistically high from month 3 onwards, whereas in S2 the anomalies do not stray too far from observed values. Local heat flux anomalies for both systems are similar, and track the (somewhat uncertain) observed estimates well. Local wind anomalies are stronger in S1 than S2, but still less than observed values. It seems the unrealistically large Nino 4 SST anomalies cannot be blamed on excessive forcing. Their most likely explanation is that the errors are simply the inverse of the mean state errors. That is, Nino 4 is normally in an upwelling regime where model errors result in excessively cold SSTs. But in 1997, the coupled system was in a very anomalous state, where the upwelling in Nino 4 had been switched off in both model and reality. In this case, the normal model errors are not operative, and if we allow for them in the usual way we get a substantial overestimate of SST. Fig 56 nicely demonstrates the inverse relationship between model bias and anomaly overestimation in this case.

For the October 97 starts, the situation is more complex (Fig 63). Nonetheless, it does seem that the tendencies of S1 (blue) diverge from those obtained using the A2 atmosphere (green and orange) in the last 3 months of the forecast, just when the mean SST error is different, and that the difference in SST tendency is equal and opposite to the difference in mean bias (Fig 56). Note also that the difference in anomaly between experiments using the high (red) and low (orange) resolution ocean also nicely matches the difference in the mean drift. In July, all of the differences in the anomalies between the experiments seemed explicable by the mean state differences, and that is clearly not true here. The SST anomalies are substantially different in the first months between S1 (blue) and A2_O1 (green), with A2_O1 being warmer. There appear to be two factors in this. The A2 atmosphere model heat flux anomalies cool the ocean less than those of A1, for the first two months only. The difference is about 15 W/m2. Together with this, A2 puts more heat into the ocean in the mean, and there may be non-linearities which help rectify this. (With the upwelling switched off, a given mean heat flux will have a greater impact on SST evolution and the lack of significant cooling may allow a shallower mixed layer to develop. This appears to be the case according to ocean diagnostics, not shown).

Although 1997 was an unusual situation, it does highlight a fundamental problem with using linear methods to correct for model errors. In particular, model errors in upwelling and non-upwelling situations are likely to be different, and as the front between the cold tongue and the edge of the warm pool moves backwards and forwards across the Pacific, our method of bias removal will be inaccurate. There is no easy way to avoid this. If we were to make flux corrections during the model integration, we would face the same problem of a time-varying region of upwelling. In principle, we should be able to define a non-linear correction scheme to help deal with this, but in practice the very limited number of data points and the multiplicity of factors in causing forecast errors make this very difficult. More accurate ocean models are clearly the best way to deal with this problem, and it seems that the ocean used in S2 is a step forward in this regard.

5.5 Equatorial evolution of the drift in coupled integrations

We can gain some insight into the zonal structure of the errors in the coupled model by looking at Hovmueller plots of ocean fields along the equator. In this section we consider the mean drift, and in the following section we will consider the evolution of the anomalies. For simplicity, we will focus on the forecasts starting in July, paying special attention to the forecasts initiated in July 1997.

All the Hovmueller diagrams presented in the following two sections will show results from the cross-coupled experiments described in Table 1. The reference experiment will be A2_O2. For the diagnostics presented in this section, experiments A2_O2 and S2 give very similar results.

Figure 64 illustrates the equatorial evolution of the mean state of relevant ocean fields. It shows the evolution of the mean state from the coupled model A2_O2 forecasts starting in July (upper row), the verifying analysis (middle row) and their difference (lower row). The fields represented are SST (left column), sea level (SL) (middle column) and the zonal component of the wind stress (right column). The SL is included because it filters the effect of the wind stress (it can tell which features of the wind variability matter for the ocean), and it is not very sensitive to surface processes, such as heat flux and mixed layer formulation

The drift of the coupled model in terms of SST shows two prominent features: a) a coastal warming, off the South American coast, that exceeds 3K, and b) a cooling in the central Pacific that reaches -2K (lower left panel). The coastal warming appears very quickly, in the first days of the integration, and intensifies over the first 3 months, remaining confined to the eastern Pacific, east of 110W. Looking at the evolution of the full SST fields (middle and upper left panels) one can see that the coastal warming is a consequence of the detachment of the cold tongue from the coastal area in the coupled model: the cold tongue, an upwelling region where the SST has a minimum, migrates from the eastern coast to $\sim$120W. Since the position of the cold tongue determines the centre of SST interannual variability, it can be anticipated that this displacement will have some impact on the ENSO prediction -see later fig 70. The cold bias starts in the central Pacific during the second month of the integration, intensifies and propagates eastward. The Atlantic is characterized by an overall SST warming, while the drift in the Indian Ocean is cold.

Why does the coupled model drift? Has the drift become worse as a result of the new ocean model or the new atmospheric model? To answer the first question, one can look at the relationship between the drift of SST and the errors in the winds and heat fluxes and to answer the second we will consider the cross-coupling experiments. We saw in the previous section that error in heat fluxes was unlikely to be the main cause of the drift in SST in the east and central Pacific, and we will not discuss this further. To asses the impact of the wind error on the ocean state, one can look at the evolution of the drift in terms of SL (middle column panels). A dipole structure (panel f, negative values west of 140W, positive values east of 100W) developes very quickly. This dipole structure, especially the positive sea level drift in the east, is broadly consistent with the drift in SST, and possibly produced by wind errors in the eastern half of the basin. However, the dipole structure dies after 2 months into the integration, with the eastward propagation of a cold signal in the sub-surface, to give way to an overall negative sea level anomaly in the equatorial Pacific. This SL deficit in the coupled model has been caused by the intensified easterlies in the Western Pacific, the effect of which reaches the Eastern Pacific months later.

The largest error in the SL happens in the central Pacific, where the wind errors change sign. Looking at the full SL fields, one can see that the SL gradient in the coupled model retreats towards the west in the first 2 months: the 50cm contour line moves from 160W to 160E. The fact that the intensity and evolution of the coastal warming drift in SST does not have such a clear signal in the sea level suggests that elements other than the wind error are involved in the SST drift, such as errors in the ocean model: the detachment of the cold tongue from the coast is a common error in coupled models and in ocean-only runs. The SL drift is also

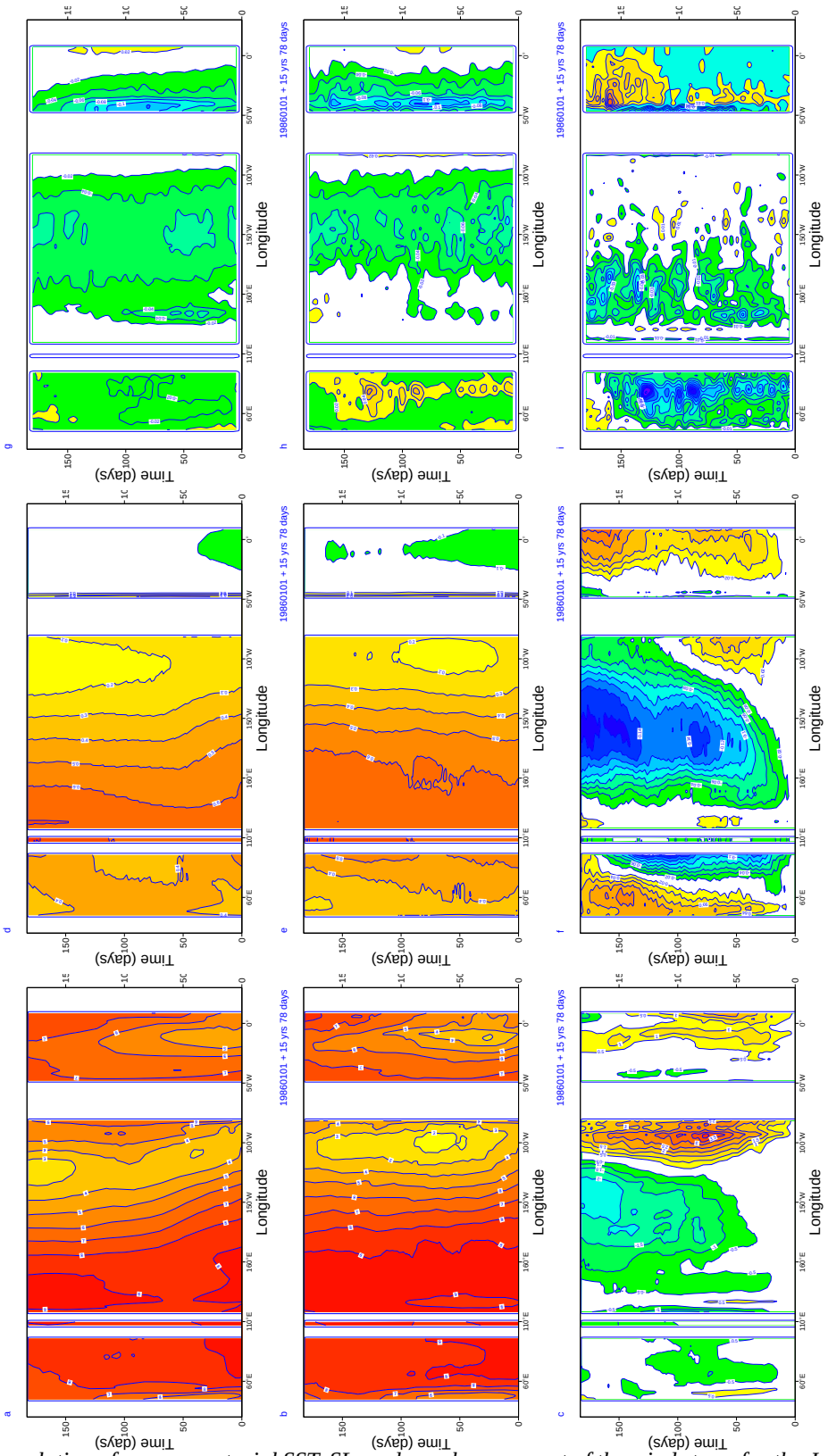


Figure 64: Time evolution of mean equatorial SST, SL, and zonal component of the wind stress for the July-start forecasts with the coupled model A2_O2 (upper row panels: a, d and g), verifying analysis (middle panels: b, e, and h), and drift of the coupled model (lower panels: c, f, and i). All the panels span a time interval of 6 months. The contour interval is 1K for the full SST fields (offset is 20K), 0.5K for the drift in SST, 0.1m for the full SL fields, 0.02m for the SL drift, 0.02 Nm^{-2} for the full zonal wind stress and 0.01 Nm^{-2} for the drift in zonal wind stress.

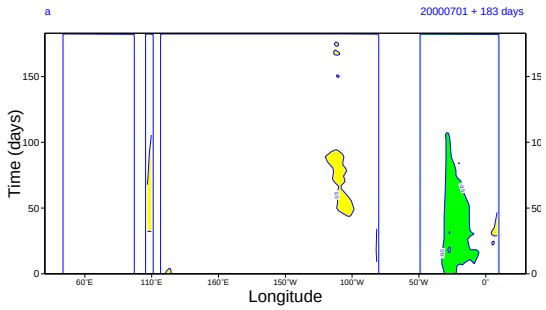


Figure 65: Contribution of the change in ocean model to the drift in SST, as measured by the difference in the 6-month evolution of the equatorial SST for the forecasts initiated in July between the experiments A2_O2 and A2_O1 The contour interval is 0.5K.

apparent in the Atlantic (overall increase) and Indian Ocean (very strong dipole structure). It is consistent with the drift in the zonal component of the wind stress (intensified easterlies west of 150W, including the Indian ocean, and weakened easterlies east of 150W, including the Atlantic ocean).

By examining the results from the cross-coupled experiments, one can assess the impact of the new ocean model or the new atmospheric model on the drift. Figure 65 shows the impact on the drift from changing the ocean model, as measured by the difference in the experiments A2_O2 and A2_O1. It shows that changing the ocean model has only a weak impact on the SST drift (barely 0.5K warmer in the Eastern Pacific).

The contribution to the drift from changes in the atmospheric component appears in figure 66, as the difference between experiments A2_O2 and A1_O2. The atmosphere A2 produces warmer SST overall. It increases the coastal warming in the Eastern Pacific, but the spatial pattern of the difference between A2_O2 and A1_O2 is not the same as the drift in SST shown in the lower left panel of fig 64: most of the warming happens west of 100W, and increases with time, while the drift happens east of 110W and peaks at around month 3. The cold tongue is warmer with the A2 atmosphere. The impact of changing the atmosphere has a substantially larger effect on SST than that of changing the ocean. On the positive side, the use of A2 reduces the cold drift in the central Pacific. The impact on the SST of the Indian ocean also seems beneficial. On the negative side, the drift in SL is made worse by the A2 atmosphere, especially in the Indian, Atlantic and central Pacific oceans. The use of A2 does not produce any noticeable change in SL in the Eastern Pacific until after the second month of the integration. The impact of the atmosphere on SL is consistent with the changes in the zonal component of the wind stress: too strong easterlies west of 150W and too weak easterlies east of 150W.

Summarizing, the introduction of A2 has a beneficial impact on the SST drift overall, with the exception of the Eastern Pacific. It seems to make the coastal warming bias worse, but it is not responsible for it. The impact on SL is detrimental. Changing the ocean model has little impact on the SST drift: it just makes the cold tongue slightly warmer. The erroneous westward shift of the cold tongue is present in all experiments. The story for the October starts is very similar.

5.6 Equatorial evolution of the July 1997 anomaly in coupled integrations

Figure 67 shows the time evolution of the equatorial SST anomaly (left column), SL anomaly (centre column) and zonal wind stress anomaly (right column) in the coupled forecasts A2_O2 (upper row), the verifying analysis (middle row) and their difference (bottom row). In the Pacific, as anticipated by the discussion on the drift, the maximum of SST anomaly in the model is detached from the coast, and does not coincide with the centre of the anomaly in observations. In addition, the amplitude of the SST anomaly is weaker in the model than in

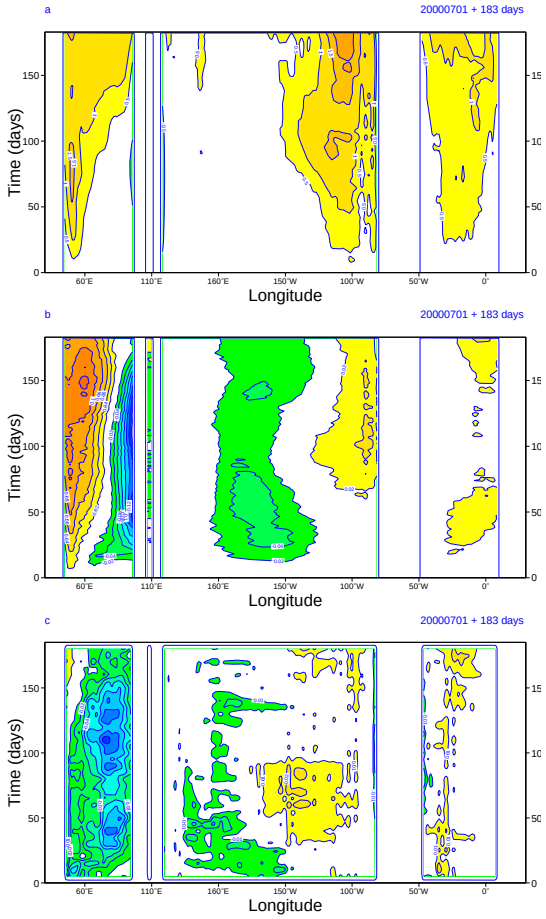


Figure 66: Contribution of the change in atmosphere model to the coupled model mean state, as measured by the difference in the 6-month evolution of the equatorial fields for the forecasts initiated in July between the experiments A2_O2 and A1_O2. The contour interval is 0.5K for SST differences (upper panel), 0.02m for the SL differences (middle panel) and 0.01Nm^{-2} for the zonal wind stress differences (lower panel).

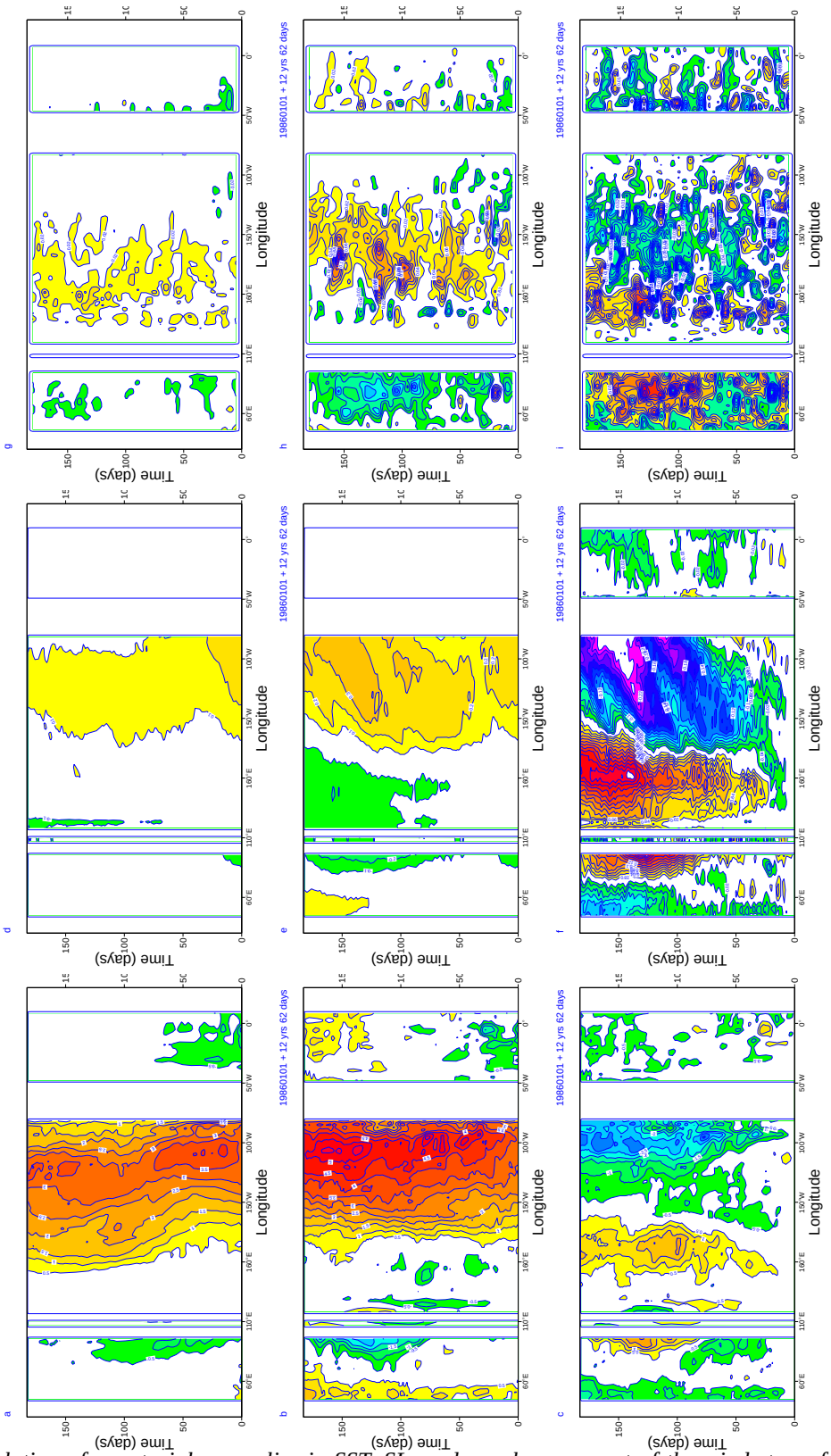


Figure 67: Evolution of equatorial anomalies in SST, SL, and zonal component of the wind stress for the July 1997 6-month forecast from the coupled model A2_O2 (upper row panels: a, d and g), the verifying analysis (from 1st of July 1997 to first of January 1997) (middle panels: b, e, and h), and the error in the coupled model anomalies (lower panels: c, f, and i). The contour interval is 0.5K for all SST fields, 0.1m for the SL anomalies, 0.02m for the SL error, 0.02 Nm^{-2} for the zonal wind stress anomaly and 0.01 Nm^{-2} for the error in zonal wind stress.

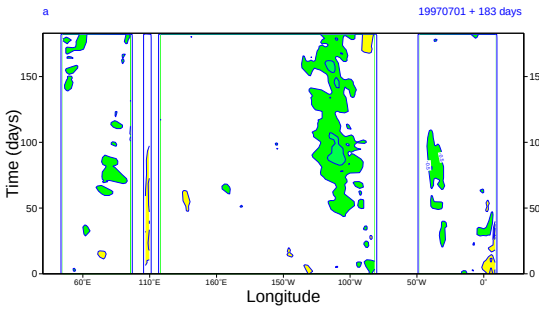


Figure 68: Contribution of the change in ocean model to the predicted SST anomaly for forecasts initiated in July 1997: the panel shows the difference between the anomaly predicted by experiments A2_O2 and A2_O1. The contour interval is 0.5K.

observations. The error in the anomaly is characterized by an overall cooling in the Eastern Pacific, with a very similar pattern to that of the bias but extending farther west, and intensifying with forecast time to over 3K. Errors are not confined to the Eastern Pacific. The Western Pacific in the coupled model is too warm by more than 1K. In the Indian Ocean, the coupled model does not predict the strong "Indian Dipole" that was observed.

Looking at the sea level anomaly, one can see immediately the lack of Kelvin wave activity over the Pacific in the coupled forecast A2_O2, which is due to the lack of westerlies in the coupled model (lower row, central and right panels). The absent Kelvin waves will never reach the eastern coast, and therefore, they will never produce the warming over the cold tongue region by shutting off the upwelling. The lack of wind variability explains the weak intensity of the predicted SST anomaly, but not its wrong location. Neither does it explain the quick development of the SST error in the Eastern Pacific (in month 1). These errors are clearly related to errors in the mean state (i.e. detachment of the cold tongue from the South American coast). In the Indian ocean, the SL anomaly in the coupled model is absent, indicating again a failure in the wind variability over this region.

Changing the ocean model has some impact on the evolution of the SST anomaly. Fig 68 shows that the new ocean model tends to produce a weaker SST anomaly over the cold tongue position (about 0.5K colder), but this is not because of the increased variability in SL or wind. The SL anomaly is almost the same. The weaker SST anomaly is probably due to weaker upwelling associated with the warmer ocean mean state of the new ocean model. The impact of changing the atmospheric model is more dramatic (fig 69): the SST anomalies are more than 1K colder east of 160E after the second month of the integration. This is mainly due to the collapse of the wind variability and therefore lack of downwelling Kelvin waves (lower panel).

It can be concluded that the main source of error for the July 1997 SST anomaly for the NINO3 is the lack of wind variability. This will affect the SST anomaly after the second month of the integration, but the error in the first 2 months of the integration (attributed to westward migration of the cold tongue in the coupled mean state) is not explained by the lack of wind variability in the new atmospheric model. Nor is it fully explained by the change of the ocean component of the coupled system. It is likely present in S1, but due to compensation of errors was not noticeable in the NINO3 index.

The story for the October starts is similar. A2 is unable to maintain the westerly wind anomaly present in the initial conditions at around the date line, leading to the fast decay of the SST anomaly in the eastern Pacific. In this case, the change of the ocean component does not seem to have any impact. Again, the error observed in the SST anomaly in the eastern Pacific is a combination of 2 facts: i) lack of wind variability in A2, in particular the inability to sustain the ongoing wind anomalies and ii) the westard displacement of the SST anomaly in the coupled model due to errors in the ocean mean state as the cold tongue separates from the coast. None of

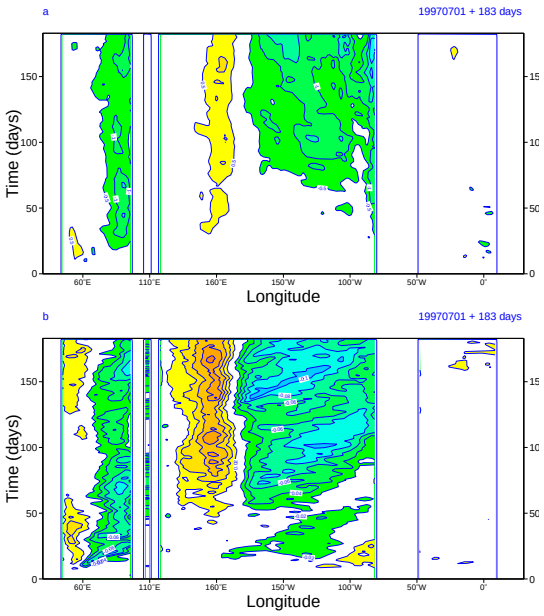


Figure 69: Contribution of the change in atmosphere model to the forecasts for July 1997. The upper panel shows the difference between SST anomalies predicted by A2_O2 and A1_O2. The contour interval is 0.5K. The lower panel shows the equivalent difference but for SL, with 0.02m as contour interval.

the changes in the oceanic and atmospheric components seems to have a significant impact on the cold tongue detachment error.

In fact, the cold tongue detachment error is also present in long integrations of the ocean model forced by analyzed winds, which indicates that this error is not a feature exclusive to the coupled model, nor it is produced by the initilization shock due to the imbalance between the ocean initial conditions (produced using data assimilation) and the winds. The error influences the position of SST variability, as can be seen in figure 70, depicting the standard deviation of monthly SST anomalies for observations and for a model run using analyzed winds, over the period 1993-1998. In observations (fig 70a), the SST variability shows a unique maximum, close to the South American Coast and extending continuously eastward at the Equator. In the model (fig 70b), the SST variability shows a double maximum: one by the South American coast, and a second, more pronounced, maximum located at around 140W. Sensitivity experiments indicate that parameterization of vertical mixing, ocean resolution and different wind products affect the intensity of the equatorial centre of SST variability, but do not greatly affect its position.

5.7 Summary of the 1997 El Nino forecasts

Drawing together the diagnostics for the atmospheric forcing in the Nino regions, the Hovmueller plots, and the outcome of the hybrid forecast experiments, we summarize the causes of the differences in Nino SST forecasts between System 1 and System 2 for start dates in the period June to October 1997. As well as explaining the difference between S1 and S2, we also try to understand the errors in the S2 forecasts.

The zonal wind anomalies in the west and central Pacific produced by A2 (Cy23r4) are significantly weaker than both A1 (Cy15r8) and observations. The weakness of the wind variability is an intrinsic property of A2, and is not caused by its coupling to the ocean, although in the later stages of some of the forecasts there may be an additional feedback from the weak SST anomalies to the weak wind anomalies. The erroneous wind

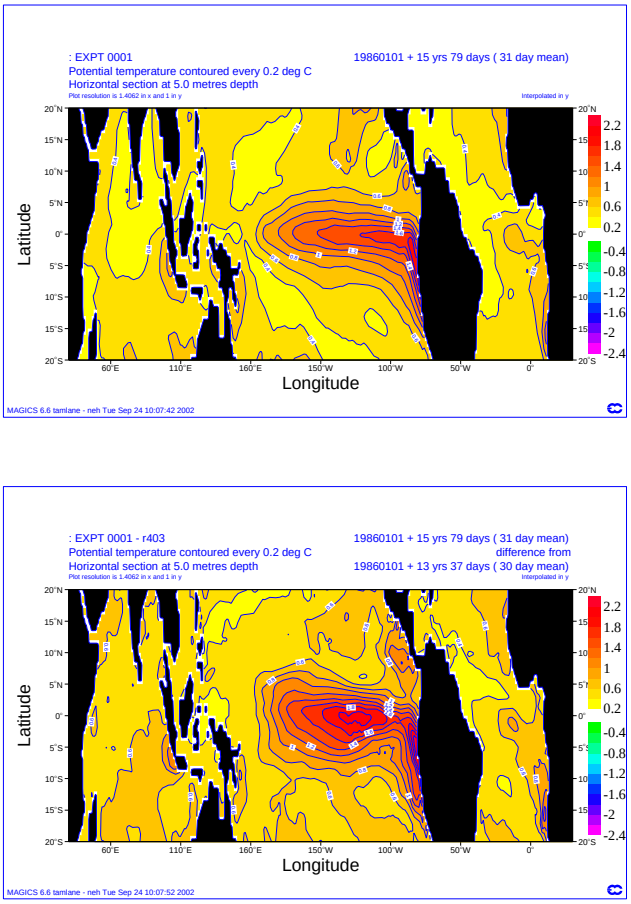


Figure 70: Standard deviation of the monthly SST anomalies during the period 1993-1998 for observations (upper panel) and an ocean-only run forced by anlyzed winds and heat fluxes (lower panel).

variability is undoubtedly the dominant factor in the difference between the S1 and S2 forecasts. It is also believed to be the biggest single cause of the underprediction of the Nino 3 SST anomaly by S2.

The cross coupling experiments demonstrate that using the A1 atmosphere gives larger amplitude Nino 3 SST anomalies in general, and larger anomalies in 1997 in particular. It does this at the expense of excessive SST anomalies in the central Pacific (Nino 3.4), although not further west (Nino 4) in the 1997 forecasts. In our cross-coupling experiments, the choice of atmospheric model has the biggest single impact.

The cross coupling experiments also show that the ocean initial conditions and ocean model explain a small part of the difference between S1 and S2 Nino 3 forecasts started in July, and a very small part in October. The effect of changing the ocean is to introduce a shift in the forecast Nino 3 index which does not grow in amplitude with time. The relative importance of the ocean model/initial conditions on Nino 3 SST is thus highest in the first month or so.

It is clear from the Hovmueller plots that errors in SST in the east Pacific at around 100W will impact the Nino 3 values. Only the July cases are shown (Fig 67), but similar results are seen for the October forecasts. In this specific region around 100W, there is a strong similarity between the errors in the mean state and the errors seen in the forecast anomalies, particularly in the first 3 months. These are largely unaffected by swapping the oceanic or atmospheric model components, and the most likely explanation for the errors is that they are due to error inherent in the ocean models. Physically, we expect the mean state error to impact the anomalies in two ways: the apparent shift of the upwelling to the west should reduce variability in the far eastern Pacific, and the warm bias of itself may limit the amplitude of large warm anomalies.

Together with specific, rapidly-appearing errors at 100W, there are clearly more general problems with SST in the east Pacific. Analysis of forcing fields suggests that the large bias in the Nino 1+2 region is not caused by errors in the forcing fields, but is inherent in the ocean. The difference in Nino 1+2 SST bias between S1 and S2 is due to the change in ocean model. SST bias closer to the equator (for example in the EQ1 region) appears fairly insensitive to the ocean model, but does depend on the atmospheric model. The Hovmueller plots show that although there is a general correspondance between a warm mean state in the east Pacific and the underestimation of warm anomalies in 1997, the relationship is not direct in terms of the specific changes introduced by changing model components or applying artificial heat flux perturbations.

Note also that S1 and S2 have a different mean zonal wind in the eastern equatorial Pacific. Everything else being equal, the stronger mean wind of S1 would lift the thermocline slightly and this might tend to increase SST variability somewhat. It is possible that the mean wind difference explains part of the difference in Nino 3 SST anomaly forecasts in October, but errors in mean wind in the east Pacific are not important for the actual error in the S2 forecasts, since the S2 winds in this region are realistic.

In total, the evidence suggests that significant errors are present in the ocean model itself in the eastern Pacific – as far as we can judge, it does not seem that we would obtain a correct ocean mean state even with the correct forcing, and the variability of SST in the ocean model with observed forcing does not have the correct spatial pattern in the far east Pacific. The biggest part of the apparent ocean model error has not changed between S1 and S2, but although it is not a dominant reason for the difference between S1 and S2, it is an important issue to be addressed. The fact that comparison of differences in our coupled experiments shows a complex story in the east Pacific (both on the equator and to the south of it in Nino 1+2), suggests that the interactions in this region are complex, and it is likely to be difficult to produce reliably good SSTs in either the mean state or the forecast anomalies.

Although the far east Pacific is the area of largest errors in SST (as seen in the Hovmueller plots, for example), it may not be the most important in terms of its global impact. We might expect a modest error in the central Pacific to have much stronger teleconnections than a large amplitude but spatially restricted error in the rela-

tively cold water of the far east Pacific. Thus although errors in the east Pacific SST are likely to be a challenge for many years to come, this will not necessarily prevent us from developing a high quality seasonal forecasting system.

We also consider the west-central Pacific, exemplified by the Nino 4 region. Here we have possible evidence of the failure of our bias removal strategy, in that there is a strong correlation between the mean state error and the error in the SST anomalies in 1997. The models differ in more than their mean states, of course, and some of the differences in the forecast anomalies are likely to be caused by other factors, such as the stronger wind anomalies in S1. Nonetheless, there is argument to support the idea that changes in the ocean model result in changes in the anomalies *because* of the change in the mean state that they produce. If the mean state has a too-extensive cold tongue, then the SST bias correction will account for this (mean) state of affairs. If another model version has a less extensive cold tongue, the bias correction will be less. In both models, as the warm pool extends eastward to regions which were previously in the upwelling cold tongue, the corrections for the usual model error are no longer appropriate, and result in erroneous SST anomalies when they are applied. The bigger the bias correction, the bigger the resulting error in SST anomaly.

Fundamentally, our assumption that there is no interaction between errors in the mean state and the anomalies may be breaking down. While it may be possible to ameliorate the effect in some of the forecast products (for example by attempting to calculate flow dependent bias corrections), the ultimate solution is to reduce the errors in the mean state. As regards the west and central Pacific, it is encouraging that the high resolution ocean used in S2 has made progress in this regard.

Quantitative partition of both the error in S2 and the difference between S1 and S2, in terms of the error in the coupled ocean mean state, errors inherent in the ocean model and its variability, ocean initial condition errors and atmosphere forcing errors has not been possible in this paper. What we have been able to do is identify some of the major errors in our forecast system which contribute to the errors in the 1997 El Nino forecast. This gives a good basis from which to assess and develop the model components, so as to enable us to improve our forecasts in the future.

6 Summary and conclusions

Overall coupled behaviour

We have described the two forecast systems S1 and S2 and showed some overall SST scores for them. In general S2 is better than S1 in Nino 3 but not in Nino 4. During the 1997 event the skill was reversed: S1 did better in Nino 3 but S2 did better in Nino 4. A series of hybrid experiments was performed in which the oceanic and atmospheric components were interchanged and forecasts performed every season for the years 1987-2001. These showed that the switch from O1 to O2 was beneficial, especially in Nino 3. Whether this resulted from changes in the ocean parameterisations or in the ocean data assimilation system was not examined. The switch from A1 to A2 was not beneficial. There was a noticeable degradation in the Nino 4 region, with a smaller effect in Nino 3. S1 was a relatively well-balanced system in terms of the amplitude of SST anomalies with a tendency to be slightly overactive. By contrast S2 is underactive. This seems to be a consequence of changes in both the atmospheric and oceanic models. With respect to the atmospheric change, interannual wind variability is damped. The winds are also damped in S1 but the damping starts earlier in S2, and is more marked.

The increase in ocean resolution which is included in S2 does not help the overall forecast statistics. This may be because of compensation of errors. The low resolution ocean has a stronger cold bias in the central Pacific, and a stronger sensitivity to wind anomalies. It is likely that this helps compensate for the too-weak wind variability of the Cy23r4 atmosphere. Although we hope that the high resolution ocean can be of benefit when

coupled to a more realistic atmosphere model, this has not been demonstrated.

For S2, both ocean and atmosphere have errors on their own, and these seem to dominate the errors in the coupled system. The atmosphere model error in the tropical low level winds is characterized by stronger-than-observed trades in the the western half of the Pacific basin, an easterly bias over the Indian ocean and a westerly bias over the Atlantic. The bias in the winds in the western Pacific occurs in association with a substantial underprediction of interannual wind variability near the dateline. The weak variability of the surface winds is believed to be the main reason for error in the Nino 3 and Nino 4 SST coupled forecasts.

The ocean model error is characterized by a westward shift of the cold tongue upwelling area, resulting in a warm bias near the South-American coast, and in a westward displacement of the area of maximum SST variability. These ocean model errors change the structure, and perhaps the amplitude, of coupled SST anomaly forecasts.

1997 El Nino forecasts

Forecasts for the 1997/8 El Nino show that changing the atmosphere has a bigger effect than changing the ocean. Comparison with other model forecasts for 1997/8 shows a huge range in behaviour, with no model capable of doing a satisfactory job throughout. This strengthens the case for multi-model forecasts.

Previous work has shown that the rapid growth of El Nino from May to July 1997 was not well captured by any of our forecast systems, and was due to wind anomalies whose predictability is not clear. That we are unable to reproduce the observed winds with any ensemble member of any model version we have tested tells us we have a problem. In this paper we have concentrated on forecasts starting in the July to October period, when there were substantial differences between S1 and S2, and the S1 Nino 3 SST anomalies were very successful. Consideration of all of the diagnostics for forecasts started in this period bring out the following points.

The interannual variability of the SST produced by the coupled model depends primarily on the wind variability produced by the coupled model, and on the mean state of the ocean. The wind variability influences the intensity of the SST anomalies. The zonal wind anomalies produced by Cy23r4 in 1997 are significantly weaker than both Cy15r8 and observations. This is undoubtedly the dominant factor in the difference between the S1 and S2 forecasts. The weak wind anomalies are also believed to be the biggest single cause of the underprediction of the Nino 3 SST anomaly by S2, although in the October starts other factors may also be important. The weak wind anomalies of S2 are inherent in the atmosphere model.

The mean state of the ocean depends on ocean model parameterization and on mean forcing (winds, heat flux). It affects the SST variability in several ways. The spatial location of the upwelling cold tongue determines where the maximum of remotely forced SST variability will occur. The strength of upwelling influences the amplitude of SST anomalies that develop. Mixed layer structure also affects the evolution of SST anomalies. The ocean mean state in general, and the strength of upwelling in particular, is quite sensitive to the wind forcing and ocean parameterizations, and errors are particularly prevalent in the eastern equatorial Pacific.

Substantial errors in the 1997 SST forecasts at around 100W appear very quickly in the forecasts, and seem to be inherent in the ocean model. These particular errors are similar between S1 and S2, and do not explain the difference between the Nino 3 forecasts. Although unrelated to S1/S2 differences, these errors should be addressed. There are other errors in the east Pacific which are dependent on the ocean and atmosphere models used, and which evolve throughout the coupled forecasts. Experiments show that sensitivities of both mean state and the 1997 anomalies to changes in atmosphere or ocean or forcing terms are quite complex, supporting the view that the East Pacific is a region where many different problems are interacting. Because the coupled forecasts have substantial errors in the wind anomalies, it is hard to assess just how much error in the S2 SST anomalies is due to errors in the ocean state. Note, though, that the ocean initial conditions and ocean model explain only a small part of the difference between S1 and S2 Nino 3 forecasts started in July and October.

It is clear that the apparent success of S1 in 1997, judged from the Nino 3 SST anomaly plumes, is in fact based on compensation of errors. The SST anomalies are too strong in the central Pacific, and too weak in the far east, with the Nino 3 region straddling these regions so as to effect a substantial compensation. Real improvements to S1 might thus be expected to result in less good Nino 3 anomalies for this case, but perhaps with better scores elsewhere. If S2 was "worse" than S1 purely for such a reason, we would not be overly concerned. However, we have been able to demonstrate that at least some of the deterioriation is a symptom of real problems with our forecasting models, problems which we hope can be addressed in the future.

Atmosphere model issues

We have also considered systematic errors in various atmospheric quantities and shown that there are serious biases in the atmosphere of the coupled system, most of which are of atmospheric origin and are present in uncoupled experiments. Even the low-level wind errors are initially largely of atmospheric origin, although coupling does make some difference from month two onward.

Spatial plots of the variability in the near-surface wind for S1 and S2 showed that both systems under-represented the variability in the SPCZ but had too much variability in the northern ITCZ. S2 was too inactive near the date-line. Analyses of upper air thermal errors showed that these also were mainly of atmospheric origin and developed in the first month.

Differences in the extratropics are of less importance for influencing ENSO forecasts but are of interest in their own right. The bias from the atmospheric simulations A2 shows a substantial reduction of mean error over the European sector relative to A1, although over the Pacific sector the excessively zonal flow is present in both versions. The major features of the mean error in the Northern Hemisphere are similar between the coupled and uncoupled simulations indicating that most of the error is of atmospheric origin. The standard deviations of A2 have a spatial distribution of variability over the Atlantic Ocean, Greenland and Iceland closer to the analysis and generally have reduced maxima of inter-annual variance with respect to A1. The position of the Atlantic and Pacific storm tracks are represented well in S2. The simulated maximum of low frequency transience over the Atlantic, associated with the position of Euro/Atlantic blocks, shows a southeastward shift compared with the analysis, but over the northeast Pacific, the low frequency transience is well represented.

In summary, S2 has a fairly good representation of many features important for seasonal forecasting. Like all numerical models, it also has errors and weaknesses which affect its performance. Many problems are common between S1 and S2, but each model also has its own particular weaknesses. The biggest single difference between S2 and S1 for El Nino forecasting seems to be weak interannual variability of equatorial zonal winds near the dateline in S2. Ocean model changes have also been significant, although they have had only a small impact on the 1997 forecast. The overall Nino 3 SST anomaly forecast skill of S2 is higher than that of S1. Nevertheless, careful analysis shows that many problems and model errors remain, and much work will be needed to create a reliable and trustworthy forecasting system.

A Appendix: Locations of the various Nino and EQ regions

Plot of the regions in fig 71 and 72.

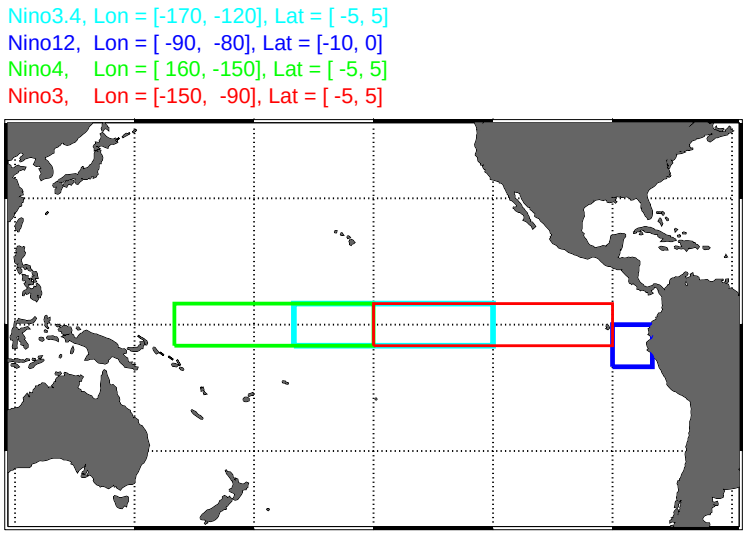


Figure 71: Plot of the Nino regions.

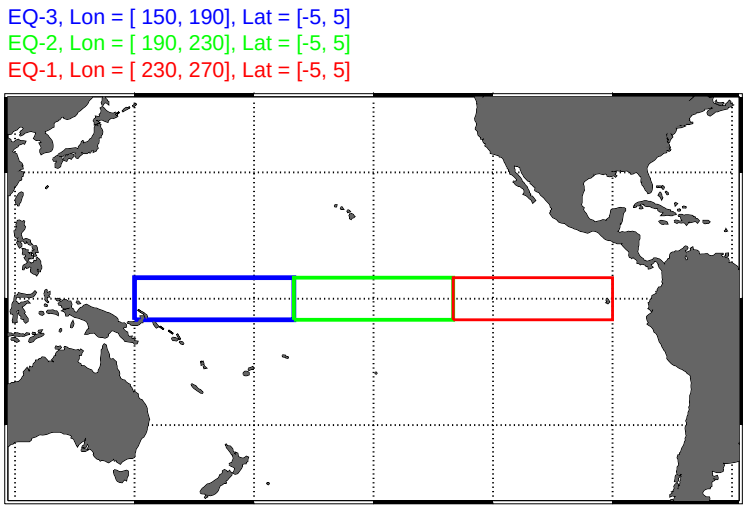


Figure 72: Plot of the EQ regions.

REFERENCES

Balmaseda M., F. Vitart, L. Ferranti and D. Anderson (2002) Westerly wind events and the 1997 El Nino event in the ECMWF seasonal forecasting system: a case study. *ECMWF Technical Memorandum*, 370. (Submitted to J Clim)

Bjerknes J. 1964: Atlantic air-sea interaction. *Adv. Geophys.*, **10**, 1-82.

Burgers G., M.Balmaseda, F.Vossepoel, G.J.van Oldenborgh, P.J.van Leeuwen, 2002: Balanced ocean-data assimilation near the equator. *J Phys Oceanogr*, **32**, 2509-2519.

Horel, J. D., and J. M. Wallace, 1981: Planetary-Scale Atmospheric Phenomena associated with the Southern Oscillation. *Mon. Wea. Rev.*, **109**, 813-829.

Gregory, D., J.-J. Morcrette, C. Jakob, A.C.M. Beljaars and T. Stockdale, 2000: Revision of convection, radiation and cloud schemes in the ECMWF Integrated Forecasting System. *Q.J. Roy. Meteor. Soc.*, **126**, 1685-1710.

Latif, M., T. Stockdale, J. Wolff, G. Burgers, E. Maier-Reimer, M. Junge, K. Arpe and L. Bengtsson, 1994. Climatology and variability in the ECHO coupled GCM. *Tellus*, **46A**, 351-366.

Latif, M., A. Sterl, M. Assenbaum, M.M. Junge, E. Maier-Reimer, 1994: Climate Variability in a Coupled GCM. Part II: The Indian Ocean and Monsoon. *Journal of Climate*, **7**, 1449-1462.

Levitus S. and T. Boyer, World Ocean Atlas 1994 Volume 4, 1994: Temperature. *NOAA Atlas NESDIS* , **4**, U.S. Department of Commerce, Washington, D.C.

Levitus S NODC (Levitus) World Ocean Atlas 1998 data provided by the NOAA-CIRES Climate Diagnostics Center, Boulder, Colorado, USA, from their Web site at http://www.cdc.noaa.gov/

Namias, J., 1969: Seasonal Interactions Between the North Pacific Ocean and the Atmosphere During the 1960's. *Mon. Wea. Rev.*, **97**, 173- 192.

Pacanowski R. and S G H Philander 1981. Parameterization of vertical mixing in numerical models of tropical oceans. *J. Phys. Oceanogr.*, **11**, 1443-1451.

Peters, H, Gregg, M C and Toole, J M, 1988: On the parameterization of equatorial turbulence. *J. Geophys. Res.*, **93**, 1199-1218.

Reynolds, Richard W., 1988: A Real-Time Global Sea Surface Temperature Analysis. *Journal of Climate*, **1**, 75-87.

Reynolds, Richard W., Thomas M. Smith, 1994: Improved Global Sea Surface Temperature Analyses Using Optimum Interpolation. *Journal of Climate*, **7**, 929-948.

Reynolds R. and T. Smith 1995 A high resolution global sea surface temperature climatology. *J. Climate*, **8**, 1571-1583.

Smith, Thomas M., Richard W. Reynolds, 1998: A High-Resolution Global Sea Surface Temperature Climatology for the 1961-90 Base Period. *Journal of Climate* **11**, 3320-3323.

Smith N., J Blomley, and G Meyers (1991) A univariate statistical interpolation scheme for subsurface thermal analyses in the tropical oceans. *Prog in Oceanography*, **28**, 219-256.

Stockdale T. N., D. L. T. Anderson, J. O. S Alves, and M. A. Balmaseda, 1998. Global seasonal rainfall forecasts using a coupled ocean-atmosphere model. *Nature*, **392**, 370-373.

Troccoli Alberto, M. Balmaseda, J. Schneider, J. Vialard and D. Anderson 2002, Salinity adjustments in the

presence of temperature data assimilation. *Mon. Wea. Rev.*, **130**, 89- 102.

Van Loon, H., and R. A. Madden, 1981: the Southern Oscillation. Part I: Global Associations with Pressure and Temperature in the Northern Winter. *Mon. Wea. Rev.*, **109**, 1150-1162.

Wolff, J., E. Maier-Reimer and S. Legutke, 1997. The Hamburg Ocean Primitive Equation Model. *Deutsches Klimarechenzentrum, Hamburg*, Technical Report No. 13.

Xie and Arkin, 1997: Global Precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs, *Bulletin of the American Meteorological Society*, **78**, 2539-2558.

Xie and Arkin, 1996: Analyses of global monthly precipitation using gauge observations, satellite estimates, and numerical model predictions, *Journal of Climate*, **9**, 840-858.