TECHNICAL MEMORANDUM

503

# Development of the ECMWF seasonal forecast System 3

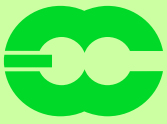
David Anderson, Tim Stockdale, Magdalena Balmaseda, Laura Ferranti, Frederic Vitart, Franco Molteni, Francisco Doblas-Reyes, Kristian Mogenson and Arthur Vidard

Research Department

Paper presented to the SAC, 35th Session

April 2007





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

Series: ECMWF Technical Memoranda

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

Contact: library@ecmwf.int



European Centre for Medium-Range Weather Forecasts
Shinfield Park, Reading, RG2 9AX, England

**Abstract**

*Overview*

ECMWF has been running a seasonal forecast system since 1997. During that time there have only been two versions of the forecast system, called System 1 (S1) and System 2 (S2). A system consists of the atmospheric and oceanic components of the coupled model as well as the data assimilation scheme to create initial conditions for the ocean, the coupling interface linking the two components and the strategy for ensemble generation. There is no dynamic sea-ice model; the initial conditions are based on the observed sea-ice limit but thereafter the sea-ice evolves according to damped persistence.

S1 went operational in late 1997, S2 started running in August 2001, although S1 continued to be run until the Fujitsu computer was decommissioned in 2003. During the last few years, work has proceded with developing S3. Major changes have taken place in the ocean analysis system, but not in the ocean model per se. At this time the definitive decision on the atmospheric model has not been taken, but assuming that cycle 31r1 passes its e-suite tests for medium range forecasting, then that is the version that will be used in S3. It is likely that S3 will run in parallel with S2 for a few months before being deemed the operational version.

As in S2, the ocean initial conditions in S3 are provided not from a single ocean analysis but from a 5-member ensemble of ocean analyses. The atmospheric initial conditions, including land conditions come from ERA40 for the period 1981 to 2002 and from Operations from 2003 onwards.

The ensemble generation strategy is not the same as S2: there are changes in the calibration period and size and in the way the ensemble is generated. In S3, the real-time ensemble set will consist of 41 members, and the calibration set will consist of 11 members spanning the 25-year period 1981-2005, so creating a calibration PDF of 275 members. Each of these ensembles has a start date of the first of the month. The initial atmospheric conditions are perturbed with singular vectors and the ocean initial conditions are perturbed by adding SST perturbations to the 5 member ensemble of ocean analyses. Stochastic physics is active throughout the coupled forecast period. S3 seasonal integrations will be 7 months long (rather than the present 6 months). Additionally, once per quarter an 11 member ensemble will be run to 13 months, specifically designed to give an ENSO outlook. Back integrations will also be made to this range, once per quarter, with a 5 member ensemble.

The data from S3 will be archived into the multi-model seasonal forecast streams (MMSF). This will give consistency in the data archive between all members of the multi-model forecasting system (called EURO-SIP). For users accessing data, the switch to the new streams should be straightforward. ECMWF is acting as a focus for the development of real-time multi-model seasonal forecast system. Currently, the participants in EURO-SIP are ECMWF, the Met Office and Meteo France, but other members are expected to join in the future.

*Ocean Analysis*

A new ocean analysis system has been implemented to provide initial conditions for S3 forecasts. The ocean analysis extends back to 1959 and provides initial conditions for both real-time seasonal forecasts and the calibrating hindcasts which are based on the period 1981-2005. (Although only the ocean analyses from 1981 onwards are used for S3, the earlier ocean analyses will be used by the ENSEMBLES project, and for analysing climate variability.)

As for S2, the ocean data assimilation system for S3 is based on HOPE-OI, but major upgrades have been introduced. In addition to subsurface temperature, the OI scheme now assimilates altimeter derived sea-level anomalies and salinity data. In S3, the observations come from the quality controlled data set prepared for the ENACT and ENSEMBLES projects until 2002 and from the GTS thereafter. The OI scheme is now 3-dimensional, the analysis being performed at all levels simultaneously down to 2000m, whereas in S2, the analysis was carried out on each model level independently and only to 400m. In S3 there is also a multivariate bias-correction algorithm consisting of a prescribed a-priori correction to temperature, salinity and pressure gradient, as well as a time-dependent bias term estimated on-line. The on-line bias correction is adaptive and allows for flow-dependent errors. Because of the a-priori bias-correction term, the subsurface relaxation to climatology has been weakened: from a time scale of 18 months in S2 to 10-years in S3.

Because of large uncertainties in the fresh water flux, the relaxation to climatology is stronger for surface salinity (approx 3-year time scale), but still weaker than in S2 (approx. 6 months).

In order to obtain a first-guess as input to the OI analysis, it is necessary to force the ocean model with atmospheric fluxes. From January 1959 to June 2002 these are taken from ERA-40, and from the NWP operational analysis thereafter. In S2 the fluxes were from ERA15/OPS. The representation of the upper ocean interannual variability is improved when using the ERA40 wind stress, although the stresses are biased weak in the equatorial Pacific. The fresh water flux from ERA-40 (Precipitation - Evaporation, denoted P-E) is known to be inaccurate. A better but by no means perfect estimate was obtained by 'correcting' the ERA-40 precipitation values (Troccoli and Kallberg 2004) as part of the EU ENACT project.

As in S2, a 5-member ensemble of ocean analyses is created using slightly different wind stress fields. The differences in these fields have been revised to take into account improvements in the accuracy with which the wind stress can be determined. The motivation for and methodology used in generating the ensemble of ocean analyses is given in Vialard et al 2005.

The bias in the both the Eastern and Western equatorial Pacific in S2 has been significantly reduced in S3, where the east-west slope of the thermocline is better represented. The assimilation of salinity data is especially beneficial in the Western Pacific. The correlation of the model currents with the observed currents at different mooring locations on the equator in the equatorial Pacific is better in S3.

The variability in the upper ocean temperature in the north Atlantic is dominated by the upward trend, starting around the mid 80's. Salinity variations occur mainly on decadal time scales, and seem to be correlated with variations in the thermohaline circulation (THC). Comparison of the time evolution of meridional transport in the North Atlantic at 30N with the 'observed' values shows that although there is broad agreement between the two, the S3 ocean analyses indicate that the decadal variability is large, and therefore sampling is an issue when drawing conclusions about the slowing down of the THC. The S3 ocean analysis will provide initial conditions for the decadal ENSEMBLES forecasts, where the capability of the coupled models to reproduce changes in the THC will be explored.

Data assimilation has a significant impact on the mean state and variability of the upper ocean heat content. In the Equatorial Pacific, it steepens the thermocline and increases the amplitude of the interannual variability. In the Indian ocean it sharpens the thermocline, making it shallower, and it increases both the ENSO-related and Indian Dipole variability. In the Equatorial Atlantic the cold phase of the seasonal cycle is more pronounced, and the amplitude of the interannual variability increased. The impact of the S3 analysis on seasonal forecasts is beneficial nearly everyehere, but especially in the west Pacific. A region where there is little impact is the equatorial Atlantic.

Observing system experiments have been carried out to determine the information provided by the recently developed ARGO ocean subsurface float network, how it interacts with the information given by the altimeter and the impact of the observing sytem on climate variability. Results indicate that the observing systems for the ocean are not yet redundant. Even in the presence of ARGO and TAO data, the altimeter information can improve the vertical temperature structure in the far eastern Pacific where there are substantial in situ temperature observations from TAO moorings. In the Indian and Atlantic oceans the contribution of altimeter and ARGO is similar, and the best results are obtained when both observing systems are included. ARGO has a large impact on the salinity field on a global scale.

*Assessment of Forecast skill*

The integrations for S3 had not yet been made at the time of writing; results are based on an rd experiment erwq, consisting of a 5 member ensemble for four start dates per year. Comparing anomaly correlation and rms errors in forecasts of Nino3.4 and Nino4 in erwq with those for S1 and S2, shows clear progress. Scatter diagrams, showing all available forecasts for which S2 and erwq can be compared, show the improvements in erwq are significant in all areas of the tropical Pacific. However, the strong improvement does not extend to all parts of the globe - outside the equatorial Pacific, changes in skill are largely neutral, and while positive in the north subtropical Atlantic, may be marginally negative in the equatorial Atlantic.

The climatology of the atmospheric component of erwq shows substantial improvements with respect to S2.

Systematic errors in geopotential height, sea-level pressure and lower-tropospheric temperature have been substantially reduced in both the tropical and the northern extra-tropical regions. Examples of systematic error reductions at month-4-to-6 range are given. For the 500-hPa height field, a notable reduction in the model bias is found over the North-Pacific, where a large positive bias exceeding 12 dam in S2 has been reduced by almost a factor of 3. Mean errors over North America, which in S2 acted to decrease the amplitude of the stationary wave pattern, have also been substantially reduced, leading to a notable improvement in the zonally-asymmetric component of the time-mean flow. A negative bias of about 6 dam over Western Europe has been shifted to the north-west, unfortunately without any noticeable reduction. The location of such a negative bias in erwq is close to the region of highest blocking frequency in the East Atlantic sector, and therefore prevents any improvement in the simulated blocking statistics. Both S2 and erwq simulate the maxima of blocking frequency over the Euro-Atlantic and North Pacific regions. Both S2 and erwq winter hindcasts underestimate the blocking frequency over most of the Northern Hemisphere. The bias is more obvious over the North Pacific, although the western Atlantic blocking is also underestimated. These differences are significant with a 95% confidence over most longitudes. The results are representative of the model behaviour in other seasons. Erwq is no better than S2 in this regard.

In the mslp field for the boreal summer, positive errors in the regions of the subtropical anticyclones over both the northern and southern oceans were present in S2, with amplitude between 4 and 8 hPa. These errors have been substantially reduced in erwq. A positive bias over the Arctic Ocean has also been reduced by about a factor of 2, but the negative bias over the southern polar regions has been partially increased.

In erwq, both the seasonal mean and the interannual variability of rainfall over the tropical oceans is generally reduced compared to S2 values, bringing the model climatology into closer agreement with GPCP observational data. The spatial distribution of modelled rainfall is notably improved in the tropical Pacific during the boreal winter. While in S2 rainfall in the eastern Pacific ITCZ exceeds observations by (at least) a factor of 2, erwq simulates a more correct ratio between rainfall in the western and eastern parts of the ocean. The improvement in the mean field is reflected in the distribution of rainfall interannual variability. Comparing the standard deviation of January-to-March (JFM) rainfall in the ensembles run respectively with S2 and the S3 prototype (erwq) shows that the S2 variability shows two distinct maxima (with similar amplitude) in the western and eastern tropical Pacific, whereas erwq simulates a single variability maximum located just west of the dateline, in closer agreement with observations.

In the northern extratropics, internal variability has increased for both the high-frequency and the low-frequency range in going from S2 to erwq. Standard deviations of 500-hPa in boreal winter (DJF) for two spectral bands, one with periods between 2 and 8 days (typical of synoptic-scale baroclinic systems), the other one with periods between 10 and 30 days (comparable to the typical duration of large-scale circulation regimes such as blocking) show that both S2 and erwq produced realistic spatial distributions of variability compared to ERA-40 data. In the high-frequency range, the maximum associated with the North Pacific storm track is better positioned in erwq than in S2, although with a larger-than-observed amplitude; erwq also shows more variability over eastern Europe and north-west Asia. Low-frequency variability shows a moderate and rather uniform increase from S2 to erwq, with erwq values in better agreement with ERA-40 data.

Internal atmospheric variability is generally higher in erwq than in S2, both in tropical and extratropical regions. For the tropics, a notable improvement is found in the amplitude of intraseasonal variability in the 20-to-70-day frequency range, which includes the Madden-Julian Oscillation. The standard deviation of tropical velocity potential anomalies at 200 hPa in the October-to-March season is calculated for ERA-40, S2 and erwq, using a bandpass filter to isolate oscillations with periods between 20 and 70 days. Although the location of the variability maxima over the Indian and west Pacific oceans is in good agreement with reanalysis data, the amplitude is underestimated by both systems. However, in erwq the amplitude difference with respect to ERA-40 is reduced by 30-40 % with respect to S2. The spectral distribution of the velocity potential variability is further analysed as a function of longitude and oscillation period. Although the erwq results represent an improvement with respect to S2 simulations, erwq fails to generate a variance maximum in the MJO frequency range, which was simulated by cycle 30r2. However, 30r2 was not an acceptable cycle as it gave inferior forecasts of west equatorial Pacific sea surface temperatures as well as developing unrealistic upper troposphere moisture distributions.

The number of tropical storms detected in S2 and erwq has been averaged over the period 1987-2004, the 5 ensemble members and the 4 annual starting dates for each individual ocean basin. Comparing the mean annual frequency of tropical storms for each ocean basin with observations suggests that erwq produces more tropical storms than S2 over all the ocean basins. The increased horizontal resolution (T159 instead of T95) is a likely explanation. erwq seems to produce a more realistic climatology of tropical storms than S2 over all the ocean basins, although erwq produces less tropical storms than observed over the western North Pacific, eastern North Pacific, North Atlantic and South Indian Ocean, and more tropical storms than observed over the North Indian Ocean, Australian Basin and South Pacific. In terms of climatology of tropical storm frequency, erwq seems to be better than S2.

The performance of erwq was also assessed in terms of the overall predictive skill of the system for seasonal means of weather parameters such as rainfall and surface air temperaure. For a seasonal prediction system, probabilistic indices are usually preferred as a measure of skill; however, given the small size of the ensemble experiments used for this preliminary assessment (5 members only), such indices may be subject to considerable sampling errors and so only a preliminary estimate can be given. Based on Relative Operative Characteristics (ROC) scores, for specific anomaly categories in extratropical and tropical regions, scores for Europe and North America were evaluated for below-average 2-m temperature anomalies in (boreal) winter (JFM) and above-average temperature anomalies in summer (JAS), while for the tropical band scores for below-average rainfall were assessed for both seasons.

The ROC scores confirm the indications which emerged from the analysis of individual processes: in general, erwq has more predictive skill than S2 for the tropical regions, while in the northern extratropics improvements are mostly evident during the summer season. The decrease of skill scores during the boreal winter appears to be related to a partial reduction of the wintertime diabatic heating anomalies in the central tropical Pacific during ENSO episodes, and to an increased level of internal atmospheric variability, both of which decrease the signal-to-noise ratio for NH interannual variability during winter. In S2, such a ratio was enhanced by larger-than observed rainfall amounts in the tropical Pacific, which partially compensated the reduction in the SST anomaly amplitude occurring during the coupled integrations.

# 1 Introduction

ECMWF has been running a seasonal forecast system since 1997. During that time there have only been two versions of the forecast system, called System1 (S1) and System 2 (S2). A system consists of the atmospheric and oceanic components of the coupled model as well as the data assimilation scheme to create initial conditions for the ocean, the coupling interface linking the two components and the stategy for ensemble generation. In the ECMWF coupled model there is no dynamic sea-ice model; the initial conditions are based on the observed sea-ice limit but thereafter the sea-ice evolves according to damped persistence.

S1 went operational in late 1997, S2 started running in August 2001, although S1 continued to be run until the Fujitsu computer was decommissioned in 2003. During the last few years, work has proceded with developing System 3 (S3). Major changes have taken place in the ocean analysis system, but not in the ocean model per se. Atmospheric model development is a continuous process and several versions of the atmospheric model have been tested. At this time the definitive decision on the atmospheric model has not been taken, but assuming that cycle 31r1 passes its e-suite tests for medium range forecasting, then that will be the version we will use in S3. Once this decision is made, S3 is ready for full operational running of the hindcast and real-time suites. It is likely that S3 will run in parallel with S2 for a few months before being deemed the operational version.

In section 2 a general description of the main components of S3 will be given together with some indications of how it differs from S2. Section 3 gives more insight into the details of the ocean initialization. In section 4 there is a history of the develpment of S3, including the testing of the different atmospheric cycles. Results from the prototype S3 (experiment erwq) will be presented and compared with S2 in section 5. Various aspects of the relative performance will be discussed. The prototype S3 gives superior forecasts of the important Nino

indices, and the climate is generally better simulated. However, it is not universally an improvement over S2, and weaknesses in S3 will also be discussed.

# 2 Description of System 3.

## 2.1 The Coupled Model of S3

The atmospheric model is cycle 31r1 which is the same cycle as is planned to be used in the interim atmospheric reanalysis. In S3, the horizontal truncation is ($T_L$)159, with a N80 reduced gaussian grid of approximately 125 km resolution; 62 vertical levels, extending up to approx. 5 hPa, are used. There are many differences between cycle 23r4 used in S2 and cycle 31r1. In section 4 some results from interim cycles will be discussed briefly, and the results presented in section 5 will use cycle 31r1 .

Although there have been no major changes in the ocean configuration used in S3 relative to S2, we include a brief description here for completeness. The S3-ocean has 29 levels in the vertical. Near the surface the level separation is 10m as for S2. The horizontal resolution in the meridional direction is 1 degree (1.4 degrees on an E-grid) and near the equator, is 0.3 degrees, as in S2. The vertical mixing is PGT (Peters, Gregg and Toole 1988), a Richardson number dependent parameterization. There are three forms of horizontal mixing: laplacian, biharmonic and shear dependent mixing. The model has a free surface to represent the sea level, which is affected by fresh water flux. The barotropic component is multi-stepped with a time step of 36 seconds compared with 1 hour for the baroclinic part. Details of the explicit free surface solver are given in Anderson and Balmaseda, 2005. In S3, the bathymetry files have been slightly modified around the Indonesian throughflow area, but without major impact (see section 4.6).

There is no dynamic sea-ice. The treatment of sea-ice has changed from S2 to S3, in order to guarantee consistency with the ice cover in the atmospheric initial conditions. In S2 the sea-ice is strongly relaxed to a damped-persistence anomaly, which is defined from the sea-ice cover in the ocean initial conditions and ERA-15 climatology, and such that after 60 days the sea-ice concentration is relaxed purely to ERA-15 climatology. In S3 the initial sea-ice cover is taken from the atmospheric initial conditions, the climatology of ice concentration comes from ERA-40 and the model for the time evolution is different: instead of persisting the anomaly, the total ice cover from the atmospheric initial conditions is persisted for 10 days, after which the persistence is damped such that the ice cover is relaxed to climatology after 30 days.

A coupler OASIS (version 2) is used to interpolate between oceanic and atmospheric grids at coupling times (once per day). In addition to SST, the ocean currents in S3 are visible to the atmospheric and wave models.

## 2.2 Initial conditions

The atmospheric initial conditions, including land conditions come from ERA40 for the period 1981 to 2002 and from Operations from 2003 onwards. The ocean initial conditions come from the ocean analysis described in section 3. As in S2, the ocean initial conditions in S3 are provided not from a single ocean analysis but from a 5-member ensemble of ocean analyses.

## 2.3 Ensemble Generation

The 41 member ensemble of real-time forecasts is generated from the ensemble of 5 ocean analyses (discussed later) by adding patterns of SST perturbations at the start of the coupled integrations. As in S2, in order to

have a 3 dimensional structure in the ocean, the SST perturbations are linearly interpolated to a depth of 40m. The perturbations are applied so as to generate no mean bias i.e for every ensemble member with a positive perturbation amplitude there is an ensemble member with the negative. This strategy was also used in S2 although the sampling strategy has been revised to ensure that ensemble sizes can always be increased cleanly.

As for S2, the differences between various SST analysis data sets are taken to reflect the inherent uncertainty in the SST analysis. The perturbations for S3 were re-evaluated based on more recent reanalyses of SST. As in S2, stochastic physics is active throughout the coupled forecast period. In S3, the atmospheric initial conditions are also perturbed using the Singular Vectors. These have an effect on the ensemble spread over the first month but little effect beyond that.

## 2.4 Structure and output of System 3

System 3 will archive its data into the multi-model seasonal forecast streams (MMSF etc). This will give consistency in the data archive between all members of the multi-model forecasting system. For users accessing data, the switch to the new streams should be straightforward - changing the stream names and adding origin=ecmf. (ECMWF is acting as a focus for the development of real-time multi-model seasonal forecast system, called EURO-SIP. Currently, the other participants in EURO-SIP are the Met Office and Meteo France, but other members are expected to join in the future.)

The period over which the back integrations are made will be 1981-2005 (compared to 1987-2001 for S2). This extended 25 year calibration period will give users more information on the skill of the system, and allow better estimation of forecast products calibrated using actual past performance. Although the mean model bias can be estimated reasonably with 15 years of data, estimates of reliability and skill of probability forecasts need as many past cases as possible. In fact, 25 years is still a long way short of what is needed, but is at least a step in the right direction. The period from 1981 to the present corresponds roughly to the period for which reasonably accurate SST and wind fields are available to initialize the coupled model forecasts. This choice of period for back integrations has been coordinated with the other members of EURO-SIP, such that all models will work towards providing this period.

The ensemble size of the S3 back integrations will be 11 members every month. This differs from S2, which has only 5 member ensembles each month, but augmented to 41 members for November and May starts. This change is cost-neutral, but allows better operational characterization of forecast performance. It is possible that for specific seasons the ensemble size might be augmented in research mode at a later date.

At the request of users, S3 seasonal integrations will be 7 months long (rather than the present 6 months). Additionally, once per quarter an 11 member ensemble will be run to 13 months, specifically designed to give an ENSO outlook. Tests with Cy29r3 gave an encouraging level of forecast performance for Nino SSTs over this forecast range, and the model version used for S3 is expected to have similar characteristics. Back integrations will also be made to this range, once per quarter, and with a 5 member ensemble.

The archived output of the model runs has been modestly upgraded. Previously, 200hPa was the highest archived pressure level, but now we have additional levels at 100, 50 and 10 hPa, which will allow a first look at the stratosphere in the seasonal forecasts. The following 2-d fields have also been added to the output list, mostly because of specific user requests:

31 sea-ice cover

33 snow density

78 total column liquid water

79 total column ice water

137 total column water vapour

183 soil temperature level 3

186 low cloud cover (6 hourly)

205 runoff

236 soil temperature level 4

243 forecast albedo.

In addition to the operational output from S3, a subset of the back integrations will have additional diagnostic output to support scientific analysis of the model. The fields will not be available as operational forecast products:

- Boundary layer height, CAPE, mid- and high- level cloud cover (6 hourly)

- TOA clear sky LW and SW fluxes (24 hourly; will allow computation of cloud forcing diagnostics)

-PV, relative humidity and vertical velocity on pressure levels.

Of particular note is the inclusion of model level output for a subset of five of the integrations (and for a limited forecast range of 5 months), to facilitate dynamical downscaling work by interested parties. Only every second model level is archived, plus level 1. Writing out large amounts of model level data slows down the model integration quadratically with the amount of data, and this has resulted in us limiting the frequency of model-level output to 12 hourly. Although not ideal, such data has been used for downscaling in DEMETER, and it was judged to be better than nothing for S3. ENSEMBLES are producing a set of integrations with full model level data at 6h intervals (with lower horizontal and vertical resolution, and only 22 start dates instead of 300). It is hoped that work with this full set of data will clarify the exact requirements in terms of space-time sampling for model level data, and allow us to find appropriate technical solutions to the performance issues for future systems. The total data archive for the S3 forecasts is expected to be about 20 Tb, with the model level output accounting for a little less than half of this.

The structure and output of S3 has been coordinated to the greatest extent possible with the requirements of the WCRP Task Force on Seasonal Prediction. It is also very much coordinated with the ENSEMBLES stream 2 integrations. It is intended that the operational S3 integrations will form the basis of the ECMWF ENSEMBLES contribution for stream 2. The integrations will be extended back to earlier start dates, but only for starts in Feb, May, August and November, as part of ENSEMBLES. It is hoped that these earlier dates will be archived as operational data, to form as seamless an archive as possible. Even though these earlier dates will not officially be part of the operational S3, they will be available to enhance knowledge and exploitation of the operational system.

# 3 The ocean analysis for S3

A new ocean analysis system has been implemented to provide initial conditions for S3 forecasts. The ocean analysis extends back to 1959 and provides initial conditions for both real-time seasonal forecasts and the calibrating hindcasts which are based on the period 1981-2005. (Although only the ocean analyses from 1981 are used for S3, the earlier ocean analyses will be used by the ENSEMBLES project, for analysing climate variability and as part of the CLIVAR Global Sysnthesis and Observation Panel (GSOP) assessment). A detailed

description and assessment of the S3 ocean analysis is given in Balmaseda et al. (2007)

When designing a data assimilation system for seasonal forecasts several considerations need to be taken into account. It is important to represent the interannual/decadal variability in the ocean initial conditions, and therefore strong relaxation to climatology is not recommended. On the other hand, in order to avoid spurious trends and signals due to the non-stationary nature of the observing system, the ocean analysis mean state should be close to the observed mean state. It is also important to avoid large initialization shocks in the coupled model, which may damage the forecast skill. In S3 we have tried to strike a balance between the above requirements: the relaxation to climatology is considerably weaker than in S2. This has been possible because an additive bias correction has been included (see section 3.1).

As for S2, the ocean data assimilation system for S3 is based on HOPE-OI, but major upgrades have been introduced. A pictorial view of the various data sets used in S3 is given in fig 1. In addition to subsurface temperature, the OI scheme now assimilates altimeter derived sea-level anomalies and salinity data. All the observations in the upper 2000m are assimilated (in S2 only the observations in the upper 400m were used). In S3, the observations come from the quality controlled data set prepared for the ENACT and ENSEMBLES projects until 2002 (Ingleby and Huddleston 2006), and from the GTS thereafter (ENACT/GTS). The OI scheme is now 3-dimensional, the analysis being performed at all levels simultaneously down to 2000m, whereas in S2, the analysis was carried out on each model level independently and only to 400m.

In S3 there is also a multivariate bias-correction algorithm consisting of a prescribed a-priori correction to temperature, salinity and pressure gradient, as well as a time-dependent bias term estimated on-line, which acts mainly on the pressure gradient (Balmaseda et al 2007). The on-line bias correction is adaptive and allows for flow-dependent errors. Because of the a-priori bias-correction term, the subsurface relaxation to climatology has been weakened: from a time scale of 18 months in S2 to 10-years in S3. Because of large uncertainties in the fresh water flux, the relaxation to climatology is stronger for surface salinity (approx 3-year time scale), but still weaker than in S2 (approx. 6 months). The analysis of SST is not produced using the OI-Scheme. Instead, the model SSTs are strongly relaxed to analyzed SST maps. The maps are daily interpolated values derived from the OIv2 SST product (Smith and Reynolds 1998, Reynolds et al 2002) from 1982 onwards. Prior to that date, we used the same SST product as in the ERA40 reanalyis.

The first-guess is obtained from integrating the HOPE ocean model from one analysis time to the next, forced by ERA40/OPS fluxes (ERA40 fluxes from the period January 1959 to June 2002 and NWP operational analysis thereafter). In S2 the fluxes were from ERA15/OPS, but the wind stresses were not directly used: instead, the wind stresses were derived from the analyzed winds using an off-line bulk formula. The representation of the upper ocean interannual variability is improved when using the ERA40 wind stress (Uppala *et al.* 2006), although the stresses are biased weak in the equatorial Pacific.

The fresh water flux from ERA-40 (Precipitation - Evaporation, denoted P-E) is known to be inaccurate. A better but by no means perfect estimate was obtained by 'correcting' the ERA-40 precipitation values (Kollberg and Troccoli 2003) as part of the EU ENACT project.

As in S2, the ocean initial conditions in S3 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. The wind perturbations have been revised to take into account improvements in the accuracy of determining wind stress and are smaller than those used in S2.

In the following section, we describe the different components of the ocean data assimilation system. Section 3.2 presents some brief assessment of S3, focusing on the comparison with S2 and climate signals. Section 3.2.3 investigates the impact of data in the ocean analysis.

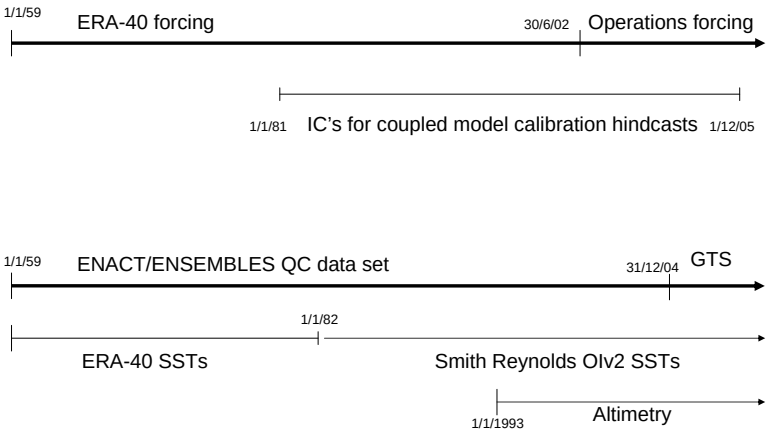



*Figure 1: Upper panel shows the surface forcing used in the ocean analysis and the initial conditions for the calibration hindcasts for S3. Lower panel shows the origin of the subsurface data surface temperature fields used.*

### 3.1 S3 Data Assimilation System

In S3, the different data streams are assimilated sequentially, as illustrated by the following scheme:

$$\begin{pmatrix} \tilde{T}^b \\ \tilde{S}^b \\ \tilde{\eta}^b \\ \tilde{\vec{u}}^b \end{pmatrix} \underset{bias}{\overset{remove}{\longrightarrow}} \begin{pmatrix} T^b \\ S^b \\ \eta^b \\ \vec{u}^b \end{pmatrix} \overset{alti}{\underset{\eta'^o}{\longrightarrow}} \begin{pmatrix} T_1^a \\ S_1^a \\ \eta_1^a \\ - \end{pmatrix} \overset{Temp}{\underset{T^o}{\longrightarrow}} \begin{pmatrix} T_2^a \\ S_2^a \\ - \\ - \end{pmatrix} \overset{Sal}{\underset{S^o}{\longrightarrow}} \begin{pmatrix} - \\ S_3^a \\ - \\ - \end{pmatrix} \overset{Geostr}{\underset{SL-trend}{\longrightarrow}} \begin{pmatrix} T^a \\ S^a \\ \eta^a \\ \vec{u}^a \end{pmatrix}$$

First of all, the model background $(\tilde{T}^b, \tilde{S}^b, \tilde{\eta}^b, \tilde{\vec{u}}^b)$ is bias-corrected according to the scheme described in section 3.1.1. Then, the detrended altimeter-derived sea level anomalies $(\eta'^o)$ are combined with the bias-corrected model first-guess $(T^b, S^b, \eta^b, \vec{u}^b)$ using the Cooper and Haines 1996 scheme (CH96 hereafter) to produce a first analysis $(T_1^a, S_1^a, \eta_1^b, -)$. This analysis is then used as a first guess for a second assimilation step, where only subsurface temperature data $T^o$ are assimilated, and salinity is updated by imposing conservation of the model temperature/salinity (T/S) relationship, while the sea level and velocity field remain unchanged. We refer to this second analysis as $(T_2^a, S_2^a, -, -)$. In a third assimilation step, the information provided by the salinity observations $S^o$ is used to modify the model T/S relationship. In this step, the T/S information is spread along isotherms following the scheme by Haines *et al.* 2006. Only salinity is modified in this step which results in the analysis $(-, S_3^a, -, -)$. After this 3rd assimilation step, velocity updates are derived from the temperature and salinity increments imposing geostrophic balance (Burgers et al 2002). Finally, the trend in global (area averaged) sea level is assimilated. By combining the altimeter-derived trend in global sea level with the model trend in global dynamic height, it is possible to make the partition between changes in the global volume and changes in the total mass. By doing so, the global fresh water budget is closed and the global surface salinity and sea level adjusted accordingly. Each of the steps is described briefly below.

The analysis is performed every 10 days. All the observations within a centered 10-days window are gathered and quality controlled. Analysis increments in temperature, salinity and velocity are calculated using the pro-

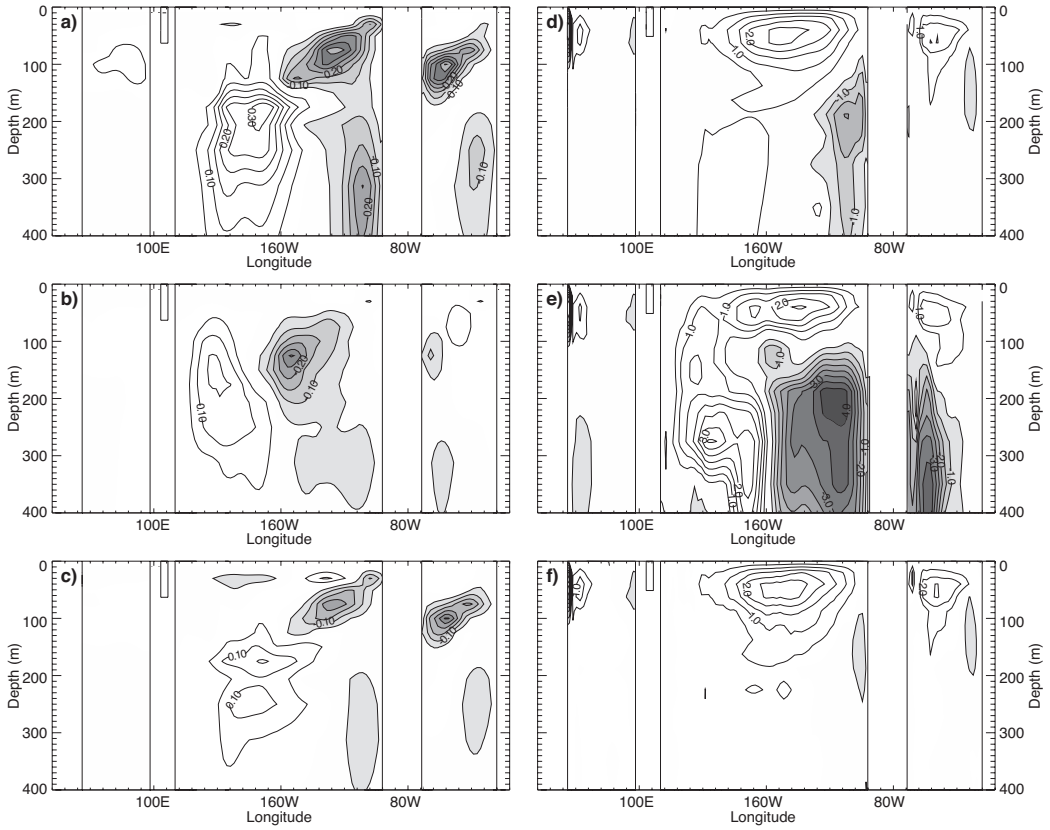



*Figure 2: Equatorial longitude-depth section of mean temperature increment (left column) and vertical velocity (right column) for an experiment without bias correction (a and d), for an experiment with bias correction directly on temperature (b and e) and for an experiment with bias correction on the pressure gradient (c and f). Contours every 0.5°C/10-days for the temperature increment and 0.5m/day for the vertical velocity. The zero countour is not plotted, and shading represents negative values. The mean corresponds to the time average during the period 1987-2001.*

cedure outlined above. To avoid exciting gravity waves, and to allow the model dynamics to adjust gradually to the changes in the density field, an Incremental Analysis Update method is used: the increment is added slowly over the subsequent 10 days, after which a new background field is available, and the cycle repeated.

### *3.1.1 Bias-correction algorithm*

The presence of bias in an ocean data assimilation scheme is a serious obstacle to the reliable representation of climate by historical ocean reanalysis. In the equatorial Pacific, the mean temperature assimilation increment in S2 is different from zero, and shows a large scale dipolar structure. Consistent with other assimilation systems, comparison with TAO currents shows that in S2 the equatorial zonal velocity in the Eastern Pacific is degraded when assimilating temperature data, even when salinity is also corrected by imposing preservation of the T-S relationship. The degradation of the zonal velocity is associated with a spurious vertical circulation underneath the thermocline, as pointed out by Bell et al 2005.

The standard procedure to deal with systematic error in a data assimilation system is to augment the model state with a set of systematic error or bias variables. The approach requires two analysis steps: one for the bias estimation and a second for the state vector. Assuming that the bias is nearly constant in time, and that the bias error covariance matrix is proportional to the forecast error covariance matrix, with the proportionality

constant small, the algorithm can be approximated so it only requires one analysis step, and thus the bias term can be updated at little extra cost ( Dee (2006)). However the requirement of proportionality between the bias and forecast covariance matrices is a limitation since the bias and the model state vector can have different control variables. An alternative approximation for the one-step bias correction algorithm for the general case was derived. The modifications included an explicit multivariate formulation which allows the control variables for the bias to be different to those for the state vector. In this context, the correction applied to the pressure gradient proposed in the Bell et al scheme can be considered as a particular choice of control variable.

Figure 2 illustrates the sensitivity of the results to the multivariate formulation of the bias covariances. It shows equatorial sections of the mean temperature increment (left column) and mean vertical velocity (right column). Results are from a standard data assimilation experiment without bias correction (upper row), an experiment where the bias is corrected directly in the temperature field (middle row) and an experiment where only the bias in the pressure gradient is corrected, following the Bell et al scheme.

The mean increment in fig 2a has the same large-scale dipolar structure in the equatorial Pacific as S2: the data assimilation is correcting the slope of the thermocline, making it deeper in the western Pacific and shallower in the eastern Pacific. Associated with the negative increment in the Eastern Pacific, is a spurious vertical circulation (fig 2d). In the experiments where the bias has been corrected online (fig 2b and c), the resulting mean increment is smaller. However, in the experiment where the bias has been applied directly in the temperature equation, the spurious vertical circulation in the Eastern Pacific is even worse (fig 2e). This degradation of the equatorial currents is consistent with that observed in experiments where the weight given to the observations is increased. In the experiment where the bias is treated by applying a correction to the pressure gradient (fig 2f), the spurious circulation does not appear.

Modifications have also been introduced in the equation for the time evolution of the bias. A simple parametric model for the time evolution of the bias was developed. The introduction of a memory term limits the influence in time of isolated or sporadic observations. A side effect is that the magnitude of the bias can be underestimated but to compensate for that, an additional constant bias term is also introduced. This term is not affected by the on-line estimation and has to be estimated *a priori*, preferably from independent information. The *a priori* term has the potential to provide a smoother analysis by preventing abrupt changes in the analysis associated with the introduction of new observing systems.

### *3.1.2 Assimilation of altimeter-derived sea level anomalies*

In S3, altimeter data are assimilated for the first time in the ECMWF operational ocean analysis. The altimeter information is given by maps of merged satellite products, provided by SSALTO/DUACS and distributed by AVISO, with support from CNES. Twice a week (on Wednesday and on Saturday mornings) $\frac{1}{3}^o \times \frac{1}{3}^o$ Maps of Sea Level Anomaly (MSLA) for a merged product combining all available satellites (Envisat, Jason, Topex/poseidon, ERS2, GFO) using optimal interpolation and accounting for Long Wavelength errors are produced (Le Traon *et al.*, 1998, Ducet *et al.*, 2000). Prior to assimilation, these maps are smoothed to remove small scale features not resolved by the model and then interpolated onto the model grid.

The anomaly maps ($\eta'^o$) distributed by AVISO are referred to a 7 year mean (1993-2000). To enable comparison with the background field ($\eta^b$), a reference mean sea level ($\bar{\eta}$) is required:

$$\delta\eta = (\eta'^o + \bar{\eta}) - \eta^b. \tag{1}$$

In S3, $\bar{\eta}$ is the 7-year mean sea level from an ocean analysis spanning the period 1993-2000. The possibility of using a reference $\bar{\eta}$ derived from the GRACE gravity mission has also been explored, but it was found

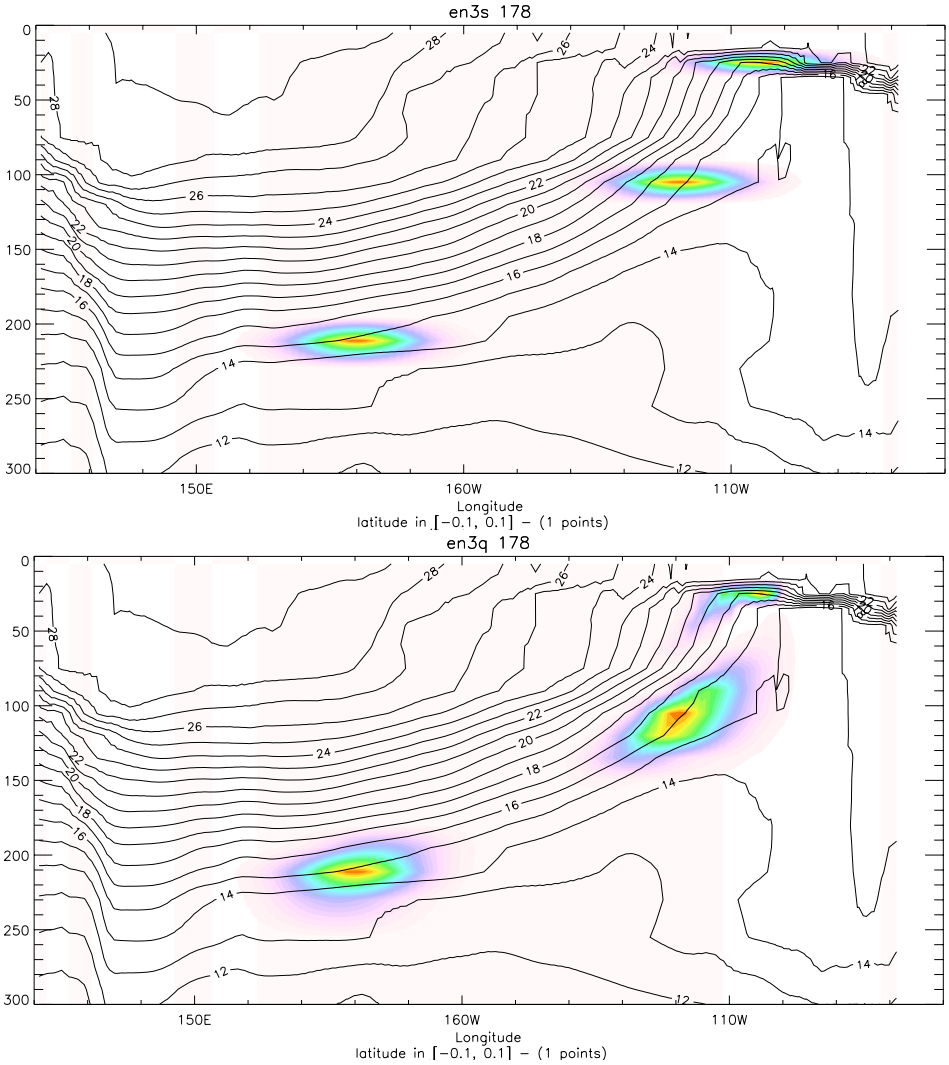



*Figure 3: Temperature increment coming from the assimilation of three observation at different locations in the equatorial pacific using the S2 (top) and S3 (bottom) covariances. Contours represent the background temperature field.*

that sensitivities to the reference mean sea level are large, and could potentially introduce abrupt jumps in the analysis.

The scheme used to project sea level information into the subsurface temperature and salinity consists of calculating the equivalent vertical displacement of the model water column to the difference in surface elevation $\delta\eta$ between background and observations, subject to the constraint that the bottom pressure is not changed.

### *3.1.3 OI assimilation of Temperature*

S3 uses an Optimal Interpolation (OI) scheme for the assimilation for subsurface temperature. The S3-OI scheme, derived from Smith *et al.* 1991, has evolved substantially: from the original univariate 2-dimensional OI scheme, where the analysis was performed on each model level separately, to a 3-dimensional scheme, where the analysis is performed at all levels simultaneously, with isopycnal formulation for the covariance matrices and a-posteriori multivariate updates of velocity and salinity. The OI interpolation is carried out on

overlapping sub-domains of the model horizontal grid (in order to reduce the cost of the matrix inversions). Where domains overlap, the analyses are blended together. The subdivisions 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 500 (for S2, the maximum number of observations was 200).

The horizontal background error correlations are represented by Gaussian functions which are anisotropic and inhomogeneous. Within 4 degrees of the equator the correlation length scale in the E/W direction is 1000 km while in the N/S direction it is 150 km. In the sub-tropics and high latitudes, polewards of 15 degrees, the correlation length scale is 300 km in all directions. Between the equatorial strip and the sub-tropics there is a smooth transition in correlation scales. In general, the observations are given half the weight of the background although the weight given to the data relative to the background field varies with depth according to a function that depends on the vertical gradients (see Balmaseda 2003). This function has been modified in S3 to preserve the vertical structure of the profiles. The net effect is that near the thermocline, the observations in S3 are given less weight than in S2 (about four times less). This reduction of the background weight near the thermocline improved the analysis of the Equatorial Atlantic.

In S3, the assimilation of subsurface temperature is performed with 3D domains down to 2000m. There are two novel aspects: the introduction of vertical scales and the introduction of a stratification dependent term which favours isopycnal spreading of information. Fig.3 shows the assimilation increment coming from 3 single observations of temperature along the equator using the covariances as in S2 (top) and S3 (bottom). Contours are background temperature isotherms. One can notice the two differences described above: the information is spread vertically in System 3[1] and the spread is not uniform and depends on stratification. At 200E/210m where the water-masses are well stratified the increment is almost an ellipsoid while at 120W/100m it has a more complicated shape.

When assimilating temperature data, the salinity increments are derived from the temperature increments by conserving the background water mass properties. The basic idea is to use locally the model background T/S relationship to reconstruct the salinity profile from temperature information only (Troccoli and Haines (1999), Troccoli *et al.*. (2002))). It has been shown that temperature assimilation can improve the salinity field of an ocean model by taking advantage of the large fraction of salinity variance that is strongly correlated with temperature variance. The correction on S is not applied in the mixed layer nor at higher latitudes where the hypothesis behind this scheme is less applicable. This is done by applying a latitude filter that reduces linearly to zero the effect of this scheme from $30^o$ to $60^o$.

### *3.1.4 Assimilation of salinity data on temperature surfaces*

Another new feature of S3 is the assimilation of salinity data: with the recent development of the ARGO network we now have an unprecedentedly good spatial coverage of salinity observations. Fig 4 shows the distribution of ARGO floats reporting salinity. The red triangles indicate TRITON moorings reporting salinity.

Getting the salinity field right is important in a number of contexts. Salinity has an impact on the density field and hence on ocean currents and transports, e.g. Cooper (1988), Roemmich *et al.*(1994), Vialard and Delecluse (1998a,b). Salinity is also important in certain places in the mixed layer where it controls the stability of the water column and hence to a degree, mixing and air-sea interaction, e.g. in the barrier layer around the western Equatorial Pacific. In addition the relationship between temperature and salinity contains important information about the nature of the thermocline and subduction rates and areas (Iselin, 1939). It is important therefore to recognize that the correct treatment of salinity data in the context of ocean data assimilation will allow analysed ocean fields to be used for more detailed studies of all of the above phenomena.

[1] It looks like there is a vertical spread in the top panel but it is an artifact of the plotting, it is confined to a model level.

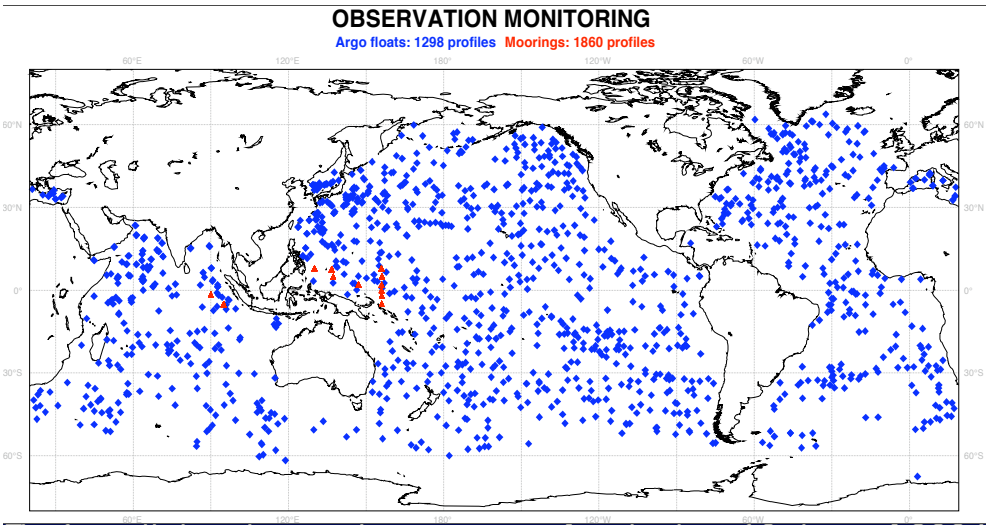



*Figure 4: salinity observation coverage for the last 10 days of 2004. Blue indicates ARGO floats. The moorings (red) measuring salinity are TRITON.*

The more conventional approach to assimilating salinity is to use covariance relationships formulated in $(x,y,z)$ coordinates. By doing the analysis in $(x,y,z)$ coordinates we are not taking advantage of the fact that the salinity increments in the Troccoli et al 2002 scheme leave the salinity unchanged on $T$ surfaces. Haines *et al.* 2006 proposed an assimilation scheme by which the temperature and salinity provide two separate pieces of information about the hydrographic structure of the ocean: the temperature information is used to correct the temperature and salinity field by preserving the T/S relationship, and the salinity information can be used to correct the model T/S relationship. They also propose a change of variable when assimilating salinity information: instead of using geographical coordinates, the salinity assimilation will be done in temperature space. We refer to this scheme as S(T) in what follows, and the conventional scheme on geographical coordinates will be referred to as S(z). Sensitivity experiments show that S(T) produces a better fit to the data than S(z), not only in salinity but also in temperature. However, the assumption that a large fraction of salinity variance is strongly correlated with temperature variance is not valid everywhere, especially at high latitudes. For that reason, using the same latitude filter as the salinity adjustment, the assimilation switches gradually between $30^o$ and $60^o$ from S(T) to S(z). Additionally S(z) is not used in the mixed layer where the search for the right temperature class may fail.

### *3.1.5 Assimilation of Global Mean Sea Level*

Figure 5 shows the time series of the global sea level from the altimeter data, for the period 1993-2006. The trend in global sea level dominates the variability. If this trend is produced by thermal expansion due to global warming, it is not represented by the ocean model sea level, as our model, in common with most ocean models used for climate activities employs the Bousinesq approximation, which means that it preserves volume. Therefore, if not treated correctly, the trend in sea level can be a problem when assimilating altimeter observations. In S3, the global sea level trend is removed from the altimeter sea level anomalies before they are assimilated via the CH96 scheme (section 3.1.2).

On the other hand there is an open debate about the attribution of sea level trend: how much is due to thermal expansion (steric) and how much is due to mass change over the ocean (isostatic) (Church and White 2006). In principle, ocean reanalysis can help to answer this question, since the idea is to use all possible information: by

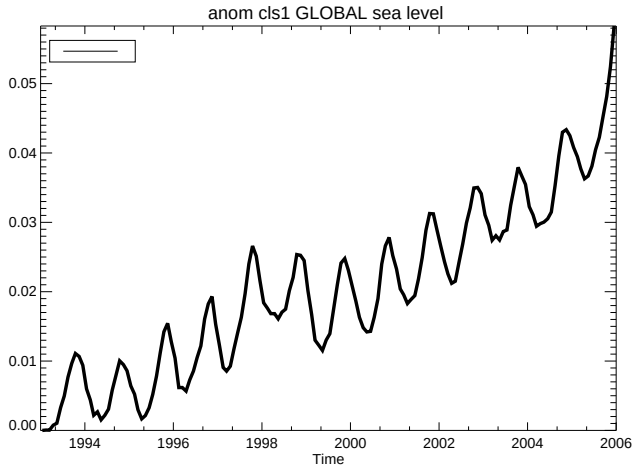



*Figure 5: Time series of the global sea level from altimeter data, relative to 1st Jan 1993. Units are metres.*

combining model first guess with subsurface data it should be possible to reduce the error in the estimation of the steric height. By comparing the total change in sea level given by the altimeter data ($\Delta\bar{\eta}$) with the change in steric height given by the ocean analysis($\Delta\bar{\eta}_s$), it is possible to estimate the component of the change due to mass variations, according to the equation:

$$\Delta\bar{\eta}_m = \Delta\bar{\eta} - \Delta\bar{\eta}_s \tag{2}$$

The residual ($\Delta\bar{\eta}_m$) is applied as a fresh water flux uniform in space. The partition between volume and mass changes is quite valuable information, since it can help to close the fresh water budget over the oceans, which is currently a problem for ocean analysis.

## 3.2 Evaluation of S3 ocean analysis

### *3.2.1 Comparison with S2*

Figure 6 shows the vertical profiles of the mean difference between the analysis and observations in the Eastern Pacific (area Niño 3) and in the Western Pacific (area EQ3). Positive/negative differences are indicative of a warm/cold bias. The bias in the both the Eastern and Western Pacific in S2 has been significantly reduced in S3, where the east-west slope of the thermocline is better represented. This is an indication that the bias correction algorithm is being effective.

A more stringent test of the quality of the analysis is given by the root mean square (RMS) of the difference between analysis and observations. Figure 7 shows the RMS for temperature in different areas of the tropical oceans. The RMS scores for S3 are systematically better than for S2. The same results hold for salinity, shown in figure 7. The assimilation of salinity data is especially beneficial in the Western Pacific (region EQ3).

The velocity measurements from the TAO moorings provide an independent data set for the validation of ocean analyses, since the currents have not been assimilated in either S2 or S3. Figure 8 shows the vertical profiles of the time correlation of the currents in S2 (blue) and S3 (red) with the TAO currents at different mooring locations. The better correlation shown by S3 indicates that not only is the density field better constrained by observations in S3, but it is also more dynamically consistent.

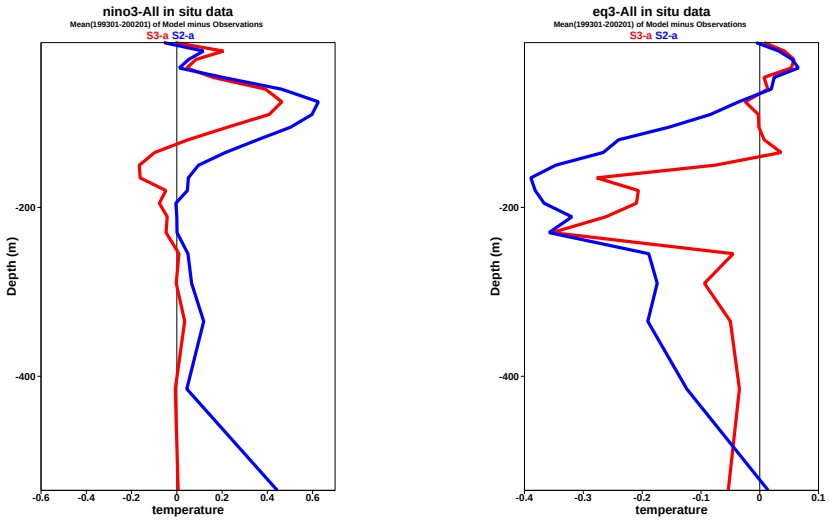



*Figure 6: Mean differences (analysis minus observations) in Nino3 (left) and in EQ3 (right). S2 is plotted in blue and S3 in red. Units are K.*

### *3.2.2 Climate variability*

Figure 9 shows the time evolution of the 5 ensemble members over the North Atlantic (30N-60N): the temperature and salinity anomalies averaged over the upper 300m are shown in the left and right panels respectively. The spread of the ensemble can be taken as a measure of the uncertainty, which may be underestimated, since only wind error is accounted for, and there is no representation of errors in the heat or fresh water fluxes, neither in the model nor in the analysis.

The variability in the upper ocean temperature is dominated by the upward trend, starting around the mid 80's. The presence of this trend should be taken into account when considering the reference climatology for the seasonal forecasts products. Salinity variations occur mainly on decadal time scales, and they seem to be correlated with variations in the thermohaline circulation (THC). Figure 10 shows the time evolution of meridional transport in the North Atlantic at 30N in the upper 1000m (upper panel), and at 3000m-5000m depth range. For comparison, the values given by Bryden et al 2005 are shown by the stars. Although there is broad agreement with the estimates by Bryden et al, the S3 ocean analyses show that the decadal variability is large, and therefore sampling is an issue when drawing conclusions about the slowing down of the THC. The S3 ocean analysis will provide initial conditions for the decadal ENSEMBLES forecasts, where the capability of the coupled models to reproduce and predict changes in the THC will be explored.

Figure 11 shows the time evolution of the global sea level from altimeter data (black line) and from the S3 analysis. The similarity between the two curves is not surprising, since the global mean sea level trend has been assimilated. The figure also shows the evolution of the global steric height (light blue) and the time evolution of the global bottom pressure (dark blue), indicative of changes in the global mass. The figure shows that till 2002 the global trend in mean sea level is mainly due to volume change probably associated with thermal expansion. The change in the period 1997-1999 is probably related to the large El Niño event. In mid 2002 and mid 2004 there are dramatic changes in the mass field. Experiments are underway to determine if this is due to changes

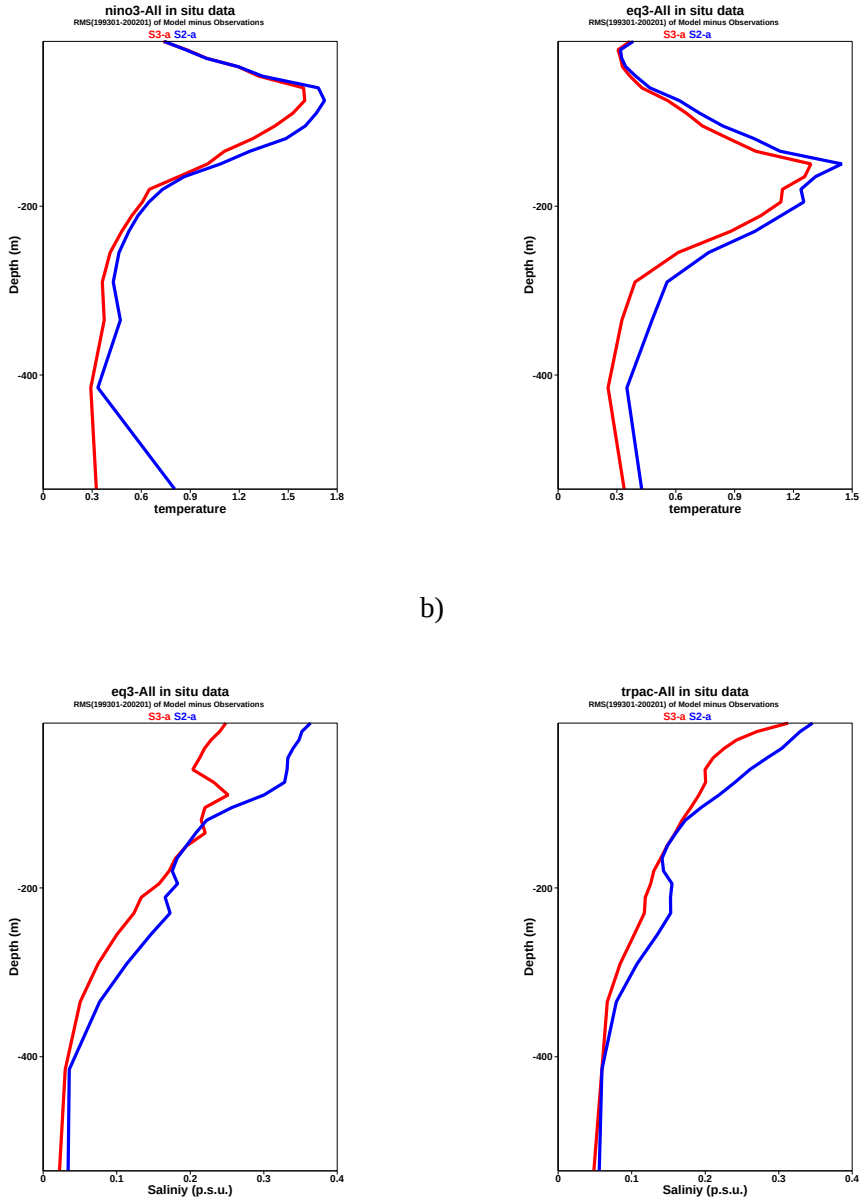



*Figure 7: RMS difference between temperature and salinity analysis and observations in different regions:S2 (blue) and S3 (red). Units are a) K, b) K, c) psu, d) psu.*

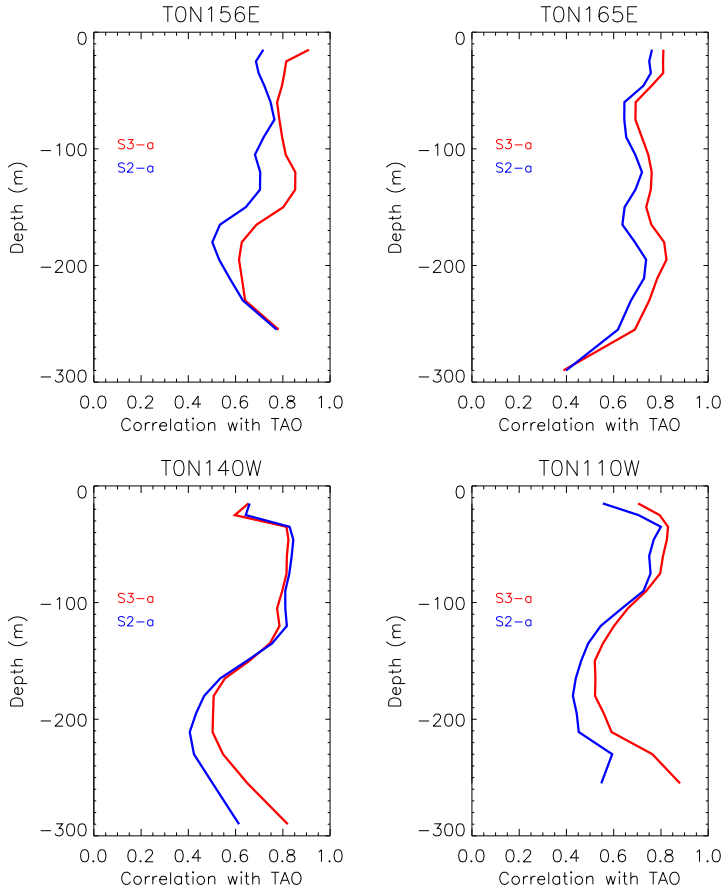



*Figure 8: Correlation with tao currents:S2 (blue) and S3 (red)*

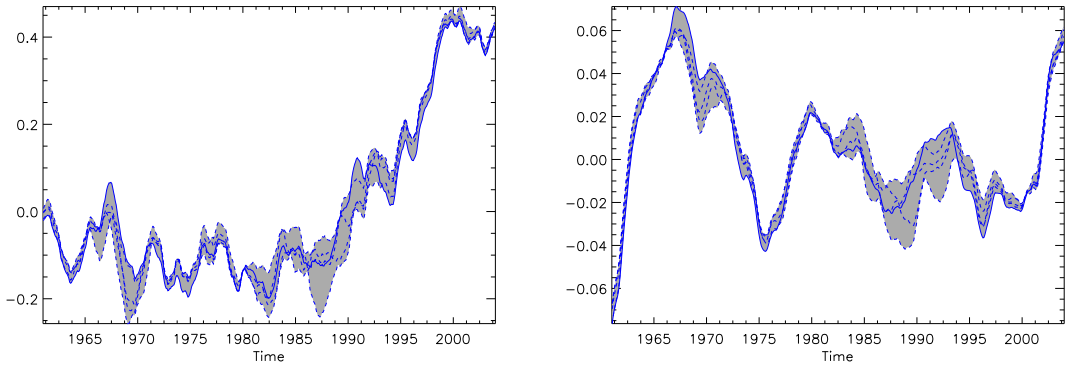



*Figure 9: Time series of temperature (left) and salinity (right) anomalies averaged over the upper 300m in the North Atlantic (30N-60N) from the S3 5-member ensemble of ocean analyses. Both curves have been smoothed with a 12-month running mean. Units are K for temperature and psu for salinity.*

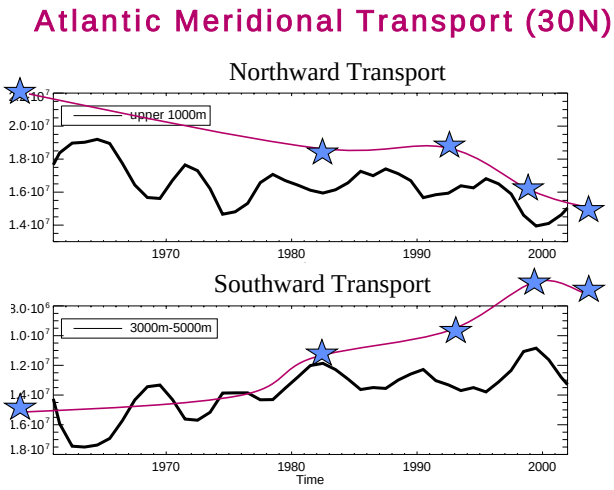



*Figure 10: Meridional transport at 30N in the North Atlantic in the upper 1000m (upper figure) and at depths between 3000m -5000m (lower figure) from S3 ocean analysis. The stars show the values given by Bryden et al 2005. Units are* $m^3/s$. *The model values have been smoothed with a 2 year running mean. The error bars in the values by Bryden et al are* $+/-6x10^6 m^3/s$.

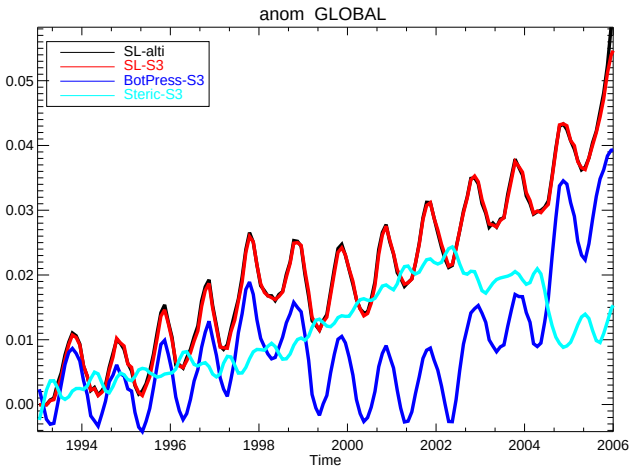



*Figure 11: Global trends in sea level, in S3 steric height (light blue), representing changes in volume, and equivalent bottom pressure (dark blue), representing changes in mass. Data from altimeter data in black, and from S3 in red but these curves are virtually indistinguishable.*

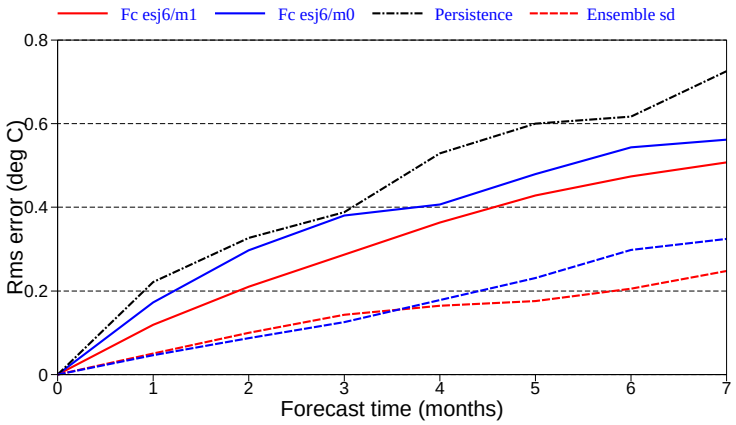



*Figure 12: Plot of the rms error in the Nino4 region as a function of forecast lead time. Two experiments are shown: forecasts from the S3 analyses (red) and forecasts from a control (blue) in which subsurface data have been withheld. For reference, the RMS error of persistence is also plotted (black). The forecast spread is shown dashed for the two experiments.*

in the observing system (increased coverage of ARGO floats), or changes in the altimeter product due to the inclusion of JASON.

### *3.2.3 Further comments*

Data assimilation has a significant impact on the mean state and variability of the upper ocean heat content. In the Equatorial Pacific, it steepens the thermocline and increases the amplitude of the interannual variability. In the Indian ocean it sharpens the thermocline, making it shallower, and it increases both the ENSO-related and Indian Dipole variability. In the Equatorial Atlantic the cold phase of the seasonal cycle is more pronounced, the amplitude of the interannual variability increases and the phase can be significantly modified in several cases. The interannual variability in both northern and southern tropical Atlantic regions is increased, especially in the later part of the record (after 2000). Comparison with the OSCAR currents (Ocean Surface Current Analysis, Bonjean and Lagerloef, 2002) shows that data assimilation improves the representation of surface currents in the system, mainly in the tropical band (not shown).

Fig 12 shows the impact of the data assimilation system on the forecast error. The red curves indicates the growth of error (solid) and the ensemble spread (dashed) for a research experiment very close to the proposed configuration of S3. This experiment is very recent and uses the S3 ocean analysis as initial conditions and Cycle 31r1 for the atmosphere. The blue curve is for an experiment which only assimilates SST information. The reduction of error using the S3 analyses is significant, especially in the west equatorial Pacific as this figure shows. The reduction of error in the tropical Atlantic is disappointingly small though there is some improvement in forecasts for the sub tropical Atlantic (not shown).

Observing system experiments have been carried out to determine the information provided by the recently developed ARGO network, how it interacts with the information given by the altimeter and the impact of the observing sytem on climate variability. Results shown in figure 13 indicate that the observing systems for the ocean are not yet redundant. Even in the presence of ARGO and TAO data, the altimeter information can improve the vertical temperature structure in the far Eastern Pacific (fig 13a). In the Indian and Atlantic oceans the contribution of altimeter and ARGO is similar, and the best results are obtained when both observing

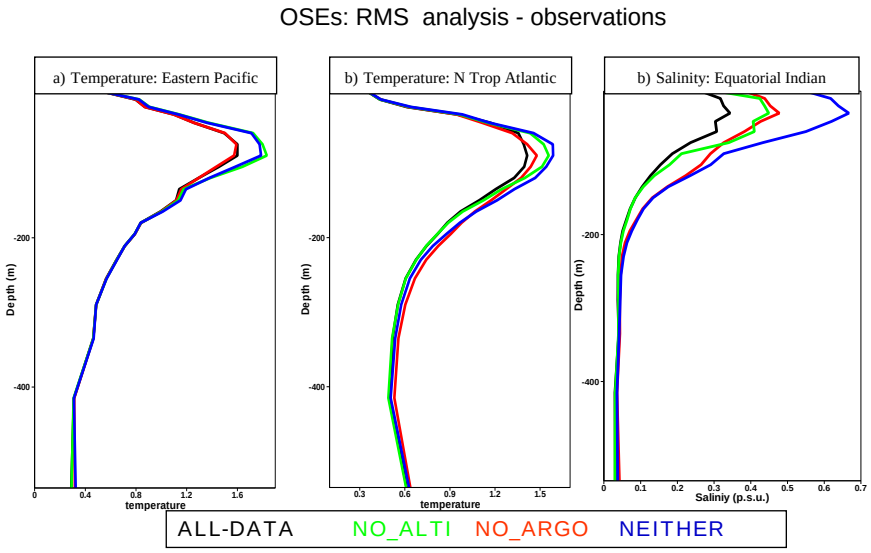



*Figure 13: Results from Observing System Experiments (OSEs). Shown are the RMS of the analysis minus observations during the period 2001-2005. The assimilation of altimeter data reduces the RMS differences of the temperature field in the Eastern Pacific (a). Both ARGO and altimeter are needed to obtain the minimum RMS difference in temperature in the North Tropical Atlantic (b) and in salinity in the Indian Ocean (c).*

systems are included (13b,c). ARGO has a large impact on the salinity field on a global scale. This can be seen in fig 14 which shows the difference in surface salinity between an analysis in which all data are presented and one in which ARGO data (both T and S) are witheld. The differences are particularly large in the vicinity of the Gulf Stream and Kuroshio, in the tropical Atlantic and Indian oceans and in the southern hemisphere. They are quite modest in the tropical and sub tropical Pacific. This figure is an average over 5 years. During that time the ARGO network has expanded considerably. The overwhelming effect of ARGO is to make the surface ocean saltier. Almost nowhere is the reverse true.

# 4 Model versions between System 2 and System 3

System 2 started operational running in August 2001, using Cycle 23r4. Since then, there have been many developments in the IFS. New cycles with significant changes have been tested in a standard seasonal forecasting configuration. Further tests have been carried out as required, to investigate specific issues such as resolution and timestep. The primary metric for these tests has been the impact on SST forecast skill in the tropical Pacific. A comprehensive account is beyond the scope of this report, but certain highlights and key cycles will be mentioned.

## 4.1 Cycle 26r3

Cycle 26r3 gave a significant degradation in the SST forecasts for the west Pacific. Investigation showed that the model had a strong timestep dependence, and that performance could be largely restored by using a 30

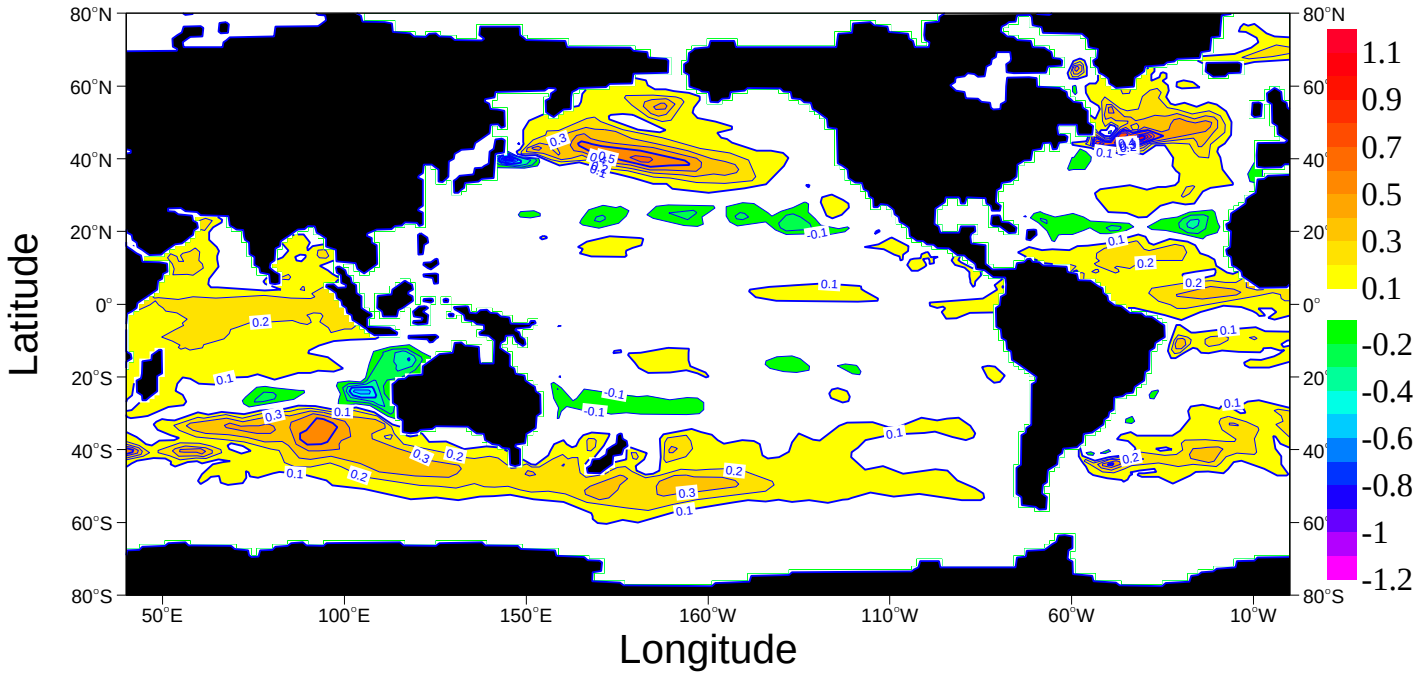



*Figure 14: Plot of the impact of ARGO on surface salinity averaged over the 5 years 2001-2005.*

minute timestep. In the several cycles that followed, further tidying of the numerics systematically reduced the timestep dependence (ie the difference in skill between forecasts using 30 minute and 60 minute timesteps). By 29r1 no significant difference in skill remained (either at T95 or T159 resolution), and beyond this cycle the impact of smaller timesteps was no longer tested. System 3 will use a 1 hour timestep. The ability to do this was an important factor in allowing us to use a T159 resolution.

## 4.2 Cycle 29r1, horizontal resolution and ocean mixing

The physics changes in 29r1 meant that, at the then-standard resolution of T95, it was on aggregate the best performing cycle to date. Various coupled and uncoupled experiments were made with this cycle to document its performance. Changes were also made to improve the pseudo ice model contained within the ocean, and associated changes in the coupling interface, so that the initial position of the ice edge could be directly set based on the initial conditions given by the atmospheric analysis system. The ocean mixing parameters were also retuned. Several tests were made of the impact of higher resolution T159. The forecast results appeared to depend somewhat on a parameter controlling ocean mixing (although sampling effects cannot be completely ruled out), but in all cases the higher resolution gave improved forecasts, and in many cases the improvements were substantial. The higher atmospheric resolution had a systematic effect on the atmospheric model climate in all runs (coupled and uncoupled), giving higher tropospheric temperatures at mid and high latitudes. This generally positive impact is believed to be due to increased eddy heat transport by the atmosphere model.

Experimentation with 29r1 also showed some sensitivity of skill scores to ocean mixing, with larger horizontal mixing giving somewhat better results in the Central-Western Pacific. The higher mixing values were used for all subsequent cycles. This means that the ocean mixing parameters are different in the coupled model and the analysis.

When using an atmosphere resolution of T159, the wave model resolution was increased to 1.5 deg. Note that for S3 we intend to keep the number of angular directions at 12. Compared to an enhanced angular resolution of 24, this does not give quite as good swell forecasts related to propagation from individual synoptic systems, but in terms of seasonal forecasts this level of synoptic detail is not thought to be sufficiently relevant to justify the additional cost.

### 4.3 Cycle 29r3 and vertical resolution

With 40 levels, using the best settings from the 29r1 experiments, and for the first time using prototype S3 ocean analyses, T159 experiments with 29r3 showed modest further improvement over 29r1. This cycle introduced the 62 level version of the IFS (enhanced vertical resolution everywhere, but still no proper representation of the full stratosphere). Tests with 62 levels gave a slightly better mean state (slightly cooler temperatures), but showed little impact on forecast scores. Nonetheless the 62 level version became the standard from this point onwards.

### 4.4 Cycle 30r1 and surface currents

Tests showed 30r1 to be essentially indistinguishable from 29r3, as expected. However, tests made with 30r1 demonstrated that the coupling of ocean surface currents to the atmosphere boundary layer was now acceptable, in that it had very little impact on the forecast scores. (Earlier testing with 29r1 had shown a significant negative impact - the difference this time is believed to be due to some subtle shifts in the interaction of mean state and forecast errors in the model, and the way in which the surface current coupling interacted with this interaction). From this point on, the surface current coupling became part of the standard model configuration. It should be mentioned that analysis of these experiments (particularly changes in surface wind and surface stress, and comparison of these fields with ERA-40) highlighted the fact that atmosphere analysis systems which assume a stationary ocean surface cannot produce simultaneously correct estimates of 10 m wind and surface stress. This is especially important given the sensitivity of ocean models to specified surface momentum forcing - a major application of reanalysis products. In order to produce good quality analyses, and interpret scatterometry and in-situ wind measurements correctly, the atmosphere analysis system needs to account for the ocean currents. Because ocean surface currents react quite rapidly to synoptic systems, this may best be done by including the ocean surface current in the outer loop of 4d-var, rather than, for example, specifying it as a fixed constant field. This is an area which will be worked on in the next two years, with the aim of having a workable system by the time of the next major reanalysis.

### 4.5 Cycles 30r2, 31r1 and the final choice of the System 3 version

It has taken some time to finalize the choice of IFS cycle to be used for S3. In the summer of 2005, the intention was to use Cy29r3, together with its associated increase in vertical resolution (to 62 levels for the reduced stratosphere version). This cycle performed relatively well in seasonal forecast tests, but failed its esuite in the medium-range forecasting system. It was decided to await the fixed version of 29r3, which eventually became available as Cy30r1. By the time it was available for us to use, an improved physics version (Cy30r2) was also available for testing. Because Cy30r2 included important physics changes to improve inter alia the representation of the MJO, it was agreed to test 30r1 and 30r2 in parallel. The results showed that 30r1 gave a creditable seasonal forecast performance, while 30r2 had both substantial advantages (in MJO activity) and substantial disadvantages, particularly in forecasting important west Pacific SST anomalies. Cycle 30r2 e-suite also developed an unrealistic upper troposphere moisture distribution. Investigation revealed that this was

precisely due to the set of modifications to the cloud scheme that also improved the MJO statistics. It was decided to build a new cycle (31r1) incorporating a further set of physics changes. Tests of this model version suggest that it is the best cycle yet in terms of SST anomaly forecasts, and that the model climate is also good. Given its overall superior performance, it has been decided that S3 will use this model version. Although we have had to wait some time for a satisfactory IFS version, the delay appears to have been rewarded by a good model version for our purposes. In much of what follows, results are based on 31r1 (actually a branch of 30r2 containing the physics changes of 31r1). Only a subset of hindcasts have been performed in the form of an rd experiment erwq: the full set of hindcasts on which an operational implementation of S3 will be based, have yet to be made.

A summary statistic of SST forecast performance, based on errors in Nino3, Nino3.4 and Nino4, shows the pace of progress in the years between S2 and the S3 prototype. (Table 1).

| Forecast System | Date of cycle | Figure of Merit |
|---|---|---|
| S1 | 1997-01 | 1.138 |
| S2 | 2001-06 | 1.157 |
| Cy25r4 | 2003-01 | 1.114 |
| Cy26r3 | 2003-10 | 1.159 |
| Cy28r3 | 2004-09 | 1.133 |
| Cy29r1 | 2005-04 | 1.084 |
| Cy30r1 | 2006-02 | 1.009 |
| erwq | 2006-08 | 0.961 |

*Table 1: Figure of merit (sum of mean absolute errors in Nino 3, Nino 3.4 and Nino 4, based on 5 member ensemble means, averaged over forecasts from 1987-2002 and over forecast lead times from 1 to 6 months) for several different seasonal forecast systems.*

In terms of this summary statistic, there was very little progress in Nino SST forecast skill in the years from 1997-2004. This period was characterized by ups and downs in the scores, by the diminution of some problems but the appearance of others, by a real sensitivity to the atmosphere model version, but difficulty in producing all-round, sustained improvement. At the end of this eight year period, there is no doubt that the IFS was a more realistic model than it had been at the start, but the improvements were somehow not enough to translate reliably into improved Nino SST scores. From Cy29r1 onwards, however, the forecast errors have been noticeably reduced compared to their earlier values. The change from 29r1 to 30r1 in the table above includes a resolution change from TL95 to TL159 - this on its own explains perhaps up to half of the improvement, but probably not more. And the latest cycle seems to give yet more significant improvements.

We should be cautious about over-interpreting this apparent change in the trend rate for model improvement, but it is certainly encouraging. The model is closer to being realistic across a whole range of physical processes than it has ever been before, and it may just be that we are starting to benefit from the strategy of real model improvement rather than an emphasis on model tuning. There is certainly plenty of scope for improving the realism of the model, particularly in terms of the tropical low-level wind field and its variability. The next few years will show whether or not we are able to sustain the recent trend of continued improvement in El Nino forecast scores.

### 4.6 Other investigations in preparation for System 3

Several other issues have been investigated in the process of deciding the configuration for S3. One of these is the specification of time-varying greenhouse gases and other forcings. Investigations showed that, especially for

forecasts made at widely separated starting dates, the correct specification of CO2 and other greenhouse gases is important. The effect is seen most strongly in large area averages such as global mean temperature. The effect becomes quite noticeable after a few months of integration - simply specifying the right global temperature in the initial conditions is not sufficient for a seasonal forecasting system. Tables of observed greenhouse gas concentrations (CO2, Methane, N2O, CFC11 and CFC12) are used up to the year 2000, and values from the A1B scenario for later dates. Solar variations (which are weak) are also accounted for.

An interesting question was how to treat volcanic aerosol. Historical data are available in terms of zonal mean optical thicknesses as a function of latitude and time. However, data for the most recent years are not readily available, and as yet we have no operational capabilities to specify volcanic aerosol loadings for the real-time forecasts. It was thus decided that time variation of volcanic aerosols should be switched off for S3. The consequences of this are readily visible in the form of small but systematic overestimations of global mean temperature, e.g. in the two years following the eruption of Mount Pinatubo in 1991. Although we have no present intention to include volcanic aerosol in S3, if there were a very large eruption (much bigger than Pinatubo, for example), then we would work fast to insert a reasonable estimate of the aerosol loading into the forecasts. In the future, it is hoped that a more operational approach to volcanic aerosols can be implemented.

All the experiments from 26r1 to 29r1 showed a marked warm bias in the Eastern Pacific. We conducted several sensitivity experiments to explore to what extent the bias could be reduced by improving the ocean component. A typical error of the ocean model when forced by analyzed fluxes is a heat deficit in the upper Equatorial Pacific, where the thermocline is too shallow. One possibility is that too much heat is exported poleward. The sensitivity of the equatorial heat storage to the horizontal/vertical mixing of heat and momentum was explored: lower values of the horizontal mixing of heat and higher vertical momentum mixing (mixed-layer) would increase the equatorial heat storage in the upper ocean, but at the expense of slowing the equatorial undercurrent. The role of the Indonesian throughflow in the equatorial heat storage was also investigated: by closing completely the Indonesian throughflow the Pacific thermocline became deeper. Therefore the ocean bathimetry was revised slightly around the Indonesian area to reduce the heat transport of Indonesian Throughflow.

### *4.6.1 Artificial adjustment of the mean state*

Investigation of the reasons for the variation in performance between different cycles highlighted the importance of the mean state of the coupled system as the base point for non-linear responses to interannual variations. This was particularly true regarding the mean extent of the cold tongue into the west Pacific warm pool. Given the sensitivity of forecasts to the model mean state, and given that even small adjustments in the atmosphere model physics could produce appreciable changes in the mean state, a natural question is as to whether better and more consistent forecast performance might be obtained by biasing the model mean state to be closer to reality, using artificial means such as flux adjustments. The idea is not necessarily to constrain the model mean state to be (apparently) close to observations by employing a 'correction', as this could easily distort the physics of the system to be worse than in the original coupled model. Rather, the emphasis is to perturb the mean state in such a way as to put the coupled model in a regime closer to reality, while not seeking a particularly close match for any particular variable.

Experiments were carried out with 28r3 and 29r1, perturbing the coupled model wind stress by adding specified offsets to it. In the case of the 28r3 experiments, the offsets were simple analytic functions with amplitude up to -0.02 N/m2. Significant improvements in forecast skill were eventually obtained, particularly in the central and west Pacific. The Nino 3 forecasts in the east were improved only slightly. It took several experiments to find an analytic form of stress correction which gave a strong improvement of skill.

Experiments with 29r1 took a slightly different approach of diagnosing a wind stress adjustment from uncou-

pled runs of the IFS using prescribed SST. By taking the difference between the uncoupled model climate and ERA40, it was hoped to make some first order stress adjustments which would mean that the ocean was operating in a more realistic regime (at least in the early part of the forecast), and would avoid the need for extensive coupled experimentation to tune the stress adjustment. In the case of 29r1, the uncoupled runs already existed. For 29r1, all the work was done at T159 resolution. The results again showed an improvement in forecast skill, and again the main improvement was in the west Pacific. This time, however, the scale of the improvement was quite a bit less than in the 28r3 experiments. Perhaps this was because the 29r1 model was improved compared to 28r3; perhaps because to get the coupled model into the state where it most accurately handles anomalies requires mean state adjustments considerably beyond those that are estimated from the observed mean state; or perhaps because the ERA-40 stresses are not sufficiently strong. (Evidence for this latter point comes both from biases in the ocean assimilation system, and first results from the interim reanalysis, which show significantly stronger equatorial wind stresses over the Pacific in the first year of assimilation compared to ERA-40).

In a different set of experiments, a bias correction algorithm, based on that developed for the ocean data assimilation, was used to estimate the bias in the ocean during a set of climatologically constrained coupled runs using cycle 28r3. The resulting 3D bias fields were then used additively to correct the ocean horizontal pressure gradient, temperature and salinity during a further limited set of coupled integrations. The methodology was successful in eliminating the biases in surface temperature, and the amplitude and patterns of interannual variability were improved. Results also showed that the bias correction estimated for the coupled model was quite different from the correction estimated for ERA40-forced ocean-only runs. The latter, when applied additively to the coupled runs, had only a small effect on the SST bias, suggesting that most of the SST bias in the coupled integrations had its origins in the atmospheric fluxes. Similar conclusions were reached by conducting coupled and ocean-only experiments from the same set of initial conditions.

Since the forecast model has improved substantially, and since the results of these preliminary experiments indicated that relatively good forecasts were hard to improve, at least in the east and central Pacific, further experiments were not made with the most recent cycles. It does seem that if a model version has a badly distorted model climate in the west-central Pacific, wind stress corrections can alleviate the problem and improve the west Pacific forecasts. Hopefully, future model versions will be good enough for this approach not to be necessary. There is still the possibility that further experimentation would enable significant improvements in skill even with the present model - the equatorial Atlantic is perhaps a case in point. Further work in this area is not ruled out, but is not a high priority at the moment.

# 5 Preliminary Results from System 3

## 5.1 Skill assessments for S3 - El Nino and SST forecasts

The expected performance of S3 is shown based on the rd experiment erwq, which uses the 30r2 physics branch used as input for 31r1, and consists of a 5 member ensemble run 4 times a year for 1987-early 2005. The ocean initial conditions are from a prototype of the S3 ocean analysis, not the definitive version. A final rd test using the definitive ocean analysis and atmosphere cycle 31r1 has just started at the time of writing. Differences between erwq (discussed here), the final rd test and the actual operational system are expected to be small. Note that our standard test period (and the period for which we have data from S1) is 1987-2002. The most critical comparisons of skill between erwq and S2 use all available data, i.e. 1987-2005.

Drift in the SST of the coupled model is generally reduced compared to earlier cycles, with an absence of tropical cooling (Fig. 15) and an improved seasonal cycle amplitude in the east Pacific (Fig 16). The interface between the cold tongue and the warm pool is reasonably represented in terms of SST, as exemplified by the

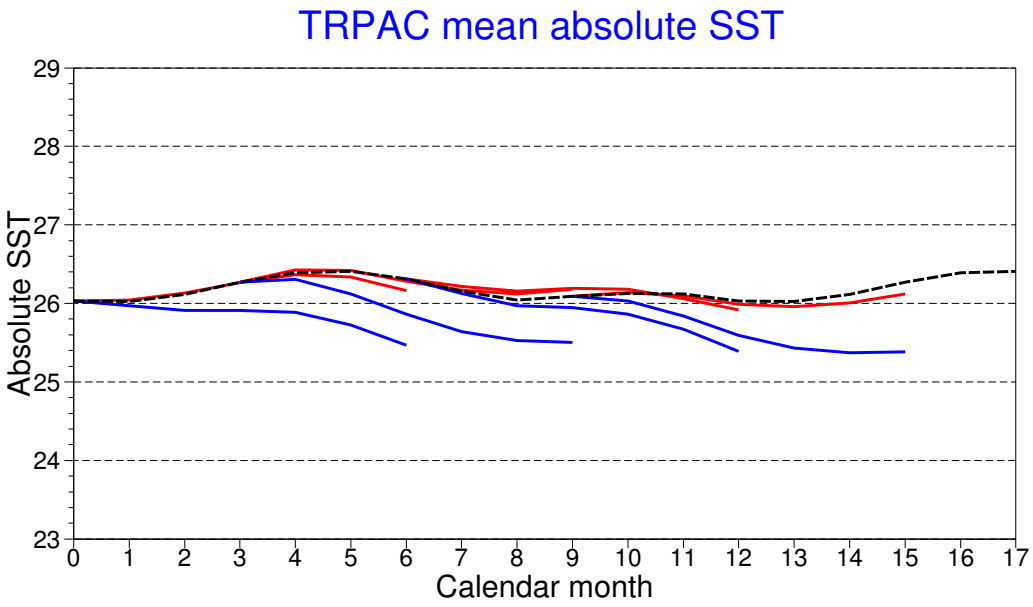



*Figure 15: Mean forecast SST averaged over the whole tropical Pacific (30N-30S), for S2 (blue) and erwq (red). The new model has very little overall drift in 6 month forecasts.*

drift in Nino 4 (Fig 17). In S2, for example, SST was too cold because the cold tongue extended too far to the west.

Figure 18 shows the rms errors in forecasts of Nino3.4 and Nino4 in erwq compared to the earlier S1 and S2. The progress is evident, both in these plots and when looking at anomaly correlation (not shown). Scatter diagrams, showing all available forecasts for which S2 and erwq can be compared, show the improvements in erwq are significant in all areas of the tropical Pacific (Figure 19 -only four regions shown). However, the strong improvement does not extend to all parts of the globe - outside the equatorial Pacific, changes in skill are largely neutral, and while clearly positive in the north subtropical Atlantic, may be marginally negative in the equatorial Atlantic. Forecast skill in the tropical Atlantic was studied at some length in a paper about to appear in J. Climate (Stockdale et al, 2006). Despite the enhancements in erwq (a more careful ocean analysis, and a slight reduction in the most problematic wind biases), predicting equatorial SST anomalies in the Atlantic is little improved, and remains a significant challenge. The very recent improvements in the observing system, although important for our long term capabilities in this region, may be a complicating factor in our operational forecasts. The last 2-3 years of forecasts consistently show too cold SST anomalies in the early summer, relative to the earlier integrations which have a mean state which is too warm. This could be an error in the forecast of interannual variablity (and may be little to do with the local initial conditions), or an improvement in the forecasting of the mean seasonal cycle, stimulated by a more realistic initial state. Future work, e.g. using winds from the interim reanalysis, may help us better understand what is going on here.

System 3 (erwq) has a slightly higher amplitude of ENSO variability than S2, although it is still below the observed value as shown in fig 20. There is a seasonal dependence to this as shown in fig 24 (discussed later). An overall impression of what is to be expected from S3 is given by Fig.21, showing the actual ensemble forecasts for Nino 3.4 SST anomalies.

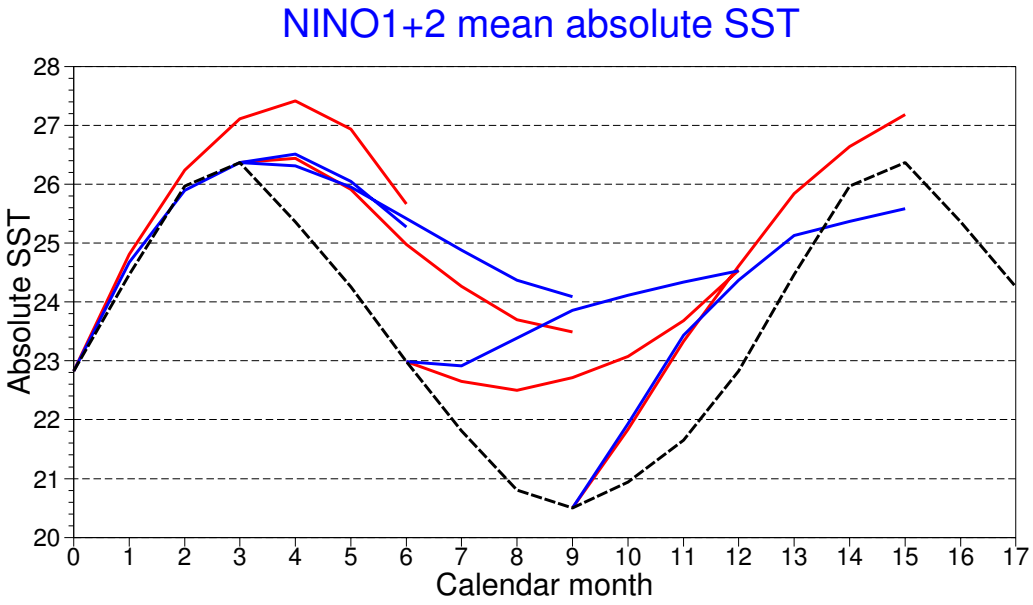



*Figure 16: Mean forecast SST in the Nino 1+2 region in the far east Pacific (0-10S, 80-90W), for S2 (blue) and erwq (red). The new model is a little warmer overall, but has a significantly better amplitude of the annual cycle.*

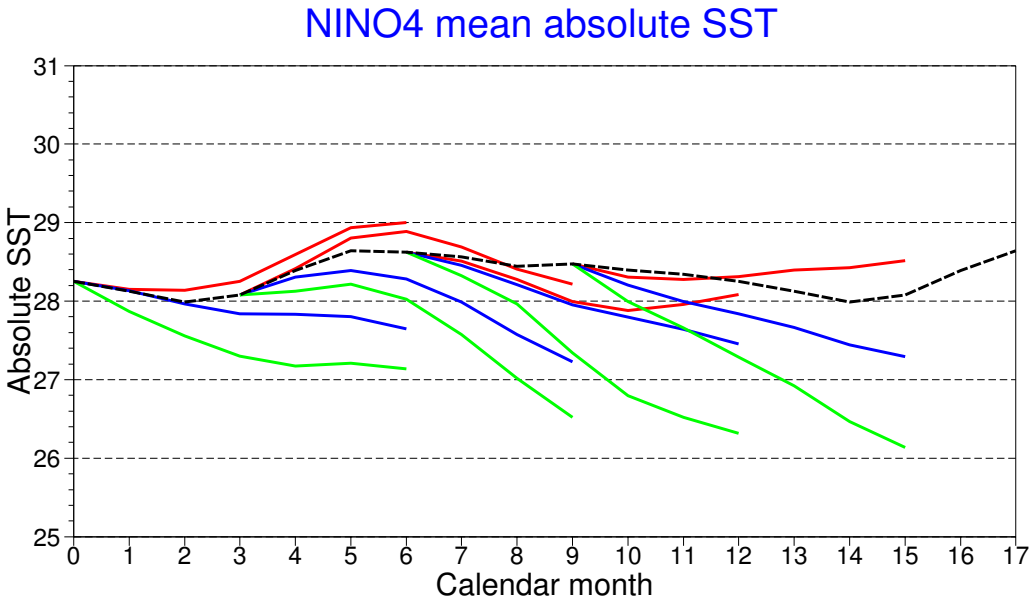
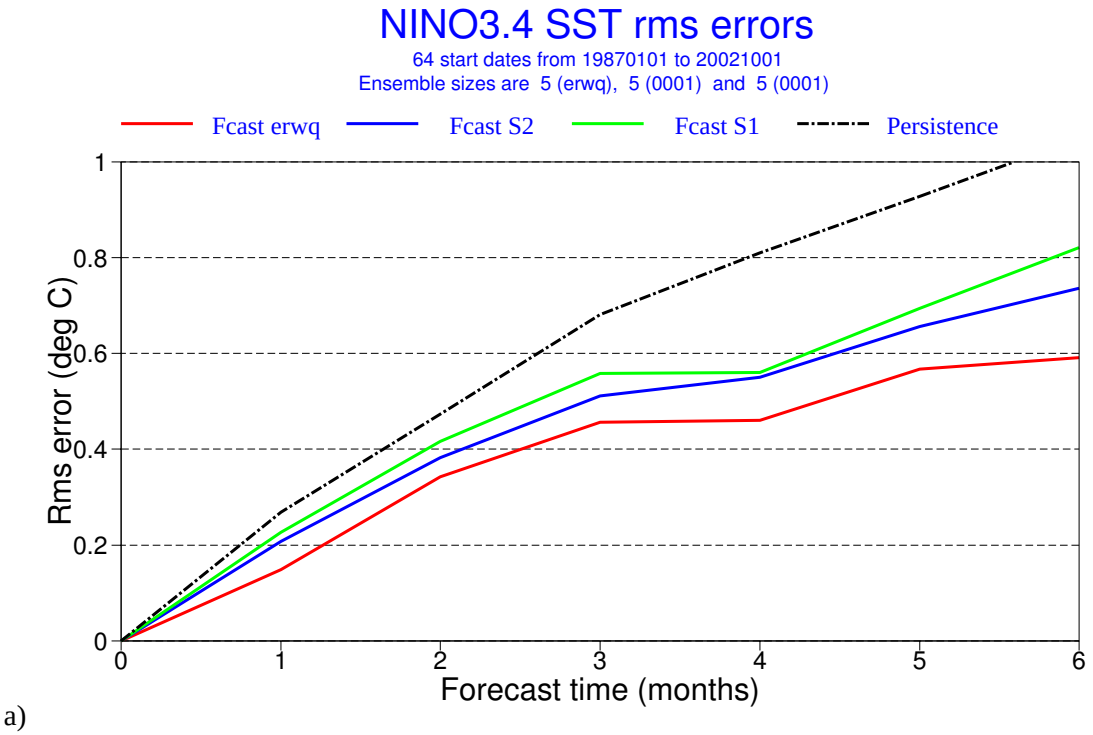
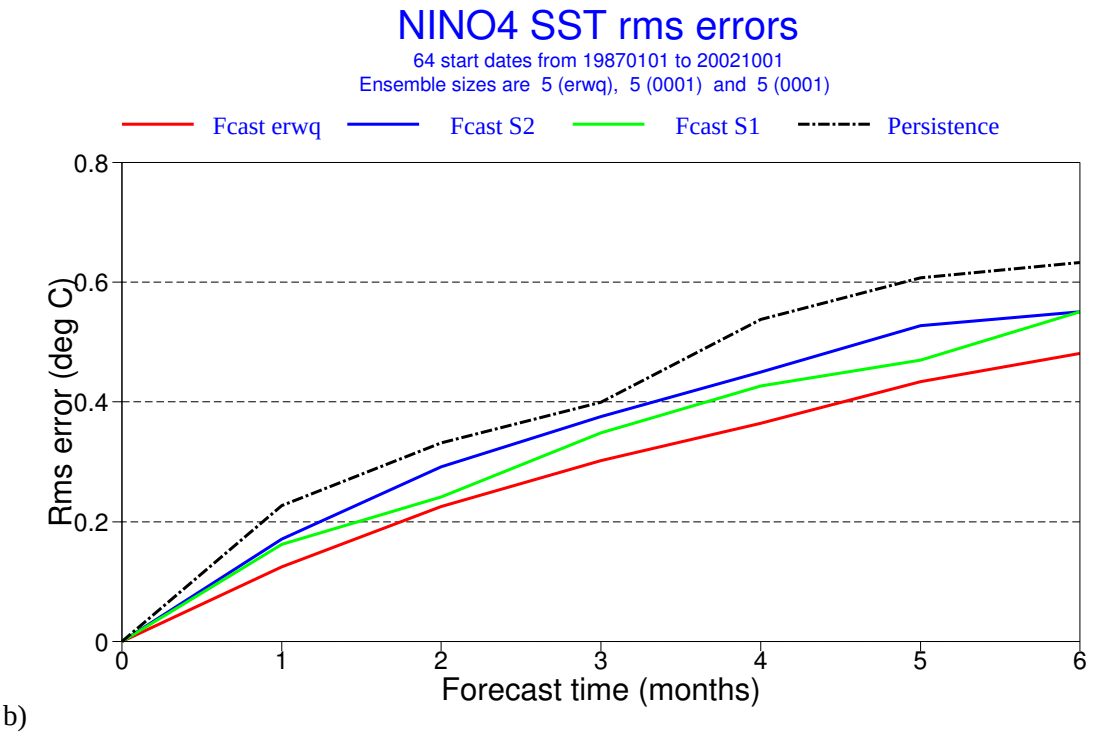



*Figure 17: Mean forecast SST averaged over the Nino-4 region (5N-5S, 160E-150W), for S1 (green), S2 (blue) and erwq (red). The new model has a reasonable balance between the cold tongue and warm pool in this region.*




a)




b)

*Figure 18: a) RMS errors for Nino 3.4 SST forecasts from S1 (green), S2 (blue) and erwq (red), for 64 forecasts in the period 1987-2002. b) RMS errors for Nino 4. Note that S2 was worse than the original S1, but this has more than been made up by erwq.)*

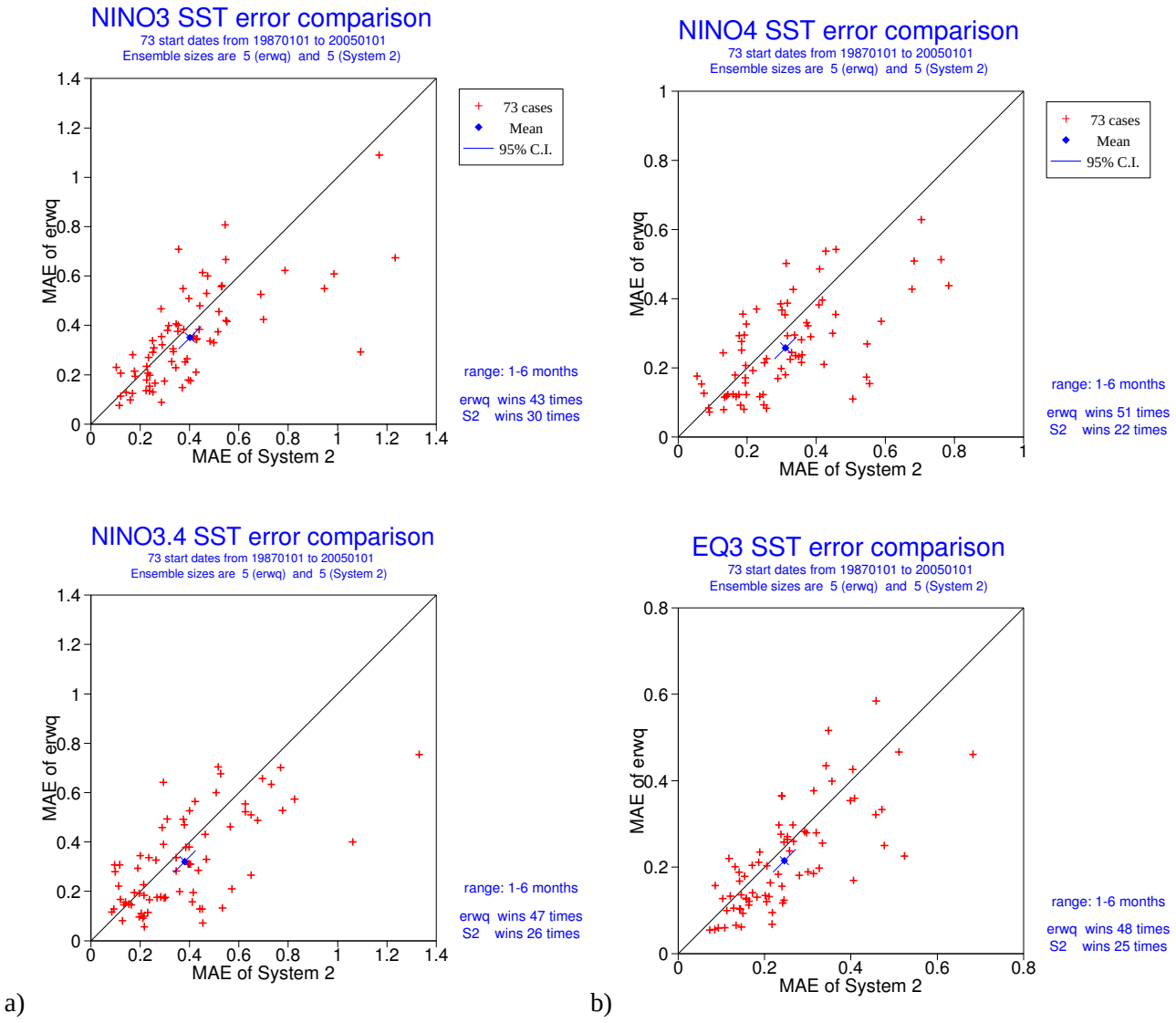



*Figure 19: Scatter diagrams for mean absolute SST error over the 6 month forecast, comparing erwq and S2 for various regions in the Pacific. In all regions, erwq is significantly better than S2.*

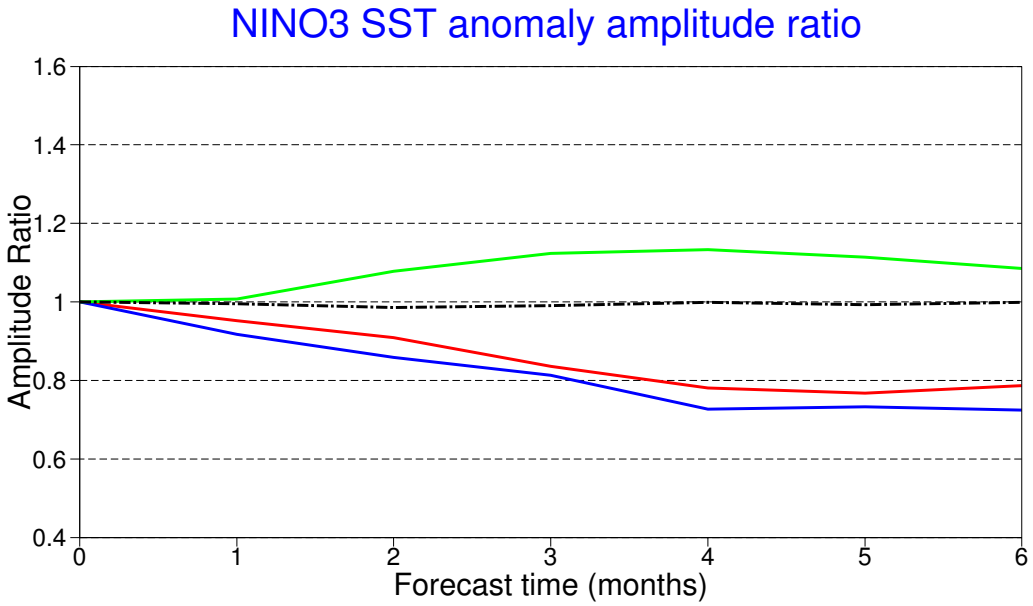



*Figure 20: Mean forecast SST variability averaged over the Nino-3 region (5N-5S,150W-90W), for S1 (green), S2 (blue) and erwq (red).*

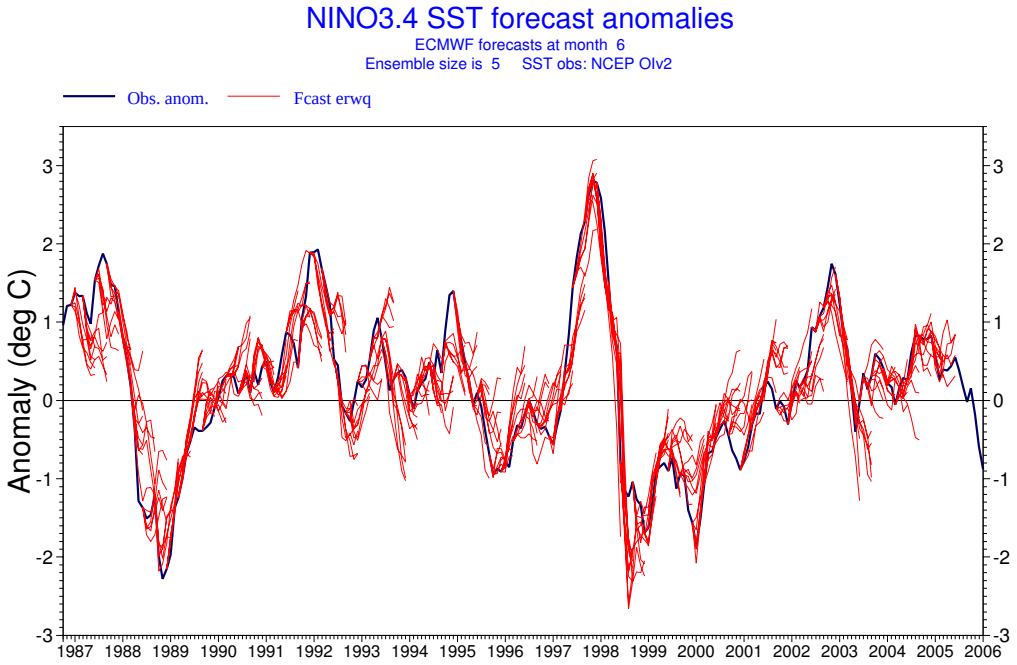



*Figure 21: Forecast trajectories for Nino 3.4 SST anomalies from experiment erwq. All 5 ensemble members are shown; forecasts are 6 months long and start at 3 month intervals. Black line shows observed SST anomalies from OIv2.*

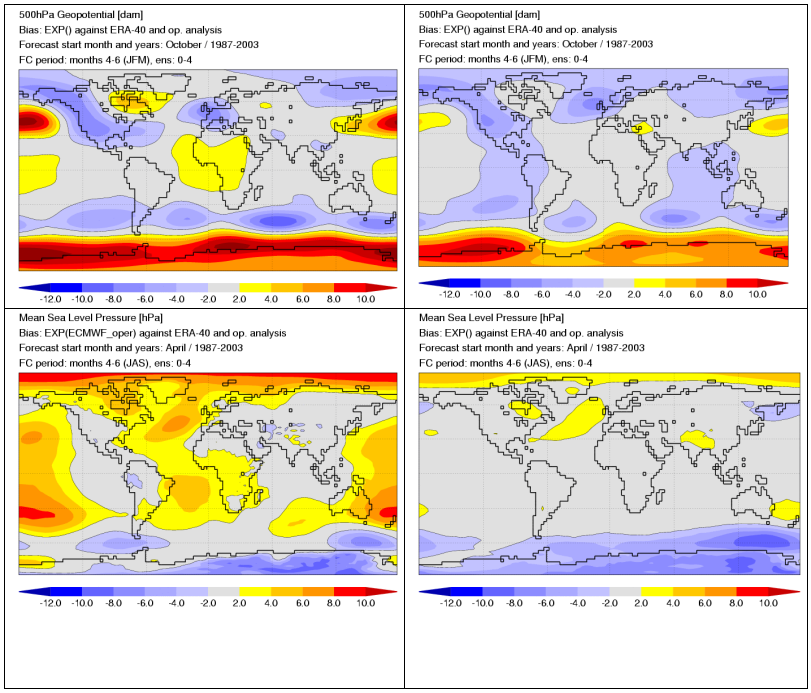



*Figure 22: Systematic errors of S2 (left) and erwq (right) experiments in the 4-to-6-month forecast range, for 500-hPa height in JFM (top) and mslp in JAS (bottom).*

## 5.2 Climatology, variability and predictability of atmospheric fields

### *5.2.1 Climatology and systematic errors*

The climatology of the atmospheric component of erwq shows substantial improvements with respect to S2. Systematic errors in geopotential height, sea-level pressure and lower-tropospheric temperature have been substantially reduced in both the tropical and the northern extra-tropical regions. Examples of systematic error reductions at month-4-to-6 range are provided by fig 22, which shows the mean error of 500-hPa height in JFM (from October initial conditions) and of mean-sea-level pressure (mslp) in JAS from April initial conditions, for S2 and S3 experiments.

For the 500-hPa height field (top row), a notable reduction in the model bias is found over the North-Pacific, where a large positive bias exceeding 12 dam in S2 has been reduced by almost a factor of 3. Mean errors over North America, which in S2 acted to decrease the amplitude of the stationary wave pattern, have also been substantially reduced, leading to a notable improvement in the zonally-asymmetric component of the time-mean flow. A negative bias of about 6 dam over Western Europe has been shifted to the north-west, unfortunately without any noticeable reduction. The location of such a negative bias in erwq is close to the region of highest blocking frequency in the East Atlantic sector, and therefore prevents any improvement in the simulated blocking statistics (see below).

In the mslp field for the boreal summer (bottom row), positive errors in the regions of the subtropical anticyclones over both the northern and southern oceans were present in S2, with amplitude between 4 and 8 hPa. These errors have been substantially reduced, and are hardly visible in the erwq bias map with the same contour level. A positive bias over the Arctic Ocean has also been reduced by about a factor of 2, while the negative

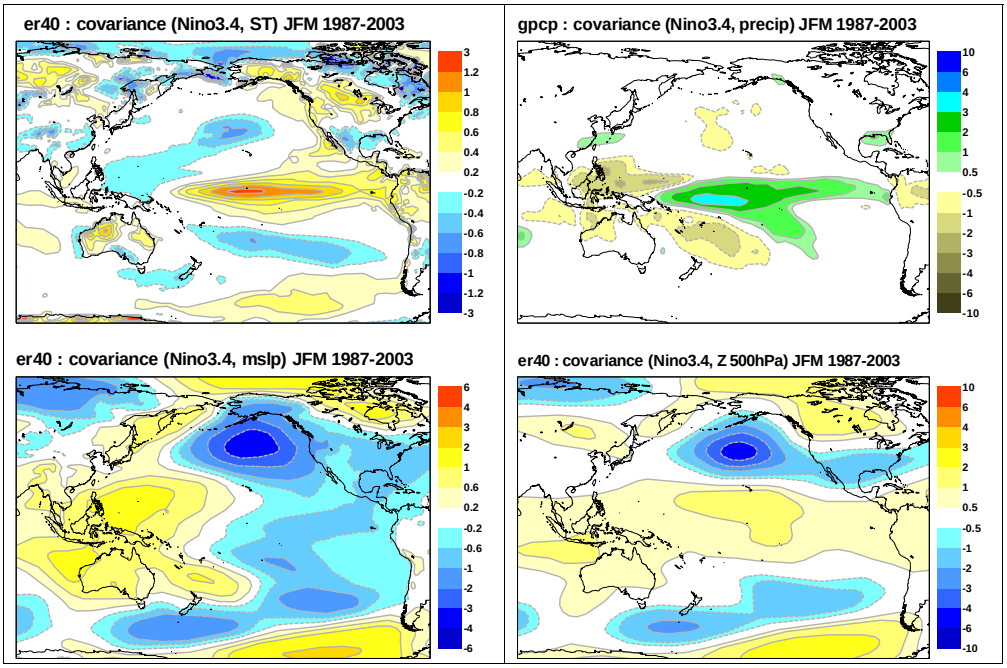



*Figure 23: Covariance of observed JFM anomalies with the standardized Nino3.4 index. Top left: surface temperature, from ERA-40. Bottom left: mean-sea-level pressure, from ERA-40. Top right: precipitation, from GPCP. Bottom right: 500-hPa geopotential height, from ERA-40.*

bias over the southern polar regions has been partially increased.

### *5.2.2 Interannual variability and teleconnections associated with ENSO.*

The interannual variability associated with ENSO can be diagnosed by plotting the covariance maps of various fields with the standardized time series of the Nino3.4 index. These maps can be interpreted as the anomalies corresponding to one standard deviation of the Nino3.4 index in the system considered. Figure 23 shows covariance maps for the January-to-March (JFM) season, for surface temperature, sea-level pressure, precipitation and 500-hPa height computed from ERA-40 data (except for precipitation, which is from the Global Precipitation Climatology Project (GPCP) dataset). These maps are used as a reference for covariances derived from the 17-year, 5-member ensemble experiments mentioned in the previous section.

Covariance maps of surface temperature and mean-sea-level pressure (mslp) anomalies from S2 and erwq ensemble experiments are shown in fig.24, and can be compared with ERA results in the left column of fig. 23. For both systems, the SST anomaly associated with ENSO has a smaller amplitude (by 20-30%) and a maximum shifted to the east with respect to observed values. For this particular season, S2 produces a slightly larger SST signal than erwq, although this is not generally true. For example fig20 based on four different start months shows S3 to be less damped than S2. A weaker-than-observed signal is also found in the sea-level pressure field for both systems, although with a fairly realistic spatial distribution. Over the western Pacific and eastern Indian ocean, a closer correspondence with observations is found in S3. Despite the reduced amplitude of the mslp covariance, normalised indices of interannual variability such as the 'equatorial Southern Oscillation index' (computed from standardised mslp anomalies averaged over two cross-equatorial boxes in

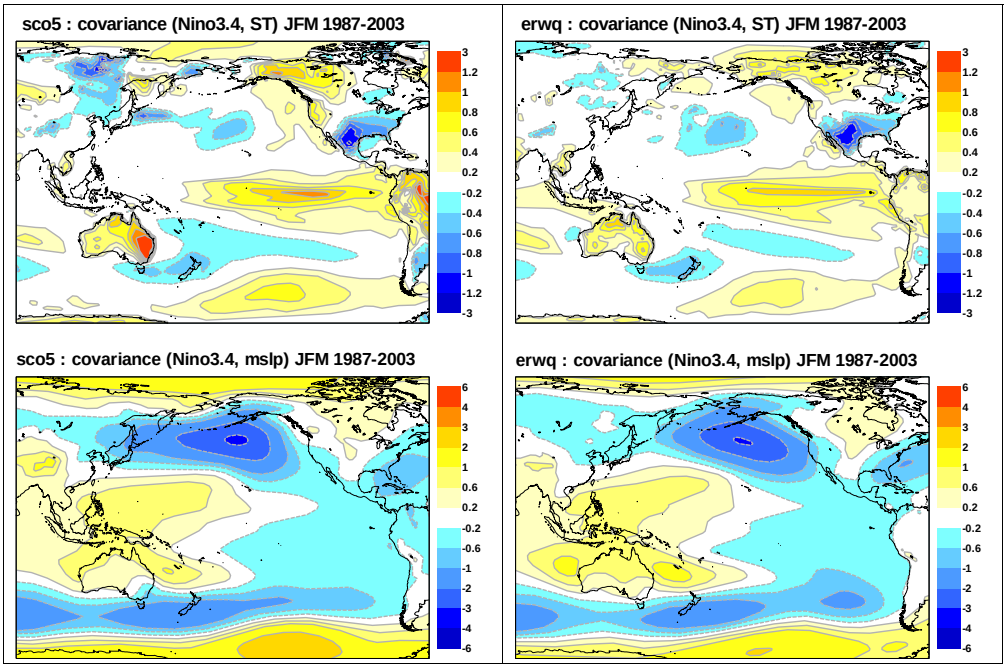



*Figure 24: Top row: JFM surface temperature covariance with the standardized Nino3.4 index, for S2 (left) and erwq (right) ensembles. Bottom row: JFM mean-sea-level pressure covariance with the standardized Nino3.4 index, for S2 (left) and erwq (right) ensembles.*

the East and West Pacific) from both S2 and erwq experiments are highly correlated with observations, with correlation coefficients close to 90% (89% for S2 and 91% for erwq, not shown) when ensemble-mean values are considered.

In erwq, both the seasonal mean and the interannual variability of rainfall over the tropical oceans is generally reduced compared to S2 values, bringing the model climatology into closer agreement with GPCP observational data. The spatial distribution of modelled rainfall is notably improved in the tropical Pacific during the boreal winter. While in S2 rainfall in the eastern Pacific ITCZ exceeds observations by (at least) a factor of 2, erwq simulates a more correct ratio between rainfall in the western and eastern parts of the ocean. The improvement in the mean field is reflected in the distribution of rainfall interannual variability. The top panels of fig. 25 compare the standard deviation of January-to-March (JFM) rainfall in the ensembles run respectively with S2 (left panel) and erwq (right panel). While the S2 variability shows two distinct maxima (with similar amplitude) in the western and eastern tropical Pacific, erwq simulates a single variability maximum located just west of the dateline, in closer agreement with observations.

The improved distribution of tropical rainfall in erwq is reflected in the patterns of rainfall anomalies associated with ENSO events. In the central row of Fig. 25, the covariance patterns of JFM rainfall with the (standardised) Nino3.4 index in the same season are displayed for S2 and erwq. For S2, the double maximum in the total variability is reproduced in the Nino3.4 covariance, while in the erwq covariance the largest positive values are correctly located close to the Equator/dateline intersection (cf. top-right panel in 23).

On the other hand, the meridional extent of the region with positive rainfall anomalies is still underestimated in erwq compared with GPCP data, and the area-averaged heat-source anomaly appears to be reduced with respect to S2. The consequent decrease in net diabatic forcing for the atmosphere causes a partial reduction in the

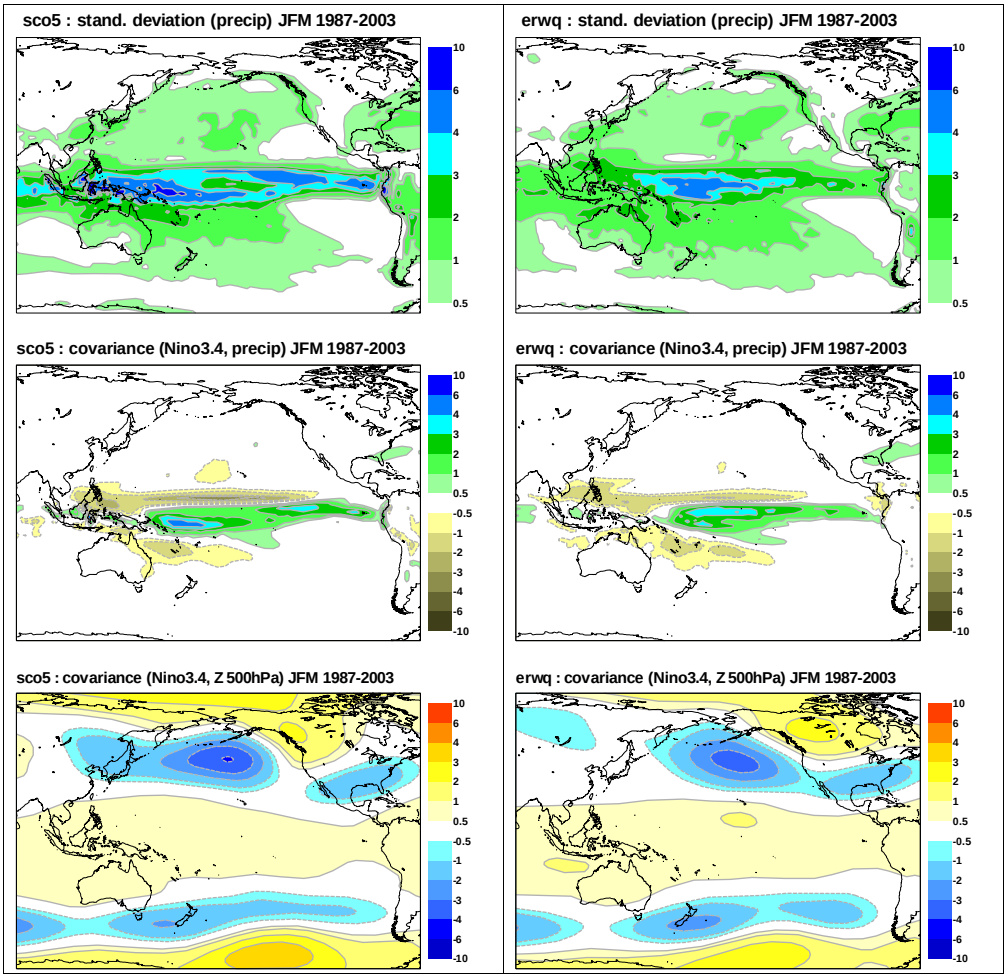



*Figure 25: Top row: standard deviation of seasonal-mean rainfall over the Pacific in JFM, for a S2 ensemble (sco5, left h.s.) and an S3 ensemble (erwq, right h.s.). Central row: JFM rainfall covariance with a standardized Nino3.4 index, for S2 (left) and erwq (right) ensembles. Bottom row: JFM 500-hPa height covariance with a standardized Nino3.4 index, for S2 (left) and erwq (right) ensembles.*

strength of the atmospheric teleconnections relative to S2 (see the 500-hPa geopotential height anomalies in the bottom panels of Fig. 25), which has a negative effect on the skill scores for the northern extratropical regions in winter in this set of experiments. In S2, some compensation occurred between the excessive amplitude of rainfall anomalies and the smaller-than-observed SST signal associated with anomalous ENSO events, while such compensation has been removed by the more realistic rainfall distribution in erwq.

### *5.2.3 Predictions of interannual variability for large-scale teleconnection patterns.*

A measure of the skill in the prediction of interannual variability of large-scale atmospheric anomalies can be gained by comparing time series of projections of observed and predicted anomalies on specific teleconnection patterns. To compare S2 and erwq performance, three specific patterns will be considered here, which are computed from the covariance of seasonal mean anomalies from ERA-40 with the following standardized indices:

* The Pacific North American (PNA) teleconnection index, defined from standardized anomalies of 500-hPa height in DJF following Wallace and Gutzler (1981);

* The North Atlantic Oscillation (NAO) index, defined as the difference of standardized mslp anomalies over Iceland and the Azores islands (e.g. Hurrell 1995);

* The All-India Rainfall (AIR) index, computed from station data over India by Parthasarathy et al. (1995)

Consistent with the definition of the corresponding indices, covariances with 500-hPa height and mslp DJF anomalies are used to characterize the PNA and NAO patterns. These well-known patterns are shown in the upper and central panel of Fig.26. For the AIR, rather than looking at rainfall predictions over India (which should be properly calibrated to account for differences between observational and model statistics) we prefer to consider the co-varying mslp anomaly over a larger domain, including the tropical Indian ocean and part of the western Pacific (bottom panel in Fig.26). This pattern (which for brevity will be referred to as the AIR-mslp) shows an anomalous depression centred over the Arabic Sea and spreading to the whole western Indian Ocean, which increases the flow of moist air from the Bay of Bengal into the Indian peninsula. Over the West Pacific, the AIR-mslp anomaly shows a dipolar structure, with positive anomalies around 15 N and negative anomalies over Indonesia.

For each seasonal-mean field from ERA-40 and the S2 and erwq ensemble members in the 1987-2003 period, the corresponding anomaly A(t) is decomposed into a component parallel to the teleconnection pattern P and a spatially-orthogonal component O(t):

$$A(t) = c(t)P + O(t) \tag{3}$$

The projection coefficient c(t) is given by the ratio of the inner products:

$$c(t) = < A(t), P > / < P, P > \tag{4}$$

computed over a specified spatial domain (which corresponds to the regions shown in the panels of Fig.26).

Comparisons of model and analysis projections onto the three patterns in Fig. 26 are given in Fig. 27 for S2 and in Fig.28 for erwq. In each panel, corresponding to a specific teleconnection, the red dots correspond to ERA-40 values, green dots to individual ensemble members and blue dots to ensemble means. The correlation between the ERA-40 and ensemble-mean time series is shown in the caption above each panel.

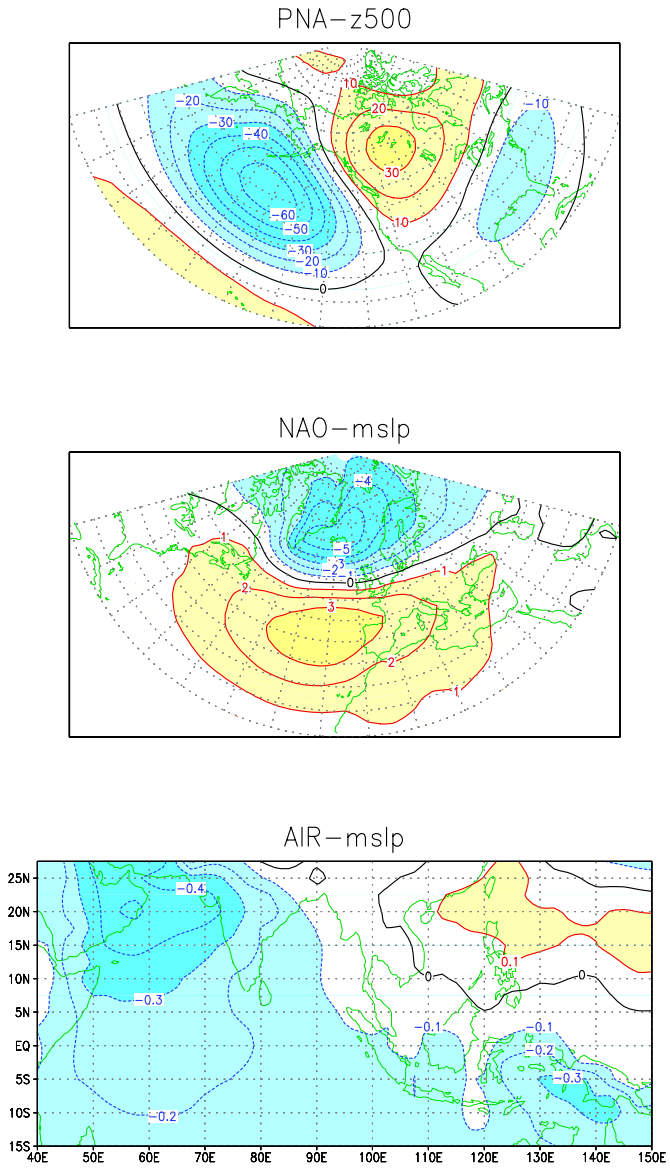



*Figure 26: Teleconnection patterns defined as the covariance of 500-hPa height anomalies with the PNA index (top), and of mslp anomalies with the NAO (centre) and AIR (bottom) indices.*

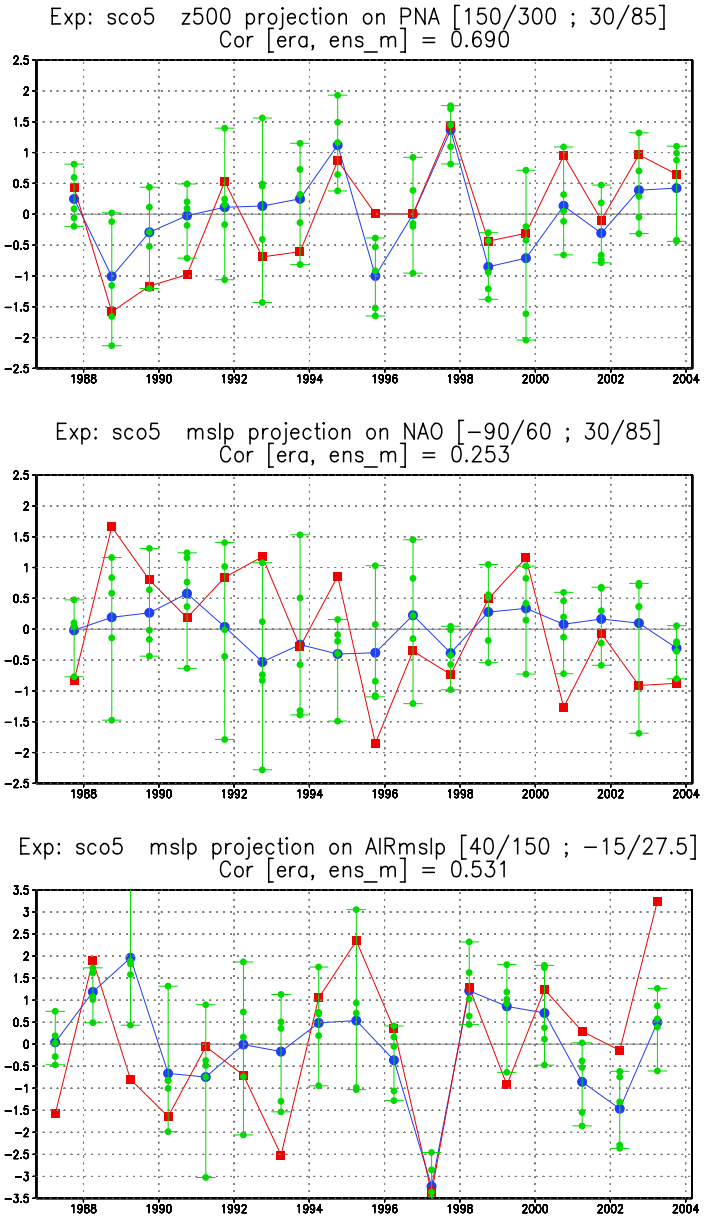



*Figure 27: Projections of seasonal-mean anomalies onto teleconnection patterns in Fig26 during the 1987-2003 period, for the S2 (sco5) ensemble experiment. (DJF data for PNA and NAO, JJA data for AIR-mslp; 500-hPa anomalies for PNA, mslp anomalies for NAO and AIR-mslp). Red dots correspond to ERA-40 values, green dots to individual ensemble members and blue dots to ensemble means. The correlation between the ERA-40 and ensemble-mean time series is shown in the caption above each panel; the lat-lon limits of the projection area are in brackets.*

As mentioned above, the smaller amplitude of the rainfall signal associated with ENSO in erwq than in S2 also reduces the amplitude of the ENSO extratropical teleconnection, which is strongly correlated with the traditional PNA pattern (see Straus and Shukla 2002 for further discussions on this point). As a consequence, the predictive skill for the PNA projection in DJF is higher for S2 than for erwq, with the latter system showing weaker projections during the strong ENSO events in the 1987-2003 period (and the ERA-ensemble mean correlation dropping from 69% to 46%).

For the NAO, erwq shows no predictive skill, which appears as a worse performance than S2. However, given the marginal positive skill of S2, this change has little practical relevance, and it is fair to say that both systems seem to be below the limit of predictive skill provided by the empirical scheme of Rodwell and Folland (2002).

On the other hand, consistent with the improvements in tropical climatology and variability discussed elsewhere, erwq delivers a better prediction of the AIR-mslp pattern variability. A notably better performance is shown during the late 80s and early 90s, and specifically during the 1987 and 1988 summers, characterized by large and opposite anomalies of the AIR index. On average, the ERA-ensemble mean correlation increases from 53% to 68%.

### *5.2.4 Intraseasonal variability*

Internal atmospheric variability is generally higher in erwq than in S2, both in tropical and extratropical regions. For the tropics, a notable improvement is found in the amplitude of intraseasonal variability in the 20-to-70-day frequency range, which includes the Madden-Julian Oscillation (although, in this respect, cycle 31r1 shows a partial degradation when compared to the previously tested cycle 30r2). The standard deviation of tropical velocity potential anomalies at 200 hPa in the October-to-March season is shown in Fig. 29 for ERA-40, S2 and erwq experiments, using a bandpass filter to isolate oscillations with periods between 20 and 70 days. Although the location of the variability maxima over the Indian and West pacific oceans is in good agreement with reanalysis data, the amplitude is underestimated by both systems. However, in erwq the amplitude difference with respect to ERA-40 is reduced by 30-40 % with respect to S2.

The spectral distribution of the velocity potential variability is further analysed in Fig. 30 where the variance of 200-hPa anomalies, averaged between 5N and 5S, is shown as a function of longitude and oscillation period. In addition to ERA-40, S2 and erwq data, results from an ensemble experiment with the 30r2 cycle are also shown. Note that differences between various systems are enhanced by the use of squared-amplitude data. Although the erwq results represent an improvement with respect to S2 simulations, cycle 31r1 fails to generate a variance maximum in the MJO frequency range, which was simulated by cycle 30r2.

In the northern extratropics, internal variability has increased for both the high-frequency and the low-frequency range in going from S2 to erwq. Fig. 31 compares standard deviations of 500-hPa in boreal winter (DJF) for two spectral bands, one with periods between 2 and 8 days (typical of synoptic-scale baroclinic systems), the other one with periods between 10 and 30 days (comparable to the typical duration of large-scale circulation regimes such as blocking). Both the S2 and erwq systems produced realistic spatial distributions of variability compared to ERA-40 data. In the high-frequency range, the maximum associated with the North Pacific storm track is better positioned in erwq than in S2, although with a larger-than-observed amplitude; erwq also shows more variability over eastern Europe and north-west Asia. Low-frequency variability shows a moderate and rather uniform increase from S2 to erwq, with erwq values in better agreement with ERA-40 data.

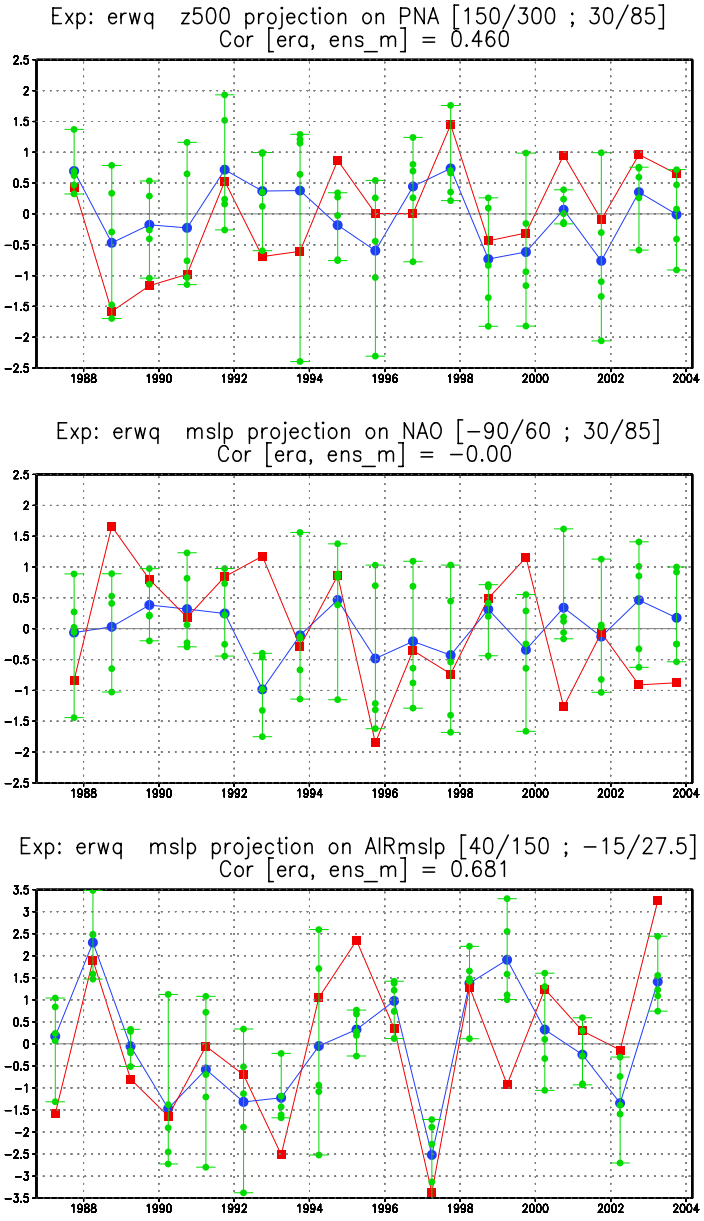



*Figure 28: As Fig. 27, but for the S3 (erwq) ensemble experiment.*

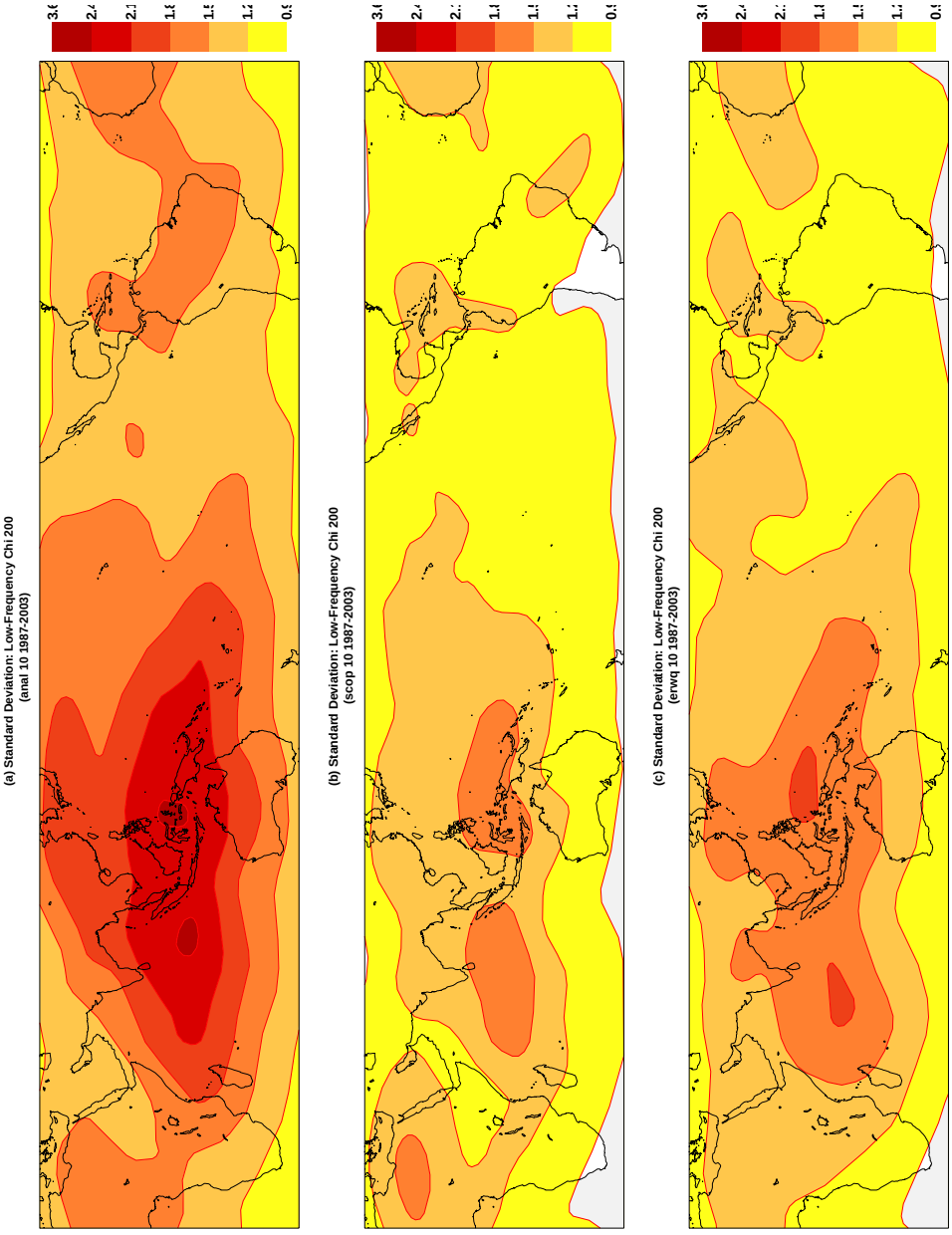



*Figure 29: Standard deviation of 200-hPa velocity potential anomalies in the 20-to-70-day period range during the Oct.-March season, from ERA-40 (top), S2 (centre) and erwq (bottom) experiments.*

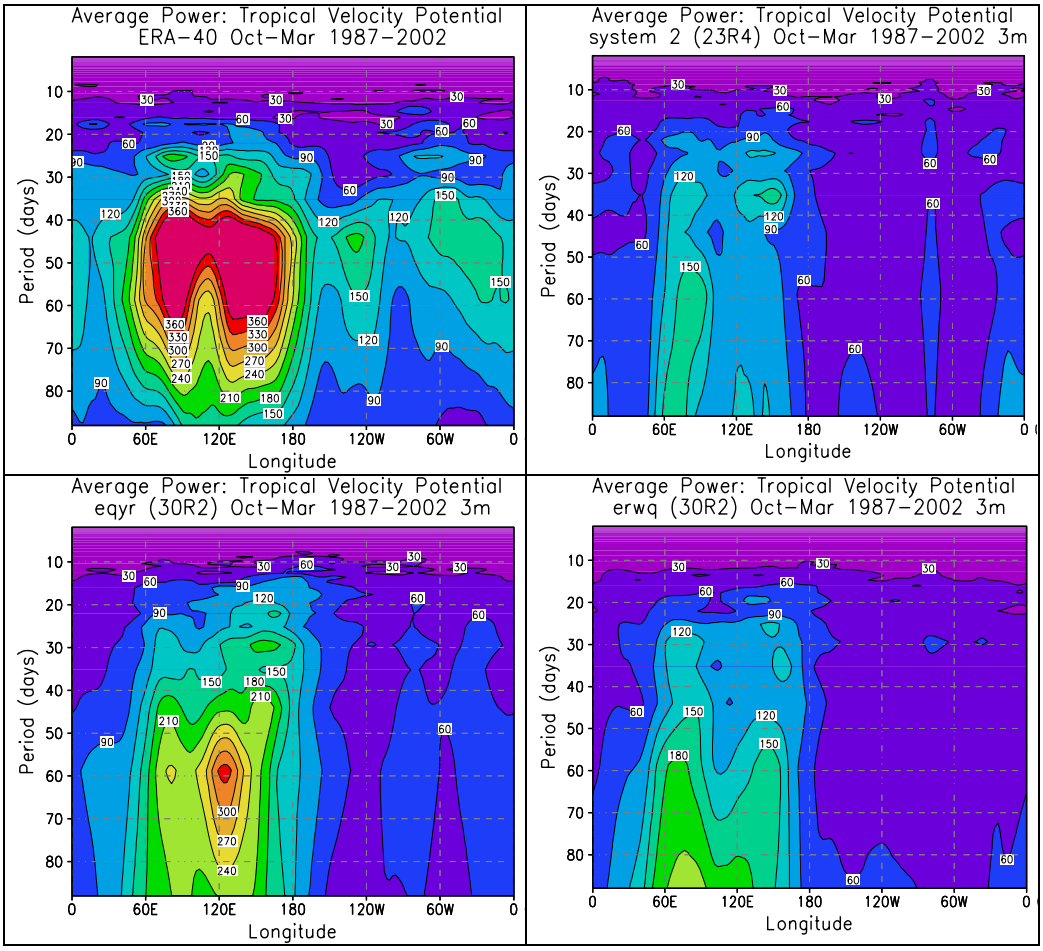



*Figure 30: Spectra of 200-hPa velocity potential anomalies in the Oct.-March season as a function of longitude and period, for ERA-40 (top left) S2 (top right), cycle 30r2 (bottom left) and erwq (bottom right) experiments.*

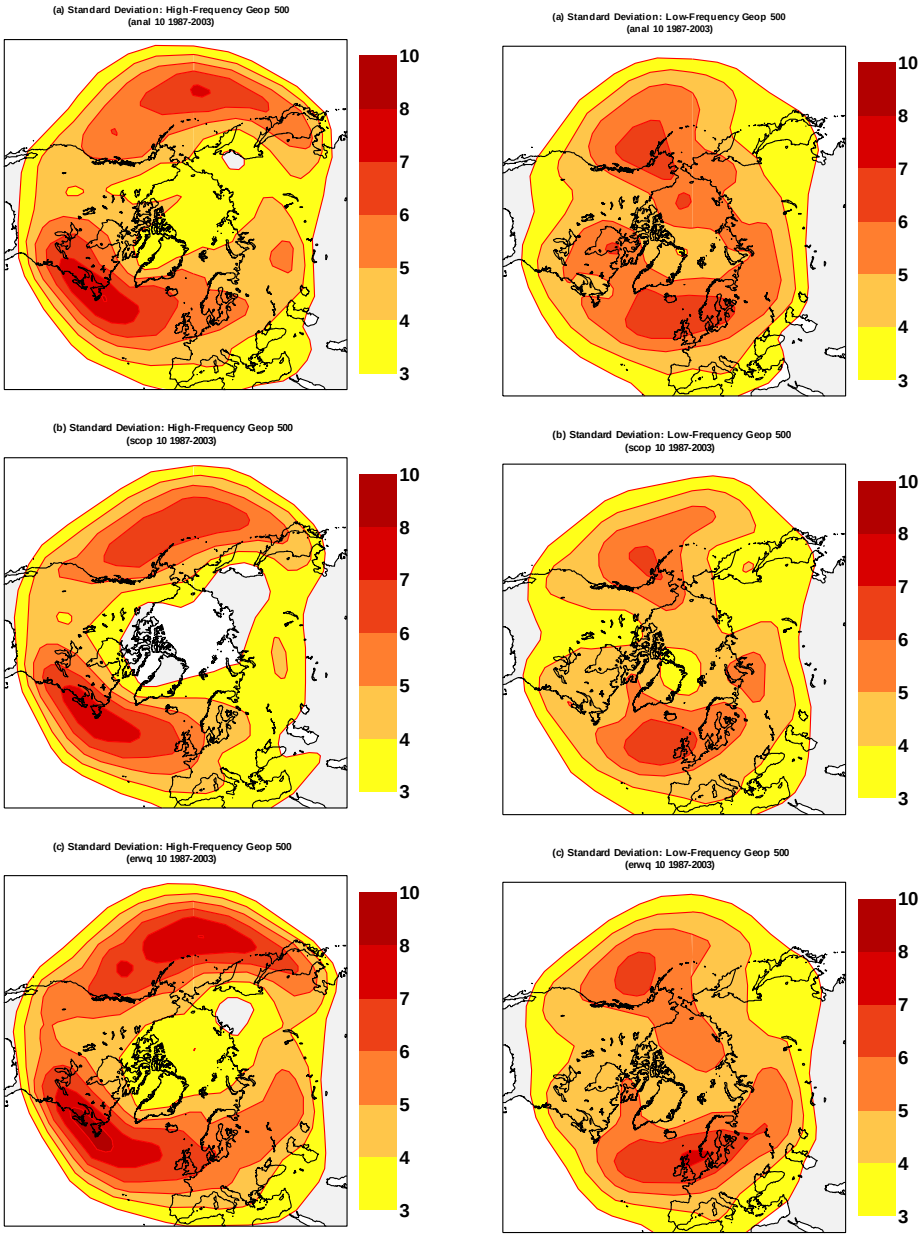



*Figure 31: Standard deviation of 500-hPa geopotential height anomalies in the 2-to-8-day (left column) and 10-to-30-day (right column) period ranges during the Dec.-Feb. season, from ERA-40 (top), S2 (centre) and erwq (bottom) experiments.*

### *5.2.5 Blocking*

A blocking pattern is characterized by a region of warm air with higher than ambient pressure, so that the situation is essentially recognized by a quasi-stationary, positive height anomaly wherein the normal eastward progression of migrating extra-tropical weather systems is deflected. Preferred Northern Hemisphere locations for the appearance of blocking events are to the north and the east of the Pacific and Atlantic storm tracks. The longest-lasting blocks are responsible for relatively long drought (summer) and cold spell (winter) episodes.

In order to make reliable long-range forecasts, a model should be able to simulate blocks with the right amplitude, location and persistence. Therefore, the ability of climate models to represent blocking as a main source of extra-tropical intra-seasonal variability needs to be assessed. Several indices have been proposed to detect blocking events in analyses and model simulations. One of the best known is the Tibaldi and Molteni (1990) index. A split-up flow pattern is detected by using a certain threshold for the meridional gradient of geopotential height at each longitude. In addition, the selected events should cover a large enough region and persist for at least five days. The specific criteria used in this study are thoroughly described in Doblas-Reyes et al. (2002).

Figure 32 shows the average blocking frequency for ERA-40, S2 and experiment erwq. Both versions of the coupled model simulate the maxima of blocking frequency over the Euro-Atlantic and North Pacific regions. Note that the confidence intervals for the model are narrower than for ERA-40, which is due to the larger sample available when using ensemble simulations and to an underestimation of the interannual variability of blocking. Both S2 and erwq winter hindcasts underestimate the blocking frequency over most of the Northern Hemisphere. The bias is more obvious over the North Pacific, although the western Atlantic blocking is also underestimated. These differences are significant with a 95% confidence over most of the longitudes. The results are representative of the model behaviour in other seasons. Blocking undersimulation in the model is mainly a consequence of a specific zonal wind bias associated with a decrease in the frequency of large-scale ridges over the main blocking regions.

### *5.2.6 General characteristics of tropical cyclones in S3 compared to S2.*

*Introduction*

The same procedure for tracking model tropical storms used for S2 has been applied to 5-ensemble member integrations which mimic the future configuration of S3. The starting dates of the forecasts are: 1st January, 1st April, 1st July and 1st October 1987 to 2004. The statistics of the model tropical storms obtained with erwq are then compared to the statistics obtained with the 5-ensemble member hindcast of S2 (from 1987 to 2001) and the first 5 ensemble members of the S2 real-time forecasts for the period 2002-2004. The starting dates for S2 are the same as the starting dates used for erwq.

*Climatology*

The number of tropical storms detected in S2 and erwq has been averaged over the period 1987-2004, the 5 ensemble members and the 4 annual starting dates for each individual ocean basin. Figure 33 displays the mean annual frequency of tropical storms for each ocean basin along with observations. Results suggest that erwq produces more tropical storms than S2 over all the ocean basins. The increased horizontal resolution (T159 instead of T95) is the most likely explanation. Experiment erwq seems to produce a more realistic climatology of tropical storms than S2 over all the ocean basins, although erwq produces less tropical storms than observed over the western North Pacific, eastern North Pacific, North Atlantic and South Indian Ocean, and more tropical storms than observed over the North Indian Ocean, Australian Basin and South Pacific. Therefore, in terms of climatology of tropical storm frequency, erwq seems to represents a clear improvement over S2.

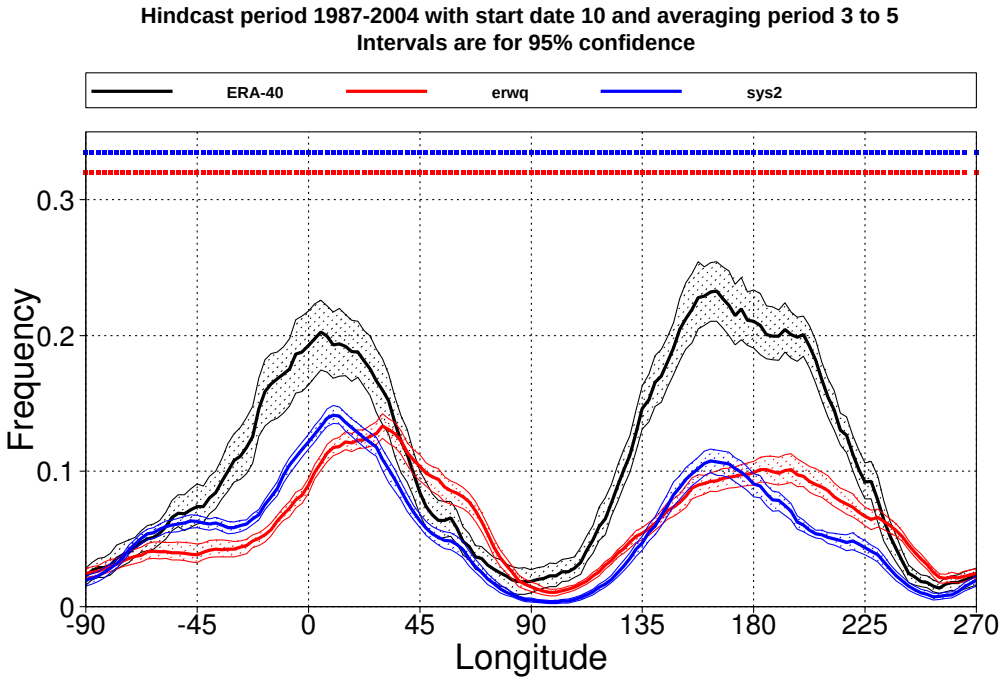



*Figure 32: Northern Hemisphere winter average blocking frequency for ERA-40 (black line), S2 (blue line) and the experimental erwq (red line) for the period 1987-2004. The model index has been computed using 5-member ensembles initialized on the 1st of October and the results shown are for the season December-to-February. The shaded areas around each bold line correspond to the 95% confidence interval computed using a bootstrap with a sample size of 500. The dots on top of the figure display the longitudes where the model blocking frequency is significantly different from the ERA-40 estimate with 95% confidence. A 2-sample test based on a 500-sample bootstrap method was used for the inference test.*

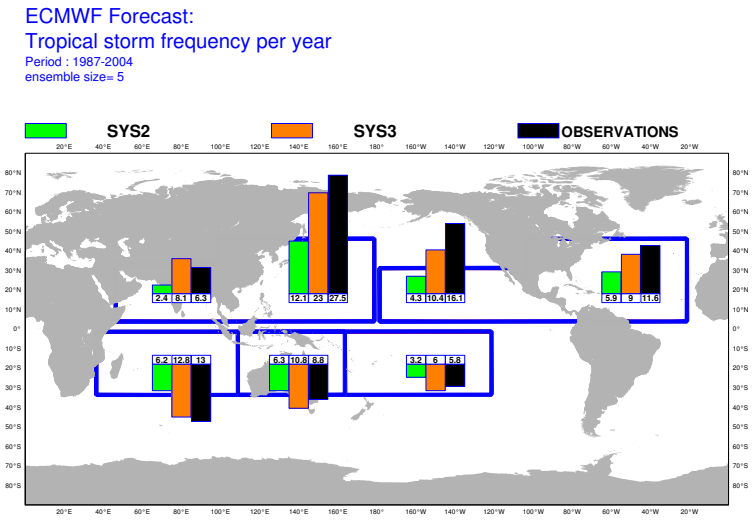



*Figure 33: Tropical storm annual frequency for the period 1987-2004 in observations (black bars), S2 (green bars) and erwq (orange bars)*

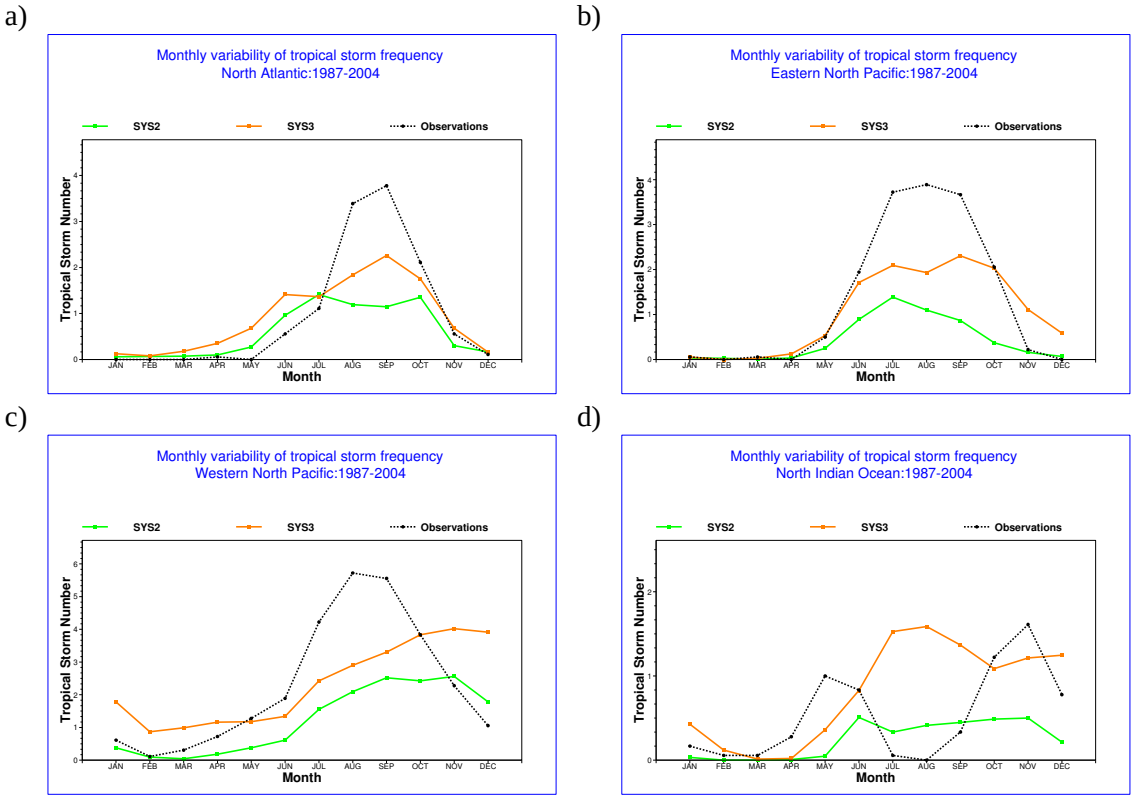



*Figure 34: Tropical storm seasonal cycle over a) the Atlantic basin b) the eastern North Pacific c) the western North Pacific d) the North Indian ocean . The black line represents the observed seasonal cycle, the orange line represents erwq and the green line represents S2.*

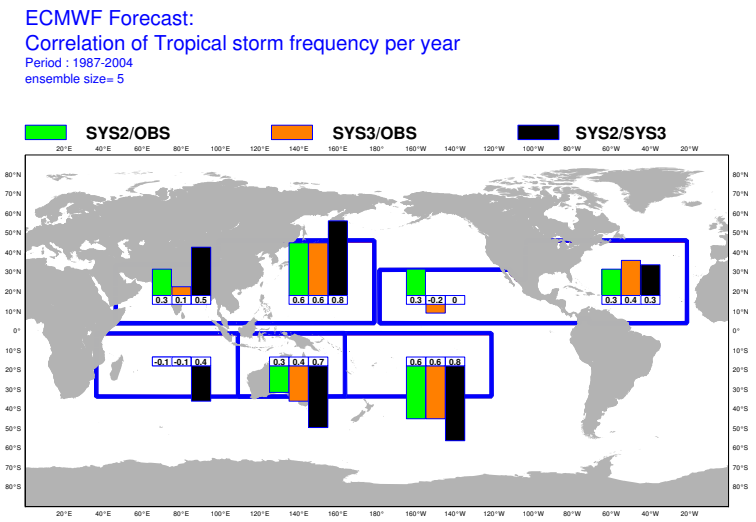



*Figure 35: Linear correlation between the observed interannual variability of tropical storms from 1987 to 2004 and the model ensemble mean. Over the North Atlantic, the eastern North Pacific and the western North Pacific, the starting date is 1st April, over the North Indian Ocean, the starting date is 1st July and over the Southern Hemisphere the starting date is 1st October. The black line represents the linear correlation between S2 and erwq*

*Seasonal cycle*

The number of tropical storms detected in S2 and erwq has been averaged over the period 1987-2004, the 5 ensemble members and the 4 annual starting dates for each individual ocean basin for each individual month. As fig 34 shows, the seasonal cycle of tropical storm frequency over the Atlantic is more realistic with erwq than with S2. In particular, erwq produces a peak activity in September as observed. However erwq produces too many tropical storms in March, April, May and June, when very few storms are observed. Over the eastern North Pacific, erwq also displays a more realistic seasonal cycle of tropical storm frequency. The model still produces too few tropical storms in July, August and September, but this is better than in S2. Over the western North Pacific and North Indian Ocean, the seasonal cycle is not well reproduced by either system, and erwq is not really an improvement.

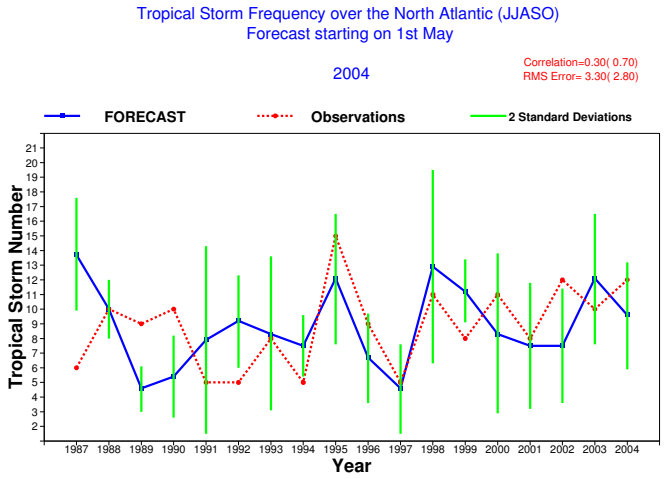
*Interannual variability*

The linear correlations between the interannual variability of observed tropical storms and the ensemble mean for S2 and erwq have been computed.Fig (35) suggests that erwq does not really improve those scores. The black bar represents the linear correlation between the interannual variabilities of S2 and erwq and shows that over the western North Pacific and South Pacific both models produce very similar interannual variabilities of tropical storms. As an example, Figure 36 displays the interannual variability of Atlantic tropical storms. The model is still not very successful over the Atlantic. Forecasts starting on 1st July are as poor. Both UKMO and Meteo-France model perform much better over the Atlantic than the ECMWF model.

*Tropical storm tracks*

The tropical storm tracks are clearly much longer and realistic in the erwq experiment than in S2, as expected from the increased horizontal resolution. Figure 37 shows an example of tracks over the eastern North Pacific and the North Atlantic. Figure 38 also shows an example over the South Indian Ocean. These figures show

a)




b)

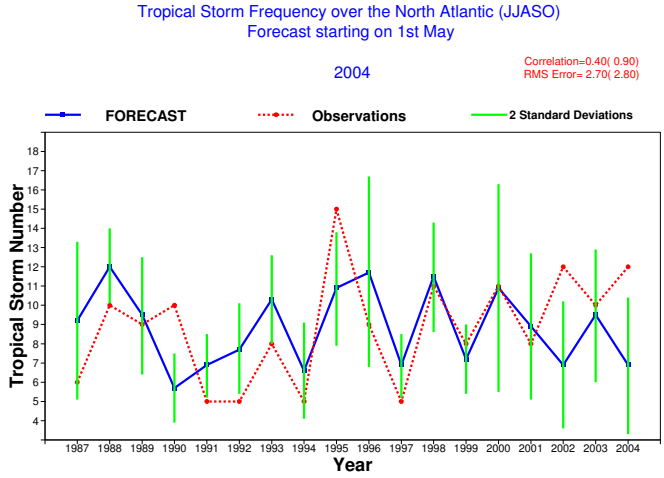



*Figure 36: Interannual variability of Atlantic tropical storms from the ensemble mean of a) S2, b) erwq (blue line) and observations (red dotted line). The vertical green line represents 2 standard deviations.*

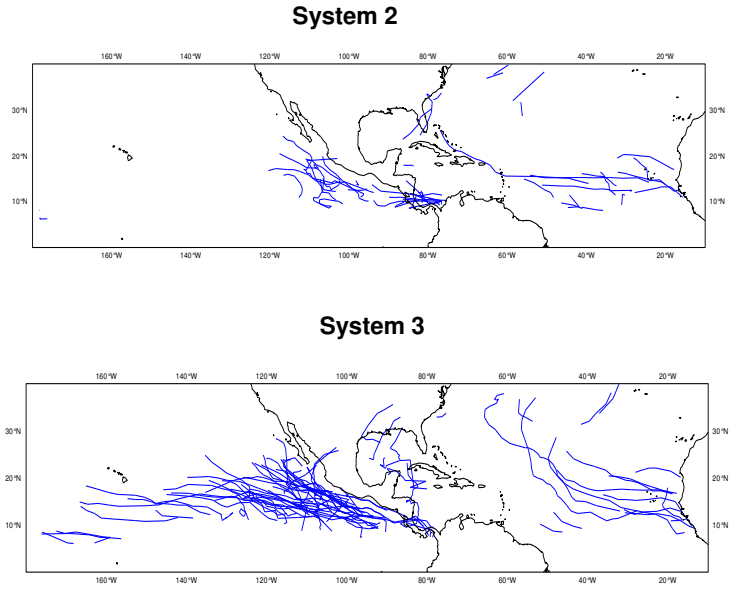



*Figure 37: Model tropical storm tracks in the Atlantic and Eastern North Pacific for S2 (top panel) and erwq (bottom panel). Forecasts starts on 1st July 2004.*

storms that last longer and recurve, and in the case of Figure 37 the tropical storms are more likely to make a landfall.

As an example of the possibilities of issuing landfall forecasts, Figures 39 and 40 show all the tracks predicted by the 5 ensemble members starting from July 1999 to April 2000 (La Nina conditions) and from July 1997 to April 1998 (strong El-Niño conditions). It is well known that ENSO has an impact on the tracks and position of storms (see the difference in the Central Pacific for example), and this has an impact on the risk of landfall. Both systems show roughly the same impact of ENSO on the position and location of storms: tropical storms forming more eastward in the western Pacific, and more tropical cyclone activity east of Madagascar during the El-Niño event than during the La-Niña event. However with S2 too few storms make a landfall, which makes it very difficult to issue explicit forecasts of tropical storm landfalls. Although there is a feeling that there is less landfall over South China during the 1997 El-Niño event with S2 than in 1998, the difference is not statistically significant. With erwq on the other hand the number of storms making a landfall is much larger, and the difference in terms of landfalls between 1997 and 1998 over South China, Mozambique, Australia is much more striking. Therefore it should be possible to look at the risk of landfall in erwq by just looking at the number of model tropical storms that make a landfall.

However there are a few concerns about tropical storm tracks in erwq: tracks may be too zonal. Too few storms make a landfall over Japan for instance. Over the Atlantic there is a surprising lack of tropical storm activity between 20N and 30N. That may be linked to the vertical wind shear being too strong for cyclogenesis in that region. There are too few landfalls to make it possible to look at the risk of landfall over the US. Tropical storm tracks are too westward over the Indian Ocean, and the model generates an unrealistically high number of tropical storms in the Central Pacific during the El-Niño year. Those problems may have a negative impact on the skill of the system to predict the risk of landfall. For instance if the tracks in the model are too zonal, then the model may generate more landfalls in the wrong region. This skill will need to be evaluated region by region. However, Figures 39 and 40 suggest that the model displays the right impact of ENSO on the risk of landfall over Mozambique, South China, and Australia at least.

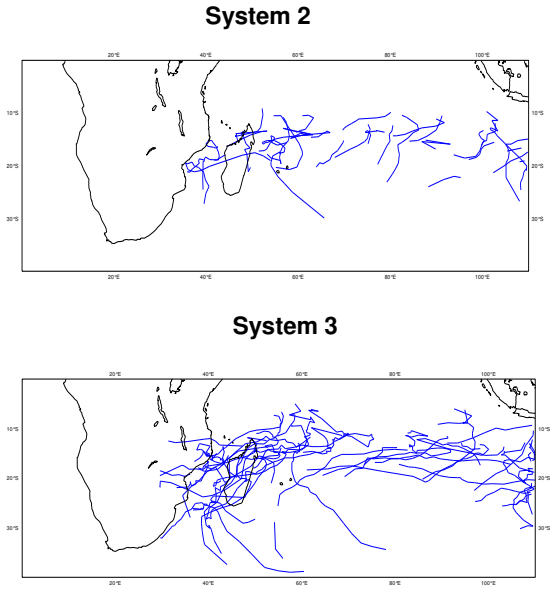



*Figure 38: Model tropical storm tracks in the South Indian Ocean for S2 (top panel) and erwq (bottom panel). Forecasts start on 1st January 2000.*

## 5.3 Examples of probabilistic skill scores

In the previous subsections, the performance of erwq has been discussed in terms of its ability to represent phenomena covering different spatial and temporal scales, from ENSO to tropical cyclones. Potential users of the system may also be interest to have a quantitative assessment of the overall predictive skill of the system for seasonal means of weather parameters such as rainfall and surface air temperaure. For a seasonal prediction system, probabilistic indices are usually preferred as a measure of skill; however, given the small size of the ensemble experiments used for this preliminary assessment (5 members only), such indices may be subject to considerable sampling errors.

A proper evaluation of the skill of probabilistic forecasts produced by S3 will be performed from the full set of back-integrations, covering 25 years with 11-member ensembles. Here, as a preliminary estimate, we show in table 2 a selection of Relative Operative Characteristics (ROC) scores for experiment erwq and S2, for specific anomaly categories in extratropical and tropical regions. In order to highlight potential high-impact events, scores for Europe and North America are shown for below-average 2-m temperature anomalies in (boreal) winter (JFM) and above-average temperature anomalies in summer (JAS), while for the tropical band scores for below-average rainfall are listed for both seasons. As in the previous subsections results from S2 and erwq are compared.

The scores above confirm the indications which emerged from the analysis of individual processes: in general, erwq has more predictive skill than S2 for the tropical regions, while in the northern extratropics improvements are mostly evident during the summer season. An analysis looking at all seasons, different lead times, and both upper and lower terciles and northern hemisphere and tropical regions supports the inferences drawn from those particular examples. The consistency of the tropical results gives some confidence that the tropical improvements are not just an artefact of sampling. The reductions in the northern hemisphere winter scores appears to be fairly consistent but the sample size is small when looking at single seasons. As discussed above, the decrease of skill scores during the boreal winter appears to be related to a partial reduction of the

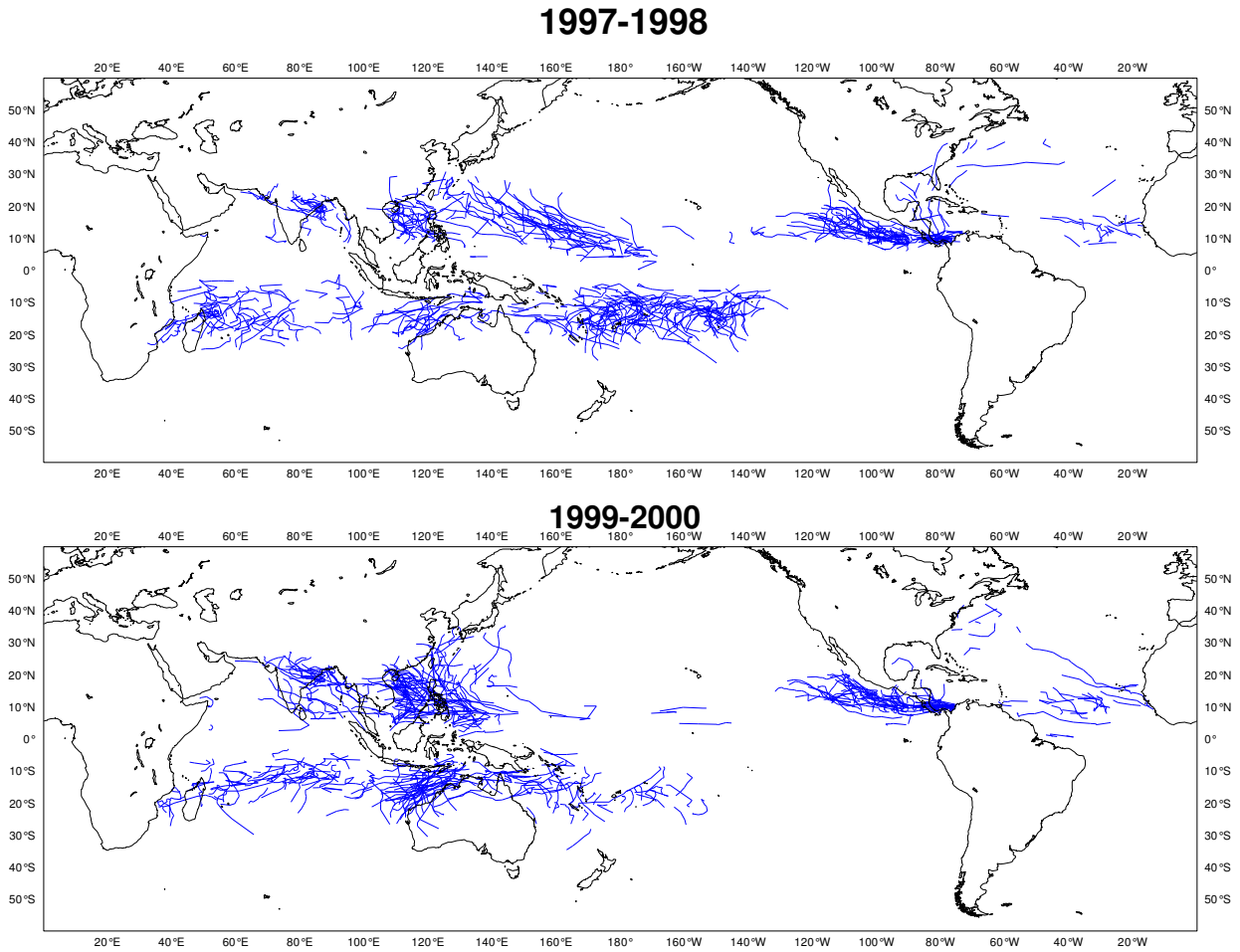



*Figure 39: Model tropical storm tracks for S2 for the all the forecasts (5 members) starting in July, October 1997, January and April 1998 (top panel, El-Niño conditions) and July October 1999 and January and April 2000 (bottom panel, la Nina conditions)*

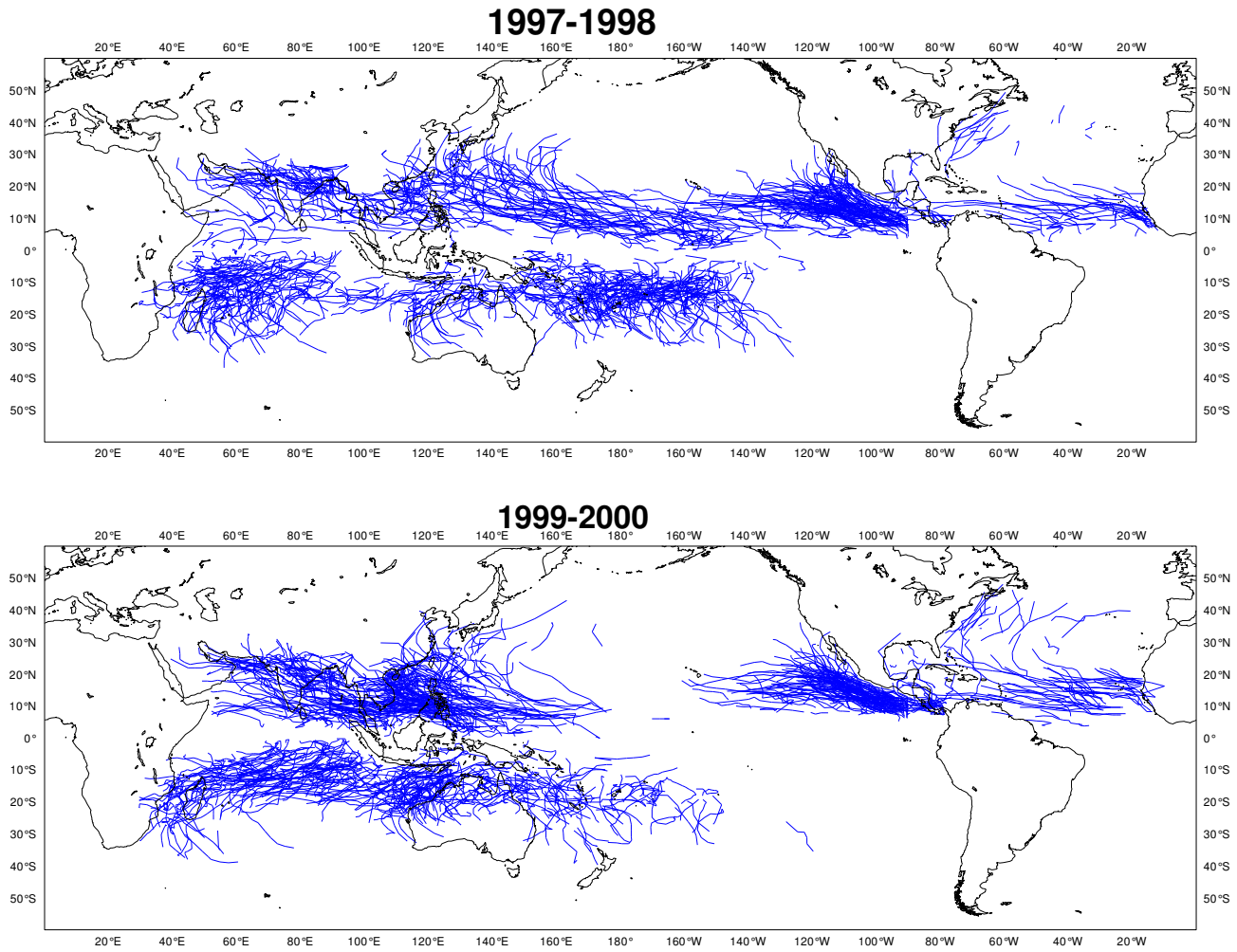



*Figure 40: Same as Figure 11 but for erwq*

| Region and season | Anomaly category | System-2 | erwq |
|---|---|---|---|
| Europe JFM | 2m T lower tercile | 0.08 | 0.06 |
| Europe JAS | 2m T upper tercile | 0.08 | 0.09 |
| N. America JFM | 2m T lower tercile | 0.14 | 0.08 |
| N. America JAS | 2m T upper tercile | 0.06 | 0.11 |
| Tropics JFM | prec.lower tercile | 0.08 | 0.09 |
| Tropics JAS | prec.lower tercile | 0.07 | 0.10 |

*Table 2: Fractional area between actual ROC curve and the zero-skill diagonal line for selected anomaly categories in different regions and seasons, for S2 and erwq experiments. This score varies from 0 for a system with no predictive skill to 0.5 for a perfectly predictive system*

wintertime diabatic heating anomalies in the central tropical Pacific during ENSO episodes, and to an increased level of internal atmospheric variability, both of which decrease the signal-to-noise ratio for NH interannual variability during winter. In S2, such a ratio was enhanced by larger-than observed rainfall amounts in the tropical Pacific, which partially compensated the reduction in the SST anomaly amplitude occurring during the coupled integrations.

# 6 Summary and conclusions

After the installation of S2, work began on the development of S3. This involved upgrading the ocean analysis and testing various atmospheric model cycles as they became available. The ocean analysis was improved in several ways; the 2D OI of S2 was upgraded to 3D OI, the subsurface temperature-only assimilation of S2 was upgraded to a scheme that could assimilate salinity from ARGO, CTD, and moorings as well as the time varying sea level from altimetry. One of the key problems in ocean data assimilation in general is that it mainly acts to correct bias related to the gradient and slope of the thermocline. To deal with this, the analysis of S3 has an explicit bias correction algorithm which allows a better representation of interannual variability than was possible in S2. An extensive ocean reanalysis was carried out covering the period from 1961 until present. The same analysis system is used throughout. The extended analysis can be used for climate studies as well as providing initial conditions for seasonal forecasts. S3 will use the analyses from 1981 as initial conditions for the calibration hindcasts. ENSEMBLES will use the earlier analyses for further seasonal and decadal hindcasts. The impact of the S3 analysis on seasonal forecasts is beneficial nearly everywhere, but especially in the west Pacific. An exception is the equatorial Atlantic, a region where the forecasts are not particularly skillful. (Stockdale et al 2006)

S3 will not only cover a longer hindcast calibration period than S2 (25 years v 15 years), it will have a larger ensemble size for the calibration (11 members v 5 members). Both features should allow better product calibration. The other members of Euro-SIP intend to follow ECMWF and so a multi-model set of hindcasts for the whole period should be possible. S3 will have a longer forecast range- 7 months compared to 6 months as well as a quarterly El Nino outlook to 13 months.

Throughout the development period various atmospheric model cycles were tested as they became available. Progress was not monotonic. Although each cycle improved the medium range forecasts this was not so for the seasonal forecast range. However, in several respects, recent cycles are better than earlier cycles and S3 (erwq) is noticeably better than S2 in terms of predictions of the important El Nino indices.

The latest cycle is not the best in all respects. For example, 31r1 has a poorer representation of the intraseasonal oscillation than 30r2, (though still better than S2). On the other hand, SST predictions in the west Pacific are

substantially better in 31r1 than in 30r2. In the northern hemisphere extratropics the improvements in erwq are mainly seen in the summer season. The ensemble size is only 5 at this stage of assessment, so a proper evaluation for regions where the signal to noise ratio is small must await the full set of calibration forecasts.

Blocking in the northern hemisphere is not well handled in either S2 or erwq, and remains a serious model deficiency which should be given more attention in future developments. Although the MJO is better represented in erwq than in S2, it is still not as well represented as we would like, and continued effort on improving this model deficiency is desirable. Improvements in blocking and the MJO would be beneficial to the extended VAREPS and monthly forecast systems as well as to the seasonal range. The coupled model is now quite well integrated into the ECMWF system making it easier to test model changes on the seasonal (and monthly) range at an earlier stage. The equatorial Atlantic is not very well forecast and improvements in both the model and initial conditions appear to be needed. Model error remains a major source of concern requiring close links between the seasonal, the physics, and the diagnostics groups to fully address this. Mean state correction, not used in S3, can not be ruled out in future though there are no current plans to use it. Initialisation of the coupled model remains an issue; it is not clear that the current strategy of initialising the two components separately is the best approach. We expect S3 to be the last operational system to use the HOPE ocean model. Future work will involve integrating the OPA ocean model, developing variational ocean assimilation and testing future cycles of IFS.

REFERENCES

Anderson, D. and Balmaseda M.A., 2005: Overview of ocean models at ECMWF. Recent developments in numerical methods for atmospheric and ocean modelling, 6-10 September 2004, 103-111.

Balmaseda, M.A., 2004: Ocean data assimilation for seasonal forecasts. ECMWF Seminar Proceedings. Seminar on Recent developments in data assimilation for atmosphere and ocean, 8-12 September 2003, 301-326.

Balmaseda, M.A., 2005: Ocean analysis at ECMWF: from real-time ocean initial conditions to historical ocean reanalysis. ECMWF Newsletter, Aurtumn 2005,24-32.

Balmaseda, M.A., D. Dee, A. Vidard and D. Anderson, 2005: A multivariate treatment of bias for sequential data assimilation: application to the tropical oceans.ECMWF Technical Memorandum No 480. Submitted to QJ.

Balmaseda, M., A. Vidard and D. Anderson, 2007: The ECMWF System 3 ocean analysis system. ECMWF Technical memorandum no. 508.

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

Bonjean F. and G.S.E. Lagerloef, 2002: Diagnostic Model and Analysis of the Surface Currents in the Tropical Pacific Ocean, Journal of Physical Oceanography, Vol. 32, No. 10, pages 2938-2954.

Bryden HL, HR Longworth, SA Cunningham, 2005: Slowing of the Atlantic overturning circulation at 25N, Nature, Vol 438, doi:10.1038/nature0438

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.

Church J. A. and N. White, 2006: A 20th century acceleration in global sea-level rise. GRL, 33, L01602.

Cooper N.S. 1988: The effect of salinity on tropical models. J Phys Ocean 18, 697-707.

Cooper, M.C. and K. Haines, 1996: Data assimilation with water property conservation, J. Geophys. Res 101, C1, 1059-1077.

Dee, D. P.,2006:Bias and data assimilation. To appear in Q. J. R. Meteorol. Soc.

Doblas-Reyes, F.J., M.J. Casado, and M.A. Pastor (2002). Sensitivity of the Northern Hemisphere blocking frequency to the detection index. J. Geophys. Res. D, 107, ACL 6 1-22.

Ducet, N., P.-Y. Le Traon, and G. Reverdin, 2000: Global high resolution mapping of ocean circulation from TOPEX/Poseidon and ERS-1 and -2. J. Geophys. Res., 105, 19477-19498.

Haines K., J. Blower, J-P. Drecourt, C. Liu, A. Vidard, I. Astin, X. Zhou. Salinity Assimilation using S(T) relationships. Mon. Wea. Rev. Vol. 134, No. 3, pages 759–771. 2006

Hurrell, J. W., 1995: Decadal trends in the North Atlantic Oscillation: regional temperatures and precipitation. Science, 269, 676-679.

Ingleby, B and M. Huddleston, 2006. Quality control of ocean temperature and salinity profiles - historical and real-time data. J. Mar. Sys., in press.

Iselin, C.O., 1939. The influence of vertical and lateral turbulence on the characteristics of the waters at mid-depths. Trans. of the Amer. Geophys. Union, 1939, 414-417.

Le Traon, P.-Y., F. Nadal, and N. Ducet, 1998: An improved mapping method of multisatellite altimeter data.

J. Atmos. Oceanic Technol., 15, 522-534.

Parthasarathy B., A.A. Munot, and D.R. Kothwale, 1995: 'Monthly and seasonal rainfall series for all-India homogeneous regions and meteorological subdivisions:1871-1994', Contributions from Indian Institute of Tropical Meteorology, Pune-411 008, INDIA.

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

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

Reynolds R., N Rayner, T Smith, D Stokes, W Wang 2002: An improved in situ and satellite SST analysis for climate. J Clim , 15, 1609-1625.

Roemmich, D., M. Morris, W.R. Young and J.R. Donguy, 1994 Fresh equatorial jets. J. Phys. Oceanogr., 24, 540-558.

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.

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

Stockdale, T.N., M. A. Balmaseda and A. Vidard, 2006: Tropical Atlantic SST prediction with coupled ocean-atmosphere GCMs. *J. Climate* **19**, 6047-6061.

Straus, D. M., and J. Shukla, 2002: Does ENSO force the PNA? J. Climate, 15, 2340 - 2358.

Tibaldi, S. and F. Molteni (1990). On the operational predictability of blocking. Tellus A, 42, 343-365.

Troccoli, A., and K. Haines, 1999: Use of the Temperature-Salinity relation in a data assimilation context, J. Atmos. Ocean Tech., 16, 2011-2025.

Troccoli A., Balmaseda M., Segschneider J., Vialard J., Anderson D.L.T., Haines K., Stockdale T., Vitart F. and Fox A.D. (2002) Salinity adjustments in the presence of temperature data assimilation, Monthly Weather Review, 130, 89-102.

Troccoli A. and Kallberg P. (2004), Precipitation correction in the ERA-40 reanalysis, ERA-40 Project Report Series, 13.

Uppala S.M., P. Kallberg. A simmons and 43 others 2005. The ERA-40 reanalysis. Quart J Royal Met Soc. See also ECMWF Newsletter 101.
http://www.ecmwf.int/publications/newsletters/

Vialard and Delecluse 1998a An OGCM study for the TOGA decade. Part I: Role of the salinity in the Physics of the western Pacific Fresh Pool. J. Phys. Oceanogr., 28, 1071-1088.

Vialard J., F Vitart, M Balmaseda, T Stockdale, D Anderson 2005: An ensemble generation method for seasonal forecasting with an ocean-atmosphere coupled model. Mon Weather Rev., 133, 441-453.

Vidard, A., D.L.T. Anderson, M. Balmaseda, 2006: Impact of ocean observation systems on ocean analysis and seasonal forecasts. Mon. Wea. Rev., in press.