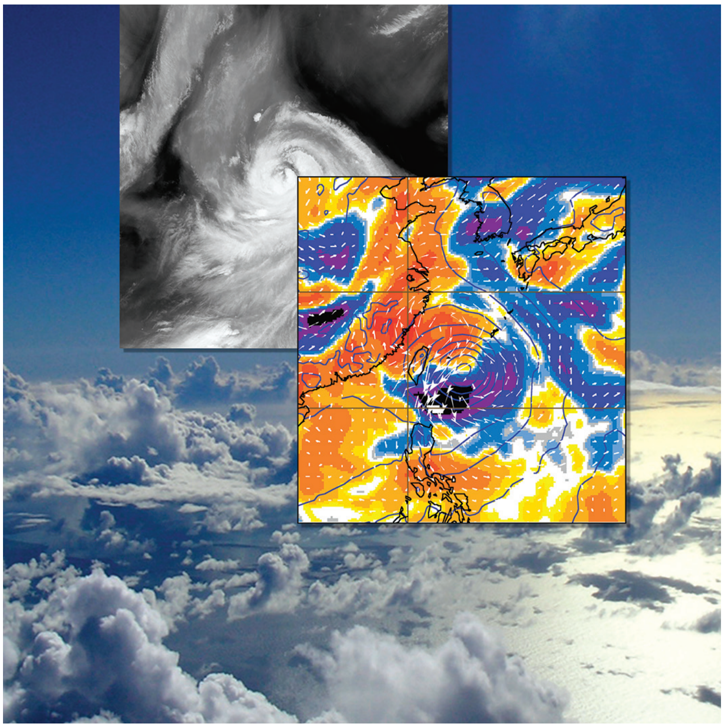
ECMWF Feature article

from Newsletter Number 110 – Winter 2006/07

METEOROLOGY

Seasonal Forecast System 3



www.ecmwf.int/en/about/news-centre/media-resources

doi:10.21957/kc2i5xvfzp

This article appeared in the Meteorolgy section of ECMWF Newsletter No. 110 – Winter 2006/07, pp. 19-25.

Seasonal Forecast System 3

David Anderson, Tim Stockdale, Magdalena Balmaseda,
Laura Ferranti, Frédéric Vitart, Franco Molteni, Francisco Doblas-Reyes,
Kristian Mogensen, Arthur Vidard

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. For all systems so far, 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 became effectively operational in late 1997, and S2 started running in August 2001. During the last few years, work has proceeded with developing System 3 (S3). Major changes have taken place in the ocean analysis system for S3, though not in the ocean model itself. The atmospheric model is cycle 31r1 (Cy31r1). The horizontal resolution has been increased from TL95 to TL159 (with the corresponding grid mesh reduced from 1.875° to 1.125°), and the vertical resolution is increased from 40 levels to 62 levels, extending up to ~5 hPa.

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 ERA-40 for the period 1981 to 2002 and from Operations from 2003 onwards.

For S3 the forecast ensemble generation is not the same as S2: there are changes in the calibration period and size and in the way the ensemble is generated. The real-time ensemble set consists of 41 members in S3, and the calibration set consists of 11 members spanning the 25-year period 1981–2005, so creating a calibration probability distribution function 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 sea surface temperature perturbations to the 5 member ensemble of ocean analyses. Stochastic physics is active throughout the coupled forecast period.

S3 seasonal integrations are 7 months long (rather than the present 6 months). Additionally, once per quarter an 11-member ensemble runs to 13 months, specifically designed to give an “ENSO outlook”. Back integrations have also been made to this range, once per quarter, with a 5-member ensemble.

The data from S3 are archived into the multi-model seasonal forecast streams (MMSF). This gives 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 a real-time multi-model seasonal forecast system. Currently, the participants in EURO-SIP are ECMWF, the Met Office and Météo-France, but other members are expected to join in the future.

S3 is presently running in parallel with S2, and will become the operational ECWMF seasonal forecast system in early 2007. For technical information on the system, including the latest status and how to retrieve data, please see www.ecmwf.int/products/changes/system3.

 doi:10.21957/kc2i5xvfzp

The 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 for analysing climate variability, and by the ENSEMBLES project for seasonal and decadal forecasts.

As for S2, the ocean data assimilation system for S3 is based on HOPE-OI (i.e. the optimum interpolation scheme developed for the Hamburg Ocean Primitive Equation model), but major upgrades have been introduced. In addition to subsurface temperature, the optimum interpolation (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 three-dimensional, the analysis being performed at all levels simultaneously down to 2000 m, whereas in S2, the analysis was carried out on each model level independently and only to 400 m. 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. Due to the large uncertainties in the fresh water flux, the relaxation to climatology is stronger for surface salinity (approximately 3-year time scale), but still weaker than in S2 (approximately 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.

- In S2 the fluxes were from ERA-15 until 1992 and then from the NWP operational analysis.
- In S3 the fluxes were from ERA-40 from January 1959 to June 2002, and then from the NWP operational analysis.

The representation of the upper ocean interannual variability is improved when using the ERA-40 wind stress, although the stresses are biased weak in the equatorial Pacific. The fresh water flux from ERA-40 (precipitation minus evaporation) is known to be inaccurate. A better but by no means perfect estimate was obtained by 'correcting' the ERA-40 precipitation values (*Troccoli & Kållberg,* 2004) as part of the EU ENACT project.

The temperature 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 now better represented. Also assimilating the 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 a warming trend, starting around the mid 1980s. Figure 1 from *Balmaseda et al.* (2007) shows the time evolution of the global sea level from altimeter data and from the S3 analysis. The similarity between the two curves is not surprising, since the global mean sea level trend has been assimilated. Figure 1 also shows the evolution of the global steric height, associated with changes in sea level due to thermal expansion and the time evolution of the global bottom pressure, indicative of changes in the global mass. It can be seen that till 2002 the global trend in global mean sea level (2 mm/year) is mainly due to thermal expansion. In mid 2002 and mid 2004 there are dramatic changes in the mass field as indicated by the bottom pressure. The change in 2002 may be due to the switch in the forcing from ERA-40 to operational analyses, where there is a noticeable increase in the global fresh water flux; the change in mid 2004 is less understood. Observing system experiments are being 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 system on climate variability.

Comparison of the time evolution of meridional transport in the North Atlantic at 30°N 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. This means that sampling is an issue when drawing conclusions about the slowing down of the thermohaline circulation from the sparse observations.

doi:10.21957/kc2i5xvfzp

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 it makes the cold phase of the seasonal cycle more pronounced, and increases the amplitude of the interannual variability.
- ARGO has a large impact on the salinity field on a global scale.

The impact of the S3 analysis on seasonal forecasts is beneficial nearly everywhere, but especially in the west Pacific. A region where there is little impact is the equatorial Atlantic. A fuller description of the ocean analysis system can be found in *Balmaseda et al.* (2007).

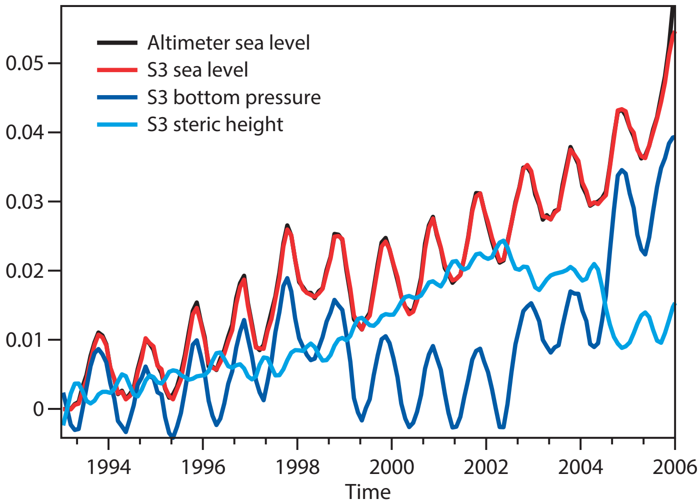


Figure 1 Global trends in sea level and associated quantities. Sea level data from the altimeter (black) and from S3 analysis (red) are almost indistinguishable. S3 steric height (light blue) represents the analysed change due to expansion of water, while the diagnosed equivalent bottom pressure (dark blue) represents changes in mass.

Assessment of forecast skill

At the time of writing, the full set of operational reforecasts for S3 is not yet complete. Results presented here are instead based on an experiment carried out in the ECMWF Research Department, which we will refer to as S3TEST, consisting of a five-member ensemble for four start dates per year, starting in 1987. The starting point for a seasonal forecasting system is its skill in predicting sea surface temperature (SST). Comparing anomaly correlation and rms errors in forecasts of Nino 3.4 and Nino 4 SST in S3TEST with those for S1 and S2 shows clear progress.

Figure 2 shows the rms error for the Nino 4 index, indicating that S3 is considerably better than S1 and S2 in this region, a more difficult region to predict than those to the east (Nino 3 and Nino 3.4). Scatter diagrams, showing all available forecasts for which S2 and S3TEST can be compared, indicate that the improvements in S3TEST 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 SST forecast skill are largely close to neutral, although there is a clear positive benefit in the north subtropical Atlantic.

The climatology of the atmospheric component of S3TEST 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. As an example, Figures 3(a) and 3(b) show systematic errors in 500 hPa height for January-March at a 4–6 month forecast range for S2 and S3TEST. 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.

 doi:10.21957/kc2i5xvfzp

The location of the negative bias over Western Europe in S3TEST 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 S3TEST simulate the maxima of blocking frequency over the Euro-Atlantic and North Pacific regions, but 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. Experiment S3TEST is no better than S2 in this regard, as can be seen in Figure 4.

Figures 3(c) and 3(d) show biases in the msl pressure field for the boreal summer for S2 and S3TEST. 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 S3TEST. 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 S3TEST, both the seasonal mean and the interannual variability of rainfall over the tropical oceans are generally reduced compared to S2 values, bringing the model climatology into closer agreement with observational data from GPCP (Global Precipitation Climatology Project). 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, S3TEST 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–March rainfall in the ensembles run with S2 and S3TEST shows that the S2 variability shows two distinct maxima (with similar amplitude) in the western and eastern tropical Pacific. However, S3TEST simulates a single variability maximum located just west of the dateline, in closer agreement with observations.

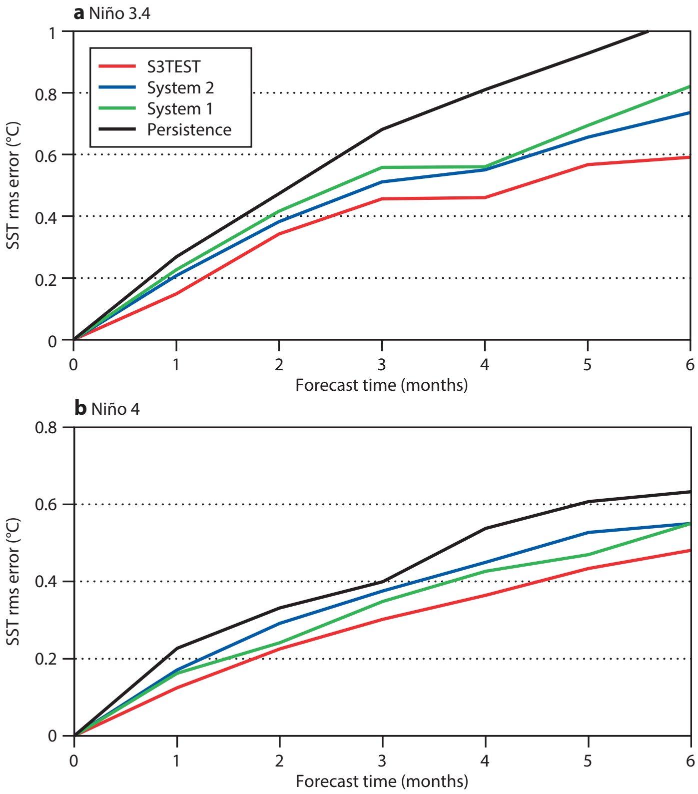


Figure 2 (a) RMS errors for Nino 3.4 SST forecasts from System 1 (green), System 2 (blue) and S3TEST (red), for 64 forecasts in the period 1987-2002. (b) As (a) but for Nino 4. Note that System 2 was worse than the original System 1, but this has more than been made up by S3TEST.

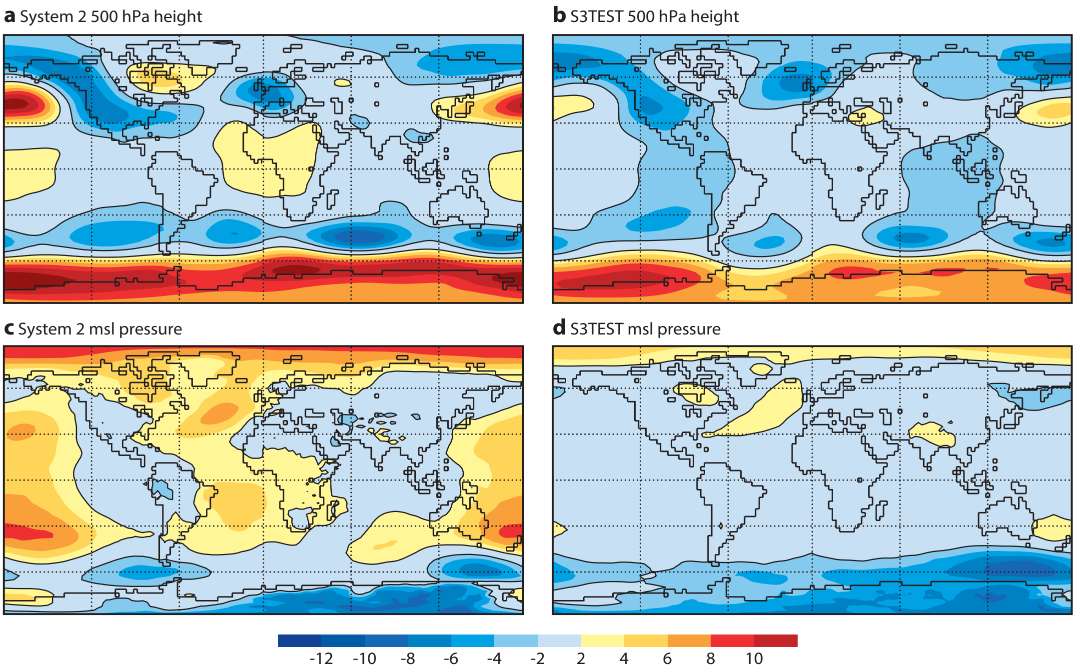


Figure 3 Systematic errors of 500 hPa height (dam) for (a) System 2 and (b) S3TEST for January–March for experiments in the 4-to-6-month forecast range. (c) and (d) As (a) and (b) but for msl pressure (hPa) in July–September.

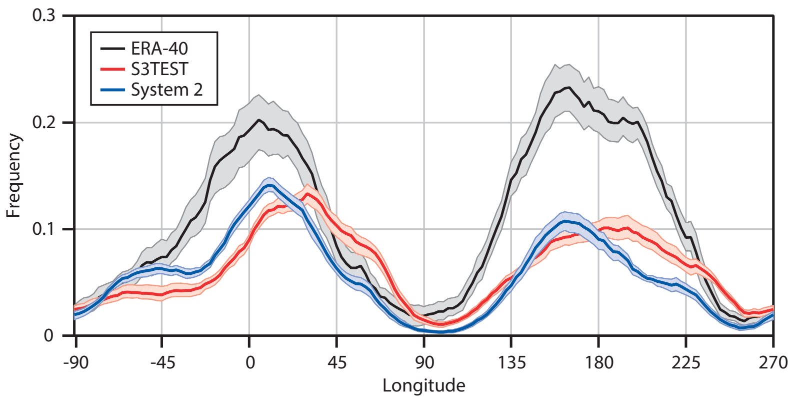


Figure 4 Northern hemisphere winter average blocking frequency for ERA-40 (black line), System 2 (blue line) and S3TEST (red line) for the period 1987–2004. The model index has been computed using five-member ensembles initialized on the 1 October and the results shown are for the season December to February. The shaded areas around each bold line correspond to the 95% confidence interval computed using a bootstrap with a sample size of 500.

Internal atmospheric variability is generally higher in S3TEST 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 frequency range of 20 to 70-days, which includes the Madden-Julian Oscillation (MJO). The standard deviation of tropical velocity potential anomalies at 200 hPa in the October-to-March season is calculated for ERA-40, S2 and S3TEST, 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 re-analysis data, the amplitude is underestimated by both systems. However, in S3TEST the amplitude is considerably closer to ERA-40 than it is in S2. The spectral distribution of the velocity potential variability is further analysed as a function of longitude and oscillation period. As shown in Figure 5, although the S3TEST results represent an improvement with respect to S2 simulations, S3TEST fails to generate a variance maximum in the MJO frequency range as good as that simulated by Cy30r2. Unfortunately, Cy30r2 was not an acceptable cycle: it was never used operationally for medium-range forecasts, and it gave substantially worse forecasts of west equatorial Pacific sea surface temperatures as well as developing unrealistic upper-troposphere moisture distributions.

doi:10.21957/kc2i5xvfzp

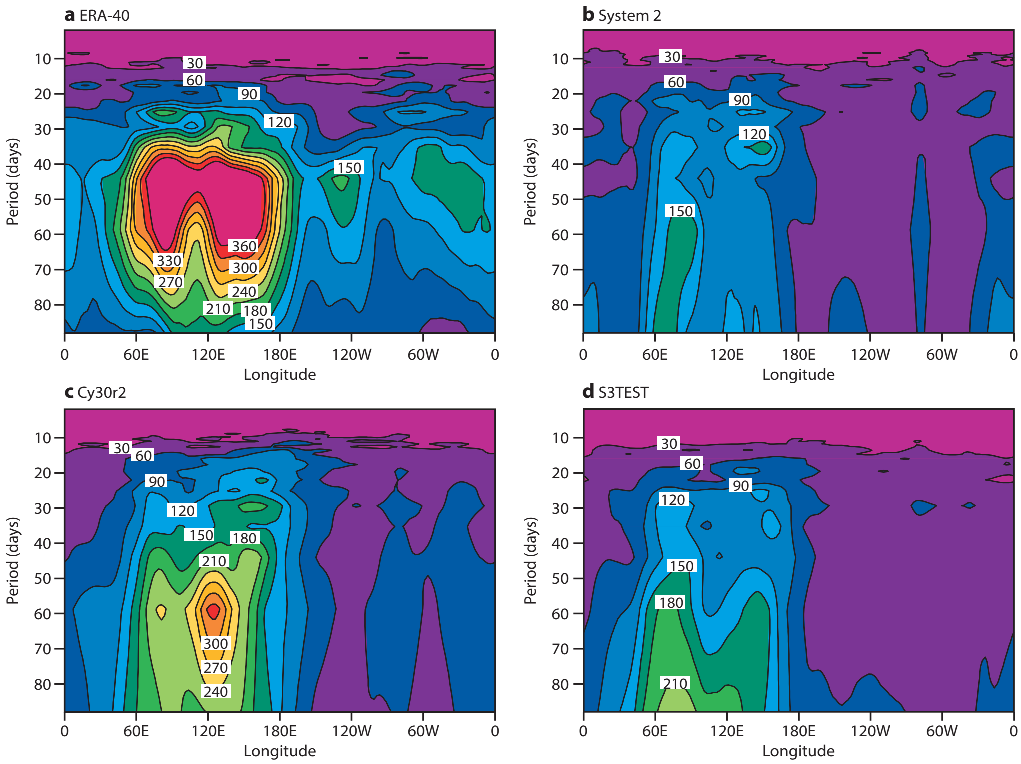


Figure 5 Spectra of 200-hPa velocity potential anomalies in the October–March season as a function of longitude and period for (a) ERA-40 (b) System 2, (c) Cycle 30r2 and (d) S3TEST experiments.

The number of tropical storms detected in S2 and S3TEST has been averaged over the period 1987–2004, the five ensemble members and the four annual starting dates for each individual ocean basin. Comparing the mean annual frequency of tropical storms for each ocean basin with observations shows that S3TEST produces more tropical storms than S2 over all the ocean basins. A possible explanation is the increased horizontal resolution (T159 instead of T95). The increased number of storms is more realistic, although in many areas the number of storms is still underestimated. The tropical storm tracks are clearly much longer and realistic in S3TEST than in S2, as expected from the increased horizontal resolution. Figure 6 shows an example of tracks over the eastern North Pacific and the North Atlantic. These figures show storms that last longer and recurve. This property seems to hold good in other areas also and may indicate that it is feasible to assess the model's ability to predict when tropical storms are more likely to make landfall. Currently only information on frequency and genesis region is made available.

The performance of S3TEST 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 temperature. 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 (five members only), such indices may be subject to considerable sampling errors and so only a preliminary estimate can be given. A variety of scores were examined for different seasons and lead times, but a fair summary of the results can be given based on a subset of the Relative Operative Characteristics (ROC) scores. Specifically, the subset includes scores for:

- Below-average two-metre temperature anomalies in (boreal) winter (January–March) and above-average temperature anomalies in summer (July–September) for Europe and North America.
- Below-average rainfall for both of these seasons for the whole tropical band.

The ROC scores confirm the indications which emerged from the analysis of individual processes: in general, S3TEST has more predictive skill than S2 for the tropical regions, while in the northern extratropics improvements are mostly evident during the summer season. The slight but possibly real decrease of skill scores during the boreal winter may 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 these decrease the signal-to- noise ratio for northern hemisphere 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.

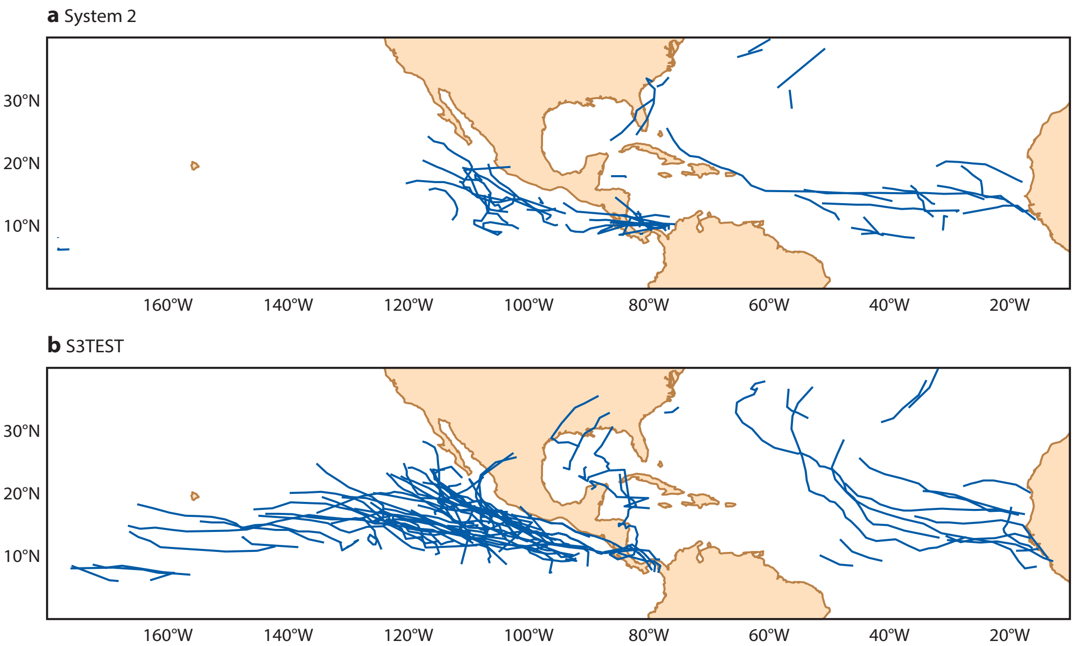


Figure 6 Model tropical storm tracks in the Atlantic and Eastern North Pacific for (a) System 2 and (b) S3TEST. Forecasts start on 1 July 2004.

Summary of System 3 performance

Throughout the extensive development period of System 3 various atmospheric model cycles were tested as they became available. Progress was not monotonic. Although each cycle improved or was at least neutral for the medium-range forecasts this was not so for the seasonal forecast range, where new cycles sometimes led to a significant drop in skill. However, the last few cycles have resulted in strong and significant gains in SST prediction skill, and the model version used in System 3 is the best we have yet seen when assessed by its ability to predict the important El Nino SST variations in the Pacific.

System 3 still has clear deficiencies, however, and there are certain aspects where other model versions show that better performance is possible. For example, Cy30r2 had a better representation of the intraseasonal oscillation than Cy31r1, despite its other failings. In the northern hemisphere extratropics the improvements in S3TEST are mainly seen in the summer season, and it is possible that northern hemisphere winter mid-latitude forecast skill is marginally worse than S2. The ensemble size is only five at this stage of assessment, and only a subset of dates have been considered, so a proper evaluation of System 3 for regions where the signal-to-noise ratio is small must await the full set of calibration forecasts. The skill scores will be published on the ECMWF web pages when they become available.

Blocking in the northern hemisphere is not well handled in either S2 or S3TEST, and remains a serious model deficiency which should be given more attention in future developments. Although the MJO is better represented in S3TEST 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.

Several other major features of System 3 should be highlighted. The ocean analysis/reanalysis is a major product in its own right. The increased ensemble size and especially the larger set of back integrations (25 years rather than 15 years) increases the accuracy of the forecast products. This is a big step forward for those wishing to process the model output themselves to create tailored seasonal forecast products. The new experimental ENSO outlook forecasts extending to 13 months give a longer-range outlook on one of the major factors that drives seasonal climate anomalies.

There is still scope for substantial improvements in the future, but we hope that System 3 will be a useful step on the road to developing numerical systems that fully exploit the predictability that exists on seasonal timescales.

doi:10.21957/kc2i5xvfzp

Further Reading

Anderson, D., T. Stockdale, M. Balmaseda, L. Ferranti, F. Vitart, F. Molteni, F. Doblas-Reyes, K. Mogensen & **A. Vidard,** 2006: Seasonal Forecasting System 3. *ECMWF Technical Memo No. 503.*

Balmaseda, M., A. Vidard & **D. Anderson,** 2007: The ECMWF System 3 ocean analysis system. *ECMWF Technical Memo No. 508.*

Troccoli, A. & **P. Kållberg,** 2004: Precipitation correction in the ERA-40 reanalysis. *ERA-40 Project Report Series No. 13.*

© Copyright 2016

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

The content of this Newsletter article is available for use under a Creative Commons Attribution-Non-Commercial-No-Derivatives-4.0-Unported Licence. See the terms at https://creativecommons.org/licenses/by-nc-nd/4.0/.

The information within this publication is given in good faith and considered to be true, but ECMWF accepts no liability for error or omission or for loss or damage arising from its use.

doi:10.21957/kc2i5xvfzp