System 4 user guide

Version 1.1

Date: Nov. 2011

Contents

1. Introduction to seasonal forecasting

1.1 The basis of seasonal forecasting

Seasonal forecasting is the attempt to provide useful information about the "climate" that can be expected in the coming months. The seasonal forecast is not a weather forecast: weather can be considered as a snapshot of continually changing atmospheric conditions, whereas climate is better considered as the statistical summary of the weather events occurring in a given season.

Numerical Weather Prediction (NWP) provides useful information for up to approximately 10 days in the future. It is based on solving a complex set of hydrodynamic equations that describe the evolution of the atmosphere, subject to the initial atmospheric state and initial conditions at the Earth's surface. Since the initial state is not known perfectly, all forecasts begin with estimates. Unfortunately the system is very sensitive to small changes in the initial conditions (it is a chaotic system) and this limits the ability to forecast the weather beyond 10-14 days.

Despite the chaotic nature of the atmosphere, long term predictions are possible to some degree thanks to a number of components which themselves show variations on long time scales (seasons and years) and, to a certain extent, are predictable. The most important of these components is the ENSO (El Niño Southern Oscillation) cycle, which refers to the coherent, large-scale fluctuation of ocean temperatures, rainfall, atmospheric circulation, vertical motion and air pressure across the tropical Pacific. It is a coupled ocean-atmosphere phenomenon centered over the tropical Pacific but the scale of the fluctuations is quite vast, with the changes in sea-surface temperatures (SSTs) often affecting not just the whole width of the Pacific but the other ocean basins too, and the changes in tropical rainfall and winds spanning a distance of more than one-half the circumference of the earth. El Niño episodes (also called Pacific warm episodes) and La Niña episodes (also called Pacific cold episodes) represent opposite extremes of the ENSO cycle. The ENSO cycle is the largest known source of year-to-year climate variability.

Changes in Pacific sea surface temperature (SST) are not the only cause of predictable changes in the weather patterns. There are other causes of seasonal climate variability. Unusually warm or cold sea surface temperatures in the tropical Atlantic or Indian ocean can cause major shifts in seasonal climate in nearby continents. For example, the sea surface temperature in the western Indian Ocean has a strong effect on the precipitation in tropical eastern Africa, and ocean conditions in the tropical Atlantic affect rainfall in northeast Brazil. In addition to the tropical oceans, other factors that may influence seasonal climate are snow cover and soil wetness. When snow cover is above average for a given season and region, it has a greater cooling influence on the air than usual. Soil wetness, which comes into play most strongly during warm seasons, also has a cooling influence. All these factors affecting the atmospheric circulation constitute the basis of long-term predictions.

To summarize, seasonal forecasts provide a range of possible climate changes that are likely to occur in the season ahead. It is important to bear in mind that, because of the chaotic nature of the atmospheric circulation, it is not possible to predict the daily weather variations at a specific location months in advance. It is not even possible to predict exactly the average weather, such as the average temperature for a given month.

1.2 Statistical and dynamical approaches

The starting point for seasonal forecasting is a good knowledge of climate, that is, the range of weather that can be expected at a particular place at a particular time of year. Beyond a simple knowledge of climatology, statistical analysis of past weather and climate can be a valid basis for long-term predictions. There are some regions of the world and some seasons when statistical predictions are quite successful: an example is the connection between the rainfall in March-May in the Nordeste region of Brazil and the sea surface temperatures in the tropical Atlantic in the months before and during the rainy season.

In theory a very long and accurate record of the earth's climate could reveal the combined (and non-linear) influences of various factors on the weather, and analysis of many past events could average out the unpredictable parts. In practice the 50-100 year records typically available represent a very incomplete estimate of earth's climate. In addition, seasonal predictions based on past climate cannot take full account of anthropogenic or other long term changes in the earth's system, such as the potential impact of global warming.

An alternative approach is to use the numerical weather prediction method by solving the complex set of hydrodynamic equations that describe the evolution of the Earth's climate system. For a seasonal forecast it is important to consider both the atmospheric and oceanic components of the Earth's system. In fact, the air-sea interaction processes that describe the complicated interchange between the atmosphere and ocean are essential to represent the ENSO cycle. Just as for synoptic range NWP forecasts, the calculation depends critically on the initial state of the climate system, particularly the tropical Pacific ocean for ENSO. Because of the chaotic nature of the atmosphere, a large number of separate simulations are made. They will all give different answers as regards the details of the weather, but they will enable something to be said about the range of possible outcomes, and the probabilities of occurrence of different weather events.

If the numerical models were very realistic, and if very large ensembles of such calculations could be performed, then the "climate" (i.e. the probability distribution of weather) to be expected in the coming months would be accurately described. To the extent that predicted "climate" differs from normal because of the initial conditions of the ocean/atmosphere/land-surface, the ensemble calculations could predict the correct seasonal forecast "signal". Unfortunately there are a number of problems that limit the seasonal forecast skill. Numerical models of the ocean and atmosphere are affected by errors, not all aspects of the initial state are well observed, and techniques for estimating the extra uncertainty that this introduces are still incomplete.

1.3 How reliable are today's seasonal forecasts?

The principal aim of seasonal forecasting is to predict the range of values which is most likely to occur during the next season. In some parts of the world, and in some circumstances, it may be possible to give a relatively narrow range within which weather values are expected to occur. Such a forecast can easily be understood and acted upon; some of the forecasts associated with strong El Niño events fall into this category. More typically, the probable ranges of the weather differ only slightly from year to year. Forecasts of these modest shifts might be useful for some but not all users.

The benefits of seasonal forecasting are most easily established in forecasts for some areas of the tropics. This is because many tropical areas have a moderate amount of predictable signal, whereas in the mid-latitudes random weather fluctuations are usually larger than the predictable component of the weather. The point at which seasonal forecasts become good enough to be useful to a particular user will depend on the user's requirements. In some cases, today's systems are already useful, although care should always be taken to interpret model outputs appropriately. As reliability continues to improve, a wider range of applications should become possible, and the value of seasonal forecasts will further increase. More work is still needed to relate probabilities of large-scale weather patterns to detailed impacts and applications. It must be remembered, however, that there are tight limits on what it is physically possible to achieve with a seasonal forecast system. It will only ever be possible to predict a range of likely outcomes. In many cases this range will be relatively large, and there will always be a risk of something unexpected happening. In many parts of the world, most of the variability in the weather will remain unpredictable.

Some seasonal forecasts available today are issued with probabilities (or error bars) which have been properly calibrated against past cases. An example is the Canonical Correlation Analysis (CCA) prediction of El Niño variability, which is regularly shown in the NOAA Climate Diagnostics Bulletin. Such forecasts are probably fairly reliable, but they have very wide error bars: they may state that in 6 months time there might be strong El Niño conditions, or fairly strong La Niña conditions, or anything in between. The outputs of seasonal forecast models generally have less spread but are also less reliable.

A proper calibration of a forecast system against data is not always easy to do. This is primarily because of the limited availability of past data. The problem is especially severe when the level of predictability is low so that many years of data are needed. Relatively low predictability on the seasonal time scale is a feature of much of the globe, but especially in mid- latitudes, and for smaller spatial scales (several hundred km, rather than several thousand). At the moment the ECMWF seasonal forecasts are not issued with calibrated probabilities. However, information about the reliability seen in past performance is available, in plots displayed together with the forecast products. The limited number of past forecasts means that we can only give a rough estimate of the reliability, particularly for smaller regions or local values. It is clear that the direct model output is still quite some distance from being perfectly reliable, although the level of reliability is improving.

2. The ECMWF Seasonal forecast system

The system consists of an ocean analysis to estimate the initial state of the ocean, a global coupled ocean-atmosphere general circulation model to calculate the evolution of the ocean and atmosphere, and a post-processing suite to create forecast products from the raw numerical output. Detailed descriptions of the models and the post-processing are given below.

2.1 Ocean model

The ocean model used is NEMO, which replaces the HOPE ocean model used in earlier ECMWF seasonal forecast systems. NEMO is used in the ORCA1 configuration, which has a 1x1 degree resolution in mid-latitudes and enhanced meridional resolution near the equator.

2.2 Ocean analysis system

The NEMOVAR ocean analysis system is used to prepare ocean initial conditions for the seasonal forecasts. These ocean analyses use all available *in situ* temperature and salinity data, an estimate of the surface forcing from ECMWF short range atmospheric forecasts, sea surface temperature analyses and satellite altimetry measurements to create our best estimate of the initial state of the ocean. In order to sample some of the uncertainty in our knowledge of the ocean state, a 5-member ensemble analysis is created using perturbed versions of the wind forcing.

Prior to starting our coupled model forecasts, the ocean analyses are further perturbed by adding estimates of the uncertainty in the sea surface temperature to the ocean initial conditions. Thus all 51 members of the ensemble forecast have different ocean initial conditions.

The ocean analyses are an important product in their own right. The "Ocean Analysis" web pages contain the documentation of our ocean analysis system. This will be updated soon to describe the NEMOVAR system. Starting with System 4, the seasonal forecasts use the near-real-time ocean analyses, which enables us to produce the forecasts earlier in the month.

2.3 Atmospheric model and coupling

The atmospheric component of the coupled model is the ECMWF IFS (Integrated Forecast System) model version 36r4. This model version was introduced for medium-range forecasting on 9th November 2010, although for seasonal forecasts we use a lower resolution. We take 91 levels in the vertical, with a model top in the mesosphere at 0.01 hPa, or a height of approximately 74 km. The spectral horizontal resolution used for seasonal forecasts is TL255. The spectral representation is used only for the dynamical part of the model calculations. All of the model physical parameterization (including clouds, rain and the land surface) are calculated on a reduced N128 gaussian grid, which corresponds to a 0.7 degrees spacing. The atmospheric model uses a two time-level semi-Lagrangian scheme for its dynamics with a 45 minute time step.

The IFS is used with a few non-standard settings. The new FLAKE lake model is activated to provide temperature and ice data for resolved lakes, but is run only in climatological mode. Some adjustments are made to stratospheric physics. The overall amplitude of the non-orographic gravity wave drag is reduced, to give a better evolution of the QBO and a better stratospheric climate. A higher level of non-orographic wave drag is imposed at high southern latitudes, which is believed to compensate for numerical damping of highly active resolved gravity waves at these latitudes. The non-conservative action of a gravity wave drag limiter is reduced to improve the realism of the model physics. Ozone is activated as a prognostic variable, and unlike the medium-range forecasts, ozone is radiatively active. As previously, we specify time-variation of greenhouse gases, using an IPCC scenario calculation for future values. We also specify a time-varying solar cycle, following recommendations for IPCC AR5.

We also allow for stratospheric volcanic aerosol within the forecast system. Only very approximate values are specified - 3 numbers giving NH, tropical and SH amounts, together with assumed vertical profiles. Values are specified using data from the month before the forecast starts, and then damped persistence applies during the forecast. Thus major eruptions are not captured in advance (!) but the after affects can be accounted for to some extent in the forecasts. It would be preferable to have a better characterization of volcanic aerosol distribution and properties, and eventually real-time analysis systems should be able to provide such information.

For the time being, however, we specify data in the re-forecasts with a similar level of accuracy we think we can achieve in real-time.

Stochastic physics are used, both the SPPT3 scheme and stochastic backscatter. The settings are identical to those used in the medium-range EPS. Note that the SPPT3 scheme in particular is efficient at exciting a divergence in the ENSO SSTs of the coupled model forecast - the spread in ENSO forecasts from System 4 is substantially larger than in System 3.

There is no sea-ice model. Previously we have specified a long-term sea-ice climatology, but this is no longer tenable for the real-time forecasts, given the large reductions in Arctic sea-ice extent in recent years. Instead, for the forecast for a given year, we specify sea-ice by sampling from the previous 5 years. This both captures the main part of the trend in sea-ice, and also gives a representation of the uncertainty in sea-ice conditions. All integrations also use a feasible ice field that contains appropriately sharp boundaries to the sea-ice (rather than a much more smoothly varying multi-year mean). As before, sea-ice for the first 10 days of the forecast persists the initial sea-ice analysis; then over the next 20 days there is a transition towards the specified ice conditions from the previous 5 years.

The atmosphere and ocean are coupled using a version of the OASIS3 coupler developed at CERFACS. This is used to interpolate between oceanic and atmospheric grids with a coupling interval of 3 hours, which allows some resolution of the diurnal cycle. A gaussian method is used for interpolation in both directions, primarily due to the complexity of the ORCA1 grid. The gaussian method automatically accounts for the inevitably different coastlines of the two models - values at land points are never used in the coupling, since these can be physically very different to conditions over water.

2.4 Initial conditions, re-forecasts and forecasts

The atmospheric initial conditions come from ERA Interim for the period 1981 to 2010 and from ECMWF operations from 1st January 2011 onwards.

Initial conditions for the land surface are treated specially. For the re-forecast period, the HTESSEL land surface model used in Cy36r4 is run in offline mode, with forcing data (precipitation, solar radiation, near surface temperature, winds and humidity) coming from ERA interim. However, the ERA-interim precipitation is scaled for each grid point to match the monthly mean totals from GPCP data, for the years where GPCP data is available (up to mid-2008). A mean scaling is also calculated for each calendar month, and used to adjust the ERA interim precipitation data at the end of the re-forecast period when GPCP data is not available. It has been shown that forcing the HTESSEL model in this way produces good initial data for soil moisture, at least in well observed areas (where the GPCP data are reliable). The snow cover produced in this way also seems to be largely reasonable. The HTESSEL run is made at T255 resolution, which matches both the ERA interim forcing and the resolution on which we need the surface initial conditions.

For the real-time forecasts, ERA interim is not available. Thus from 1 January 2011 onwards, the land surface initial conditions are taken from the ECMWF operational analyses. Since the present ECMWF model uses the HTESSEL model and has a recently re-tuned land surface assimilation system, this is also believed to produce good quality analyses for soil moisture and snow cover, at least in areas with sufficient observations. Thus, the land surface conditions of forecasts and re-forecasts should be quite well matched in well observed areas. However, this is not guaranteed to be the case everywhere. The real-time analyses must be interpolated from T1279 down to T255. This can cause particular problems with glaciers, as was the case with System 3. For System 4, the scripts have been changed to remove glaciers from the T1279 analysis before interpolation; the appropriate glaciers for T255 are then added to the snow field once the interpolation is complete. A final safety check is applied to prevent real-time land surface initial conditions straying too far from those used in the re-forecasts, which might otherwise occur in mountainous regions and/or poorly observed areas. Limit fields are defined for each surface variable and for each calendar month. The limit fields define the maximum and minimum permitted values of the field in the initial conditions of the real-time forecast. The limits are defined as the maximum and minimum values observed at that point and calendar date for the 30 year re-forecast period, plus an extra small margin specified as a global constant for each field. This margin is generally chosen to correspond to a 50-year return period "event" in areas with high variability; in areas with low variability, the permitted value would be a more extreme event. The limit for snow depth is calculated slightly differently: it is the previously observed range plus or minus 1cm of water equivalent. In particular, this allows a modest covering of snow in areas where snow was not seen in the previous 30 years. Overall these limits allow real-time initial conditions to cover a wide range of values, including extremes somewhat beyond those observed in the 30 year re-forecast period, but still prevent any physically unreasonable anomalies being specified.

The model has radiatively interactive ozone, and needs ozone initial conditions. Unfortunately, the interannual variability of ozone in ERA interim is largely artificial, driven by changes in satellite instruments. Since these spurious changes were found to drive substantial temperature errors in the stratosphere, they cannot be used as initial conditions. Instead, a seasonally varying climatology is formed from what are believed to be the best years of the ERA interim ozone analyses (1996-2002), and the ozone initial conditions are taken from this. Although we do not provide any initial data on ozone anomalies, the ozone field is free to develop during the forecast and will develop anomalies physically consistent with e.g. temperature anomalies and specified CFC time history/projection.

The seasonal forecasts consist of a 51 member ensemble. The ensemble is constructed by combining the 5-member ensemble ocean analysis with SST perturbations and the activation of stochastic physics. The forecasts have an initial date of the 1st of each month, and run for 7 months. Forecast data and products are released at 12Z UTC on a specific day of the month. For System 4, this is expected to be the 7th.

Every seasonal forecast model suffers from bias - the climate of the model forecasts differs to a greater or lesser extent from the observed climate. Since shifts in predicted seasonal climate are often small, this bias needs to be taken into account, and must be estimated from a previous set of model integrations. Also, it is vital that users know the skill of a seasonal forecasting system if they are to make proper use of it, and again this requires a set of forecasts from earlier dates.

A set of re-forecasts (otherwise known as hindcasts or back integrations) are thus made starting on the 1st of every month for the years 1981-2010. They are identical to the real-time forecasts in every way, except that the ensemble size is only 15 rather than 51. The ensemble is again constructed by adding SST perturbations to the 5-member ocean analysis. The data from these forecasts is available to users of the real-time forecast data, to allow them to calibrate their own real-time forecast products using their own techniques. Update (July 2013): For start dates on the 1st February, May, August and November, the re-forecast ensemble size has been extended to 51 members, to allow a better assessment of skill of the system. These additional ensemble members were created in 2013 and are not considered part of the operational system as such. They are archived in MARS together with the first 15 members, and are available for use in studying the performance of System 4.

In addition to the seasonal forecast which is made every month, an annual-range forecast is made four times per year, with start dates the 1st February, 1st May, 1st August and 1st November. The range of the forecast is 13 months. The annual range forecasts are run as an extension of the seasonal forecasts, and are made using the same model but with a smaller ensemble size. Both re-forecasts and real-time forecasts and have an ensemble size of 15. The annual range forecasts are designed primarily to give an outlook for El Niño. At present they have an experimental rather than operational status.

2.5 Post-processing and product generation system

Seasonal mean climate anomalies are usually relatively small, for example temperature anomalies are often less than 1 °C. Since model biases are typically of a similar magnitude, some form of post-processing to remove model bias is needed. Two different methods are used, but in both cases the a posteriori correction is based on the assumption of a quasi-linear behavior of the atmosphere and ocean anomalies.

The post-processing system is designed to correct for mean biases in the forecast system, but in general it does not do any more than this. For example, the spatial plots of ensemble mean forecasts are not normalized to match observed variance, although when probabilities of percentiles of the climatological distribution being exceeded are given, there is an implicit mapping of the amplitude of variability in the model to the amplitude of variability in observations. Importantly, probability forecasts are not calibrated according to past forecast skill. One reason that this level of post-processing is not done is that the sample sizes are often not large enough to define such calibrations satisfactorily. Nonetheless, research on more fully calibrated products is ongoing, and experimental calibrated products may become available for certain fields.

For Niño index values (e.g. Niño 3), the mean bias of the model is estimated as a function of lead-time and calendar month from the difference between model hindcast values and observations for a set of reference years. This bias is then used to correct the model output and produce an absolute SST forecast, for example 26.1 °C. For System 4, and only for the Niño plume plots, an additional step of normalizing the forecast variance to match the observed variance is made, prior to the calculation of the absolute SST forecast value. To issue a forecast anomaly, this absolute value is then referenced against a specified climatology. At the moment we use the NCEP OIv2 climate for the 1981-2010 base period. Note that the choice of the climate base period and the choice of reference years for estimating the model bias are independent (although at present they are the same), and that this approach requires a high quality observational dataset.

For all other predicted variables, biases are removed from consideration by ignoring the true mean value of the field and considering only model anomalies with respect to the model mean state. Specifically, the values of the forecast ensemble are compared to the values of a climate reference ensemble (made up of model re-forecasts with the same lead time and calendar month, and covering a representative set of years), and the differences between model forecast and model climate are assessed and plotted. The advantage of this approach is that it is independent of observational datasets. The disadvantage is the lack of choice concerning the climate base period. Since the System 4 hindcast calibration period corresponds to a standard 30 year period (1981-2010), it is hoped that for most applications this will not be a problem.

Whichever way the forecast biases are accounted for, there will be some inaccuracy in the estimate of bias and the definition of climate. In the case of the Niño indices, the length of the reference period (30 years) and the average accuracy of the forecasts (typically 0.4 °C) determine the sampling accuracy of the mean drift estimate. If we can assume that the errors are uncorrelated, then we find that the uncertainty in the bias correction is no more than about 0.1 °C, and is thus a small contributor to the overall error in a forecast. If the forecast errors are correlated, that is, there are low frequency changes in the error characteristics of the model, then uncertainty in the bias could be larger.

For the model fields whose anomalies are shown with respect to a climatology of model forecasts, a specific issue arises if the user wants to reference the forecast to a different base period. For example, a climatology based on forecasts in the 1981-2010 period might be expected to be systematically different from a climatology based on forecasts in the 1971-2000 period, if there have been real low frequency changes in the climate system between these periods. The tricky issue is to relate differences in model climate between different periods to differences in observed climate. This is also an important issue when combining forecasts from different systems which have different base periods.

Suppose we know the model climate for the 1981-2010 period, based on a total of 450 integrations (a 15-member ensemble for 30 years). Suppose we also know that for the observed variable of interest, the 1981-2010 period was, for example, 0.2 °C warmer than the 1971-2000 period. If the forecast for the coming season shows a slight warming of 0.1 °C in the ensemble mean compared to the model climatology for 1981-2010, how do we interpret this if we are asked to produce a forecast relative to the 1971-2000 period? Do we allow for the 0.2 difference in observations between the two periods, and predict a 0.1° cooling in the ensemble mean, or do we simply insist that the model gives a 0.1 °C warming? (Of course the spread of the ensemble will in any case be bigger than this difference, but the choice will certainly affect the probability distribution).

For fields which are close to being deterministic, i.e. whose value consists of a large seasonally predictable signal and a small amount of unpredictable noise, then it is reasonable to suppose that the difference in the observed climate between the periods is due to a real change in the system, and is not just an artifact of sampling. If we further assume that the model is capable of reproducing the observed low-frequency variability (which is not certain), then we would expect the model climate for 1971-2000 to be shifted relative to the model climate for 1981-2010 by the same amount as the observations, and so we can apply the correction of 0.2 °C to adjust the base period of the forecast.

However, in other cases the level of seasonal noise may be large enough that the difference in observed values between the two periods may be largely due to chance. That is, the difference between observed temperatures in 1971-2000 and 1981-2010 might be due to chance variations. If the differences in the climate were largely unpredictable, then it might be more appropriate to take the (450 member) model climate as the best estimator of what the model would have produced for the specified thirty-year period. In this case we might not make a correction to allow for the different base period.

Temperature is a field where it is clear that there are substantial trends to warmer values over recent decades, and this is reproduced in the seasonal forecast system and needs to be accounted for in some way when considering different base periods. Nonetheless, the proper calibration of low frequency (decadal or longer) variability within the ECMWF system is still not fully resolved. For noisy fields without strong trends, such as precipitation, adjusting for base period differences is likely to be inappropriate.

2.6 Known issues

We list here known issues with System 4. These are additional to the general limits of seasonal forecast systems and the limits inherent in the specific design of System 4, which are both discussed in the general documentation.

2.6.1 Retrieval of data from System 4 to low-resolution grids

Users are reminded that System 4 uses a higher resolution model than System 3. This has consequences when retrieving data using using MARS to interpolate to a coarse resolution grid, such as 2.5 x 2.5°. MARS interpolation from a fine (0.7°) grid to a coarse (e.g. 2.5°) grid can introduce aliasing of smaller scales, leading to a slightly noisy field. Users are advised instead to retrieve data either on the original grid, or on a regular lat-long grid closer to the original model resolution, such as 0.75 x 0.75° (available via dissemination) or 1 x 1°, and then average the data to the desired low resolution grid in an appropriate way.

2.6.2 Issues which can affect quality of forecasts

2.6.2.1 Sea ice data problem and amelioration

Summary: sea-ice trends on a limited number of coastal gridpoints are not properly represented in the real-time forecasts. Skill estimates from re-forecasts are also affected.

System 4 does not have a numerical model of sea-ice. Instead, System 4 forecasts and re-forecasts use ice data from the 5 years previous to the forecast or re-forecast date, with one fifth of the ensemble members using the daily evolution of ice cover from each year. Unfortunately, the dataset used for the daily ice cover has errors in a number of grid points near coastlines, such that about 75% of the re-forecasts have ice cover that is too low on these data points. The artificially low ice cover in winter can have a large local impact on surface fluxes and 2 metre temperature. Initial tests showed that real-time forecasts, which because of the years they sample do not have the low ice cover problem, would produce much colder temperatures than the re-forecasts at these points, giving an artificial "cold" signal.

To prevent this problem, a special treatment of the affected points was introduced for the real-time forecasts (more precisely, all forecasts from 1 Jan 2011 onwards). For these forecasts, and only for the affected points, the specified ice cover is taken from a 5 year sample representing the whole of the re-forecast period, rather than the last 5 years (the years 2006, 2000,1994,1988 and 1982 are used). At these selected coastal points, the specified sea-ice approximately represents the long-term climate, both in terms of mean and terciles, so the forecasts have similar ice conditions to the re-forecasts, with no trend to lower values. The large majority of sea-ice points continue to use data from the last 5 years, and thus capture recent trends in sea-ice extent.

Although the impact of this problem on the real-time forecasts is very small, there is expected to be a more substantial negative impact on the skill estimates derived from the re-forecasts, since the re-forecasts are a mixture of some years with the problem and some years without. The exepcted skill of the real-time forecasts, which has the effect of the error neutralized, may thus be better than the skill estimated from the re-forecasts in these regions.

2.6.2.2 Inaccurate absolute values of daily values of T_{max}/T_{min} after restart points

Summary: Absolute values of T_{max}, T_{min}, and windgusts for the previous 24h reported on the 2nd day of every second month of forecast lead-time, are slightly biased. Anomalies are not affected.

The IFS calculates Tmax, Tmin and wind gusts as maximum (or minimum) values over the preceding 24 hours. Due to a bug in the IFS, the re-setting of the arrays storing the maximum values, which should happen every 24 hours, does not occur as it should during a binary restart of the model. A consequence is that the maximum value reported at 0Z on the 2nd day of some months contains the maximum value over the previous 48h instead of the previous 24h. Half of the time this will be correct (i.e. the maximum occurred on the 2nd day of the averaging period anyway), but half of the time it will be biased high (if the maximum occurs in the first half of the 48h period, the 48h maximum will be higher than the maximum of the last 24h). In System 4, binary restarts are used every 2 months, so the problem is visible every second month. Binary restarts are supposed to be completely transparent, and indeed the forecast values are bit-identical whether or not restarts are used. However, due to this bug, there is a difference in the output of these three fields. The re-forecasts and real-time forecasts have the same restart points, so the anomalies in these fields remain unbiased. If the absolute values of the daily time series of these fields is important, then users should be aware of the issue.

2.6.3 Technical issues which do not affect forecast quality

2.6.3.1 Binary identical reproduction of 1991-2010 singular vector computations

Summary: We normally ensure all calculations are binary reproducible - any part of our system can be re-run and the answers will always be identical to the last bit. For one minor part of the initialization calculation, this remains true, but particular care might be needed to reproduce this part of the calculation from certain years of the re-forecasts.

The years 1991-2010 inclusive were run with an initial version of the scripts which did not adjust snow depths when changing resolution to prepare the initial data for the singular vector computation. For the years 1981-1990 and 2011 onwards, this adjustment is made prior to the singular vector computation. Forecasts run with archived Initial Condition Perturbations will always give a binary identical reproduction of operations. However, if the singular vector calculations are repeated for 1991-2010, the relevant section in inter_fp must be disabled to get binary reproducibility of the original calculations. This issue does not affect the main calculation of atmospheric initial conditions for S4, which always uses the appropriate adjustments to snow depth prior to interpolation. There is no scientific impact on the re-forecasts - this is purely a bit-reproducibility issue.

3. ECMWF Seasonal forecast graphical products

There are two classes of product produced by the Seasonal Forecasting system at ECMWF. The first is a moderately extensive set of graphical products. These are designed to show the main features of the model predictions for the forthcoming seasons, in an easily understood way. The variety of graphical products is designed to show different perspectives on the model-predicted future - no single plot could represent the information contained in the forecast ensemble. The ECMWF web pages currently host the following sets of products from the seasonal forecast system.

3.1 Niño plumes

Forecasts of Equatorial Pacific sea surface temperature anomalies averaged over Niño 3, (5°N-5°S, 90-150°W) Niño 3.4 (5°N-5°S, 120-170°W) and Niño 4 (5°N-5S, 160°E- 150°W) areas are shown for seasonal and annual-range charts. Predicted monthly mean anomalies from each individual ensemble member are shown as red dots joined by thin red curves, and the verifying analysis, where available, is represented by a thick dashed blue curve. Forecasts start on the 1st of a month, and the monthly mean anomaly for that month is the first value plotted. This is joined to the preceding (observed) monthly mean anomaly with a dashed line to represent the continuity of the forecast with the analysis. Note that the lines do not represent the continuous time evolution of the SST, they simply connect the monthly mean values.

The Niño plots show a spread in predicted values - sometimes the spread is large, sometimes it is relatively small. The spread in the first month is largely controlled by the perturbations applied to the ocean initial conditions, in particular the SST perturbations. The growth of the spread in later months is due to the inherent unpredictability in atmospheric behaviour within the coupled system, and depends on both the time of year, the state of El Niño, and the amount of uncertainty in the ocean sub-surface analysis.

The Niño indices are used as indicators of El Niño activity. The predicted anomalies are defined with respect to the NCEP OIv2 climatology adjusted to a base period of 1981-2010, the most recent standard climate reference period available. Note that some El Niño forecasts are still made with respect to older base periods. Since the equatorial oceans have been warming in recent decades, the size of positive anomalies will be larger if an older base period is used. There is no universally agreed definition of "El Niño", although one common approach is to use a threshold of 0.5 °C applied to the Niño 3.4 SST index. The scientific community would typically characterize anomalies of 0.5 - 1 °C as a weak El Niño event. If the Niño 3 or 3.4 indices are significantly colder than normal, "La Niña" conditions are said to prevail. El Niño events can differ substantially in their spatial structure. The three different SST indices provided here give a fair description of how the SST anomalies are distributed in an east-west direction along the equator.

3.1.1 Verification plots

Together with the Niño plume plots, we show verification statistics based on past forecasts/re-forecasts. The r.m.s. error plot shows the cross-validated r.m.s. error for forecasts made with the same calendar start date in previous years. Error bars show a 95% confidence interval for the r.m.s. error for the set of start dates considered, based on the sampling uncertainty due to the finite ensemble size. Also shown is the error obtained by persisting the initial anomaly (black dashed line), and the r.m.s. spread of the ensemble. Comparison of the size of the spread with the forecast error shows the extent to which the forecast plume tends to be over- or under-dispersive. System 4 has a much better match between spread and error than did earlier forecast systems, particularly in the East Pacific. In Niño 4, however, errors are still often much larger than the width of the plume would suggest.

The mean square error skill score (MSSS) relative to climatology shows the skill of the forecast in a range between 1 (a perfect deterministic forecast) and 0 (no better than climatology). It shows how much of the variation of observed SST is being correctly forecast, and gives a sense of the lead time over which the forecast retains useful skill. (MSSS is related to ACC, but unlike ACC is sensitive to errors in amplitude. If the amplitude is correct, then ACC is simply the square root of the MSSS - a MSSS of 0.64 would correspond to an ACC of 0.8).

The Mean Absolute Error (MAE) time-series plot shows a time history of forecast errors for forecasts starting at the given calendar month. The MAE of a forecast for a given month at a given lead time is the absolute difference between the forecast ensemble mean and the verification; this is then averaged across the different forecast leads for a given start date to give the MAE for that start date. We also show what we call the Best Absolute Error (BAE), which for a given forecast date is the average across lead times of either zero (when the verification lies within the predicted range) or the absolute difference between the verification and the outer limits of the predicted range. For a perfect forecasting system, the BAE will be zero or close to zero most of the time.

The MAE and BAE time-series plots give a sense of how much variation there is in the errors, and may suggest whether or not errors have tended to decrease with time or have tended to be associated with certain phases of ENSO. This is designed to complement the other scores, which are averaged across the whole of the re-forecast period.

3.2 Spatial maps

Spatial maps are produced showing the model-predicted anomalies in seasonally averaged quantities. In most cases both global and regional plots are produced, although global and mid-latitude plots are not publicly available. Each plot is labelled with the period for which it is valid, e.g. DJF 2011/12 is the three-month period December 2011 - February 2012. The start date of the forecast is given, as is the number of model integrations in the forecast ensemble and the number used to define the climate.

The lead-time in the drop down menus is defined as the time between the forecast start reference date and the start of the verification period. For System 4 the forecast products are released on the 8th day of each month, so the "usable" lead times are slightly less than their nominal values. Plots for lead times of 1, 2, 3 and 4 months are produced each month. It is good practice to compare the forecast charts for a given target period at different lead times as they become available. The major forecast signals are usually fairly stable, but not always. Weaker signals are subject to appreciable sampling error, and so even if the model signal were to remain unchanged, plots from different months would vary just because of the sampling. The colour scale depends on the field plotted: in most cases blue is used for lower values and red for higher values of a field or probability, but for precipitation brown is used for drier and green for wetter conditions. For individual tercile and outer quintile (20%ile) categories, high probabilities are in red regardless of the field or category being plotted.

3.2.1 Terciles

For each forecast parameter, lead time and calendar start date, there are a set of 450 re-forecasts (a 15 member ensemble for 30 years). For each grid point, the 450 re-forecasts are analysed to determine the terciles of the model climate distribution at the specified lead time. The lower tercile is the value below which the outcome occurs in 1 out of 3 cases in the model climate, and the upper tercile is the value which is exceeded in 1 out of 3 cases. In the absence of any other information, and assuming the climate to be stationary, we would take the probability of a future value exceeding the upper tercile to be 1/3. Our forecast ensemble will give us a different estimate, which we normally take to be the fraction of ensemble forecast members which exceed the upper tercile of the model climate distribution. If there is no particular "forcing" acting on the system, then the proportion of forecast members exceeding the upper tercile will be about 1/3, and indeed this is often the case. However, if there is something in the climate system that "pushes" the forecast in a particular direction, then the predicted probability can be very different from 1/3, and these situations are typically of particular interest.

Plots of the probabilities of the individual tercile categories (ie below the lower tercile, between the lower and upper tercile, and above the upper tercile) are produced, with contour intervals which show both where there is an unusually high chance of a particular category and also where there is an unusually low chance of a particular category occurring. We also produce a tercile summary plot, which shows in a single figure the areas which have an increased probability of being either below the lower tercile or above the upper tercile. This plot gives a good overview of a seasonal forecast, and is listed first in the choice of plots offered to the user on the website.

3.2.2 Ensemble mean

The ensemble mean anomaly represents the shift in the first moment of the predicted probability distribution - it is not a deterministic prediction of the actual value. Tercile and other percentile plots give information on what the model is predicting in terms of the typical variation of the quantity concerned - for example, the chances of it being "unusually" warm. The ensemble mean plots give information on what the model is predicting in absolute terms - °C or mm of rainfall.

The fields are subjected to a local significance test before plotting. Points where the forecast distribution is not significantly shifted at the 10% level compared to the climate distribution are blanked out, and appear white on the plot. This is quite a lax test, and allows both areas of modest signal strength and areas of 'unlucky' sampling to be shown. A second significance test is made at the more stringent 1% level, and a solid contour line encloses areas which are significant at this level. The significance tests are made using the Wilcoxon-Mann-Whitney or 'w' test, which is non-parametric and very efficient at detecting shifts in the mean of a distribution; results are generally very similar to a 't' test. The significance test says nothing about the reliability of the forecast, it simply informs the user about the likelihood of an apparent signal being due to sampling errors in the forecast ensemble. For more on sampling errors, see the sub-section on sampling below.

3.2.3 Probability of exceeding median

Probability maps show the probability of a given model variable (e.g. precipitation) being greater than the model climate median. As with the terciles, the climate median is estimated from the set of 450 re-forecasts made for the same calendar start date and lead time during the 30 year period 1981-2010. The probabilities are shaded symmetrically above 60% and below 40%. The probability plots do have not a significance masking applied, but as for the ensemble mean plots the 1% significance level contour is shown for guidance.

3.2.4 Probability of highest/lowest 20%

We also show probability maps for excedence of the upper and lower 20th percentiles. These are useful for highlighting regions where the distribution of likely outcomes is shifted very substantially from the climatological average. The probabilities here are calculated in the normal way, by counting the number of forecast members in the relevant interval of the climatological distribution. Detailed statistical examination of the tails of the forecast would require different analysis techniques. Although this could be done, verification of how well the forecast system predicts low probability events would be a challenge, given the very limited samples available.

3.3 Spatial maps: sampling errors

Information on the likely impact of sampling errors can be given in different ways. One traditional way is to take a null hypothesis that the forecast distribution is the same as the climatological distribution, and to apply a significance test which is efficient at detecting any shift in the forecast distribution. Such a test can be helpful for screening out situations where an apparent forecast 'signal' is due to a chance fluctuation in the sampling, and not indicative of how the model would behave if a larger ensemble size were available. The results of such significance testing are shown on the ensemble mean and probability of exceeding median charts.

However, this sort of significance testing is very limited, even in telling us about sampling errors. It is not directly relevant to e.g. tercile probabilities, where the user is interested in the sampling accuracy of the probability of a particular event. Although it can warn us about the potential presence of "false positive" signals, it tells us nothing about "false negatives".

3.3.1 Type I errors

First we consider the possibility of a chart showing a signal when in fact none is present. In statistical terms this is a Type I error. If we assume the forecast and climatology distributions are the same, then we can calculate the probabilities of the forecast probability of an event falling within a certain range, allowing for the sampling errors in both the forecast and climatological distributions. Such calculations are made using the binomial distribution, and the bigger our ensemble sizes, the lower the chance of obtaining the 'wrong' forecast category. The following table gives the values (S4 prob) relevant to the tercile and 20%ile plots on the web, and also shows equivalent sampling probabilities for previous systems (S3 and S2):

Tercile plots: model signal =33.3%				
Plotted range	S4 prob	S3 prob	S2 prob	Colour
0-10%	0.0002	0.001	0.003	
10-20%	0.03	0.05	0.07	
20-40%	0.80	0.76	0.70	BLANK
40-50%	0.16	0.18	0.19	
50-60%	0.01	0.02	0.03	
60-70%	0.0002	0.001	0.003	
70-100%	0.0000	0.0000	0.0000	

20%ile plots: model signal = 20%				
Plotted range	S4 prob	S3 prob	S2 prob	Colour
0-10%	0.05	0.07	0.10	
10-30%	0.90	0.86	0.79	BLANK
30-40%	0.05	0.07	0.10	
40-50%	0.001	0.003	0.01	
50-70%	0.0000	0.0001	0.0004	
70-100%	0.0000	0.0000	0.0000	

Median plots: model signal = 50%				
Plotted range	S4 prob	S3 prob	S2 prob	Colour
10-20%	0.0000	0.0001	0.001	
20-30%	0.004	0.009	0.02	
30-40%	0.09	0.11	0.15	
40-60%	0.82	0.77	0.69	BLANK
60-70%	0.09	0.11	0.12	
70-80%	0.004	0.009	0.02	
80-90%	0.0000	0.0001	0.001	

From this table, we see that if no signal is present there is a high probability that the map will show no signal (80% in the case of the tercile plots, 90% in the case of the 20%ile plots, 82% in the case of the median plots). There is a small but not negligible chance of a weak signal being shown (e.g. one colour band either side of the blanked area). Strong signals are very unlikely to occur by chance. The table also shows how the sampling properties have systematically improved with successive forecast systems. Remember that the probabilities in this table apply locally. If we look at a plot for a hypothetical case in which no signal is present, we would expect to see a moderate amount of colour overall, even if the a priori probability of it occurring at any given location is fairly small. If many degrees of freedom are present in the plot, even locally improbable events are likely to occur somewhere

3.2.2 Type II errors

The risk of a chart falsely showing a signal to be present is not the only concern. We also face the situation in which a signal is present, but the chart does not show it. In statistical terms this is a Type II error. We can calculate the probabilities of these errors in the same way as we handled the Type I errors, again allowing for sampling error in both the forecast and the climate ensembles. This time we must specify an assumed true level of signal in order to calculate the effect of sampling errors upon it. Some example tables are given below:

Tercile plots	Model signal = 5%		Model signal = 15%		Model signal = 45%		Model signal = 55%		Model signal = 65%		Model signal = 85%	
Plotted range	Prob		Prob		Prob		Prob		Prob		Prob	
0-10%	0.93	CORRECT	0.20		0.0000		-		-		-	
10-20%	0.07		0.64	CORRECT	0.0003		0.0000		0.0000		-	
20-40%	0.0004	NULL	0.16	NULL	0.26	NULL	0.03	NULL	0.0006	NULL	-	NULL
40-50%	0.0000		0.0000		0.49	CORRECT	0.23		0.02		0.0000	
50-60%	-		-		0.22		0.49	CORRECT	0.23		0.0002	
60-70%	-		-		0.02		0.23		0.51	CORRECT	0.01	
70-100%	-		-		0.0003		0.02		0.24		0.99	CORRECT

20%ile plots	Model signal = 5%		Model signal = 35%		Model signal = 45%		Model signal = 60%		Model signal = 85%	
Plotted range	Prob		Prob		Prob		Prob		Prob	
0-10%	0.94	CORRECT	0.0001		0.0000		0.0000		-	
10-30%	0.06	NULL	0.27	NULL	0.03	NULL	0.0001	NULL	-	NULL
30-40%	0.0000		0.51	CORRECT	0.25		0.007		-	
40-50%	-		0.21		0.48	CORRECT	0.10		0.0000	
50-70%	-		0.02		0.24		0.81	CORRECT	0.02	
70-100%	-		0.0000		0.0003		0.08		0.98	CORRECT

Median plots	Model signal = 5%		Model signal = 15%		Model signal = 25%		Model signal = 35%	
Plotted range	Prob		Prob		Prob		Prob	
0-10%	0.91	CORRECT	0.19		0.01		0.0001	
10-20%	0.09		0.64	CORRECT	0.22		0.02	
20-30%	0.001		0.17		0.55	CORRECT	0.24	
30-40%	0.0000		0.007		0.20		0.51	CORRECT
40-60%	-	NULL	0.0001	NULL	0.01	NULL	0.24	NULL
60-70%	-		0.0000		0.0000		0.0004	
70-80%	-		-		-		0.0000	
Median tables not shown for signals of 65%, 75%, 85% and 95% because results are symmetrical with those given here.								

From these tables it is apparent that for signals in the middle of the range represented by a colour band, there is a high probability that either the correct colour band OR an adjacent one will be shown. Roughly speaking, the sampling resolution of our system is +/- one colour band. In some case the chances of appearing outside of this range are not negligible. For example, in a tercile plot where the model signal is 45%, there is a 0.02 probability of showing a signal in the 60-70% range, and where the model signal is 55%, there is a 0.03 probability of the signal being estimated as being in the 20-40% range, and will thus be plotted as if it were the climatological probability. Thus the risk of substantial model signals being mis-interpreted as "no signal present" is real, given that a global map contains many degrees of freedom.

In overall terms, our system has a moderate sampling resolution. To give more globally reliable estimates of the model signal would require a substantial increase in ensemble size, which would be expensive. An alternative strategy is to increase the ensemble size by pooling results from several different moderately-resolved forecasting models. This has the advantage of starting to sample over errors in the models themselves, which are typically more serious than the sampling errors discussed in this section. This multi-model approach is implemented at ECMWF in the EUROSIP project. A description of the EUROSIP system is available in the EUROSIP User Guide.

3.4 Climagrams

The climagrams show a time series of percentiles of the forecast pdf of an index, together with the corresponding percentiles of the model and observed climatology. Climagrams are created both for indices of atmospheric variability (including the Southern Oscillation, the PNA and the NAO), and for area-averaged temperature and rainfall indices.

3.4.1 Equatorial Southern oscillation

The Equatorial Southern Oscillation index is defined as the difference between the standardized monthly anomalies of sea level pressure averaged over an area of the eastern Pacific (80°W - 130°W, 5°N - 5°S) and over Indonesia (90°E - 140°E, 5°N - 5°S).

3.4.2 Northern hemisphere winter teleconnections

A variety of statistical methods have been used in the literature to define Northern-Hemisphere (NH) teleconnection patterns. For the climagrams, leading variability patterns for the NH winter are defined by an EOF analysis of monthly-mean geopotential height at 500 hPa in the December-to-March season, in the 30-year period (1981-2010) covered by the System-4 re-forecasts. The analysis has been applied to three sectors in the latitude belt 25-85 N: the Pacific/North American sector (160E-80W), the Atlantic/European sector (80W-40E), the Asian/Pacific sector (40E-160E). Data are from the ERA-40 re-analysis until winter 2001/02, from operational ECMWF analyses afterwards.

The first EOF defined in each of the three sectors corresponds to a well-documented teleconnection pattern: respectively, the Pacific/North American pattern, the North Atlantic Oscillation, and the Eurasian pattern. The positive phase of these patterns correspond to an intensification of the westerly winds over the central and eastern parts of the North Pacific and North Atlantic, and over central and eastern Siberia respectively.

The second EOF of the Pacific/North American and the Atlantic/European sector are also retained, since they modulate the intensity and position of the stationary-wave ridges over the north-eastern parts of the Pacific and Atlantic oceans. The sign convention for these EOFs (referred to as the North Pacific dipole and the East Atlantic pattern) is such that positive projections correspond to an amplification of the respective stationary-wave ridge.

3.4.3 Area averages of 2 metre temperature and rainfall

Area averages of 2m temperature and rainfall anomalies are computed over a set of 25 'grid boxes', shown in the figure below:

For 2m temperature, averages are computed using land fraction as a weight, in order to isolate temperature variations over land (2m T over sea is strongly constrained by the underlying SST). For rainfall anomalies, averages are computed over the whole area in each box. An exception is made for the 'Central tropical Pacific' grid box, which has no land points at the model resolution, and where no weight is applied for either variable.

The grid boxes were chosen to correspond to fairly homogeneous regions for seasonal-mean anomalies of both temperature and rainfall, with area-average values being positively correlated with anomalies at individual grid points over at least 80-90% of the area. However, for some areas, the box definition may be more suitable for one of the two variables, or for one particular season of the year (for example, the 'Sahel' box is optimised for summer rainfall).

3.5 Tropical storm forecasts

The seasonal forecast model generates synoptic features analogous to tropical storms. These tropical storms are identified as tropical depressions with a "warm core" structure and a threshold strength - it is the warm-core structure which makes them dynamically equivalent to observed tropical storms, despite the fact that the model does not have sufficient resolution to produce the intensity of winds seen near the centre of a hurricane or typhoon. A tracking algorithm is applied to the 12-hourly upper air fields produced by each model integration, to locate and track individual tropical storms in the various ocean basins where they occur. Statistics on these tracked tropical storms can then be compared between a specific forecast ensemble and the climatology of model forecasts for that time of year.

3.5.1 Tropical storm numbers

For most seasonal forecast products (Niño SST plumes, rainfall and temperature anomalies), forecast values are calculated as an anomaly by comparison with the model climate in an additive manner. For some ocean basins, the model climate of tropical storm numbers differs substantially from the observed numbers (e.g. by a factor of 2). Since we aim to produce forecasts of absolute numbers of storms (rather than e.g. just the number of storms as a percentile of the climatological distribution), we choose to *scale* the number of model storms multiplicatively such that the model climate matches the observed climate, rather than make an additive correction. Precisely, a scaling factor is chosen by ordering the climatological distribution of tropical storm numbers, and taking the mean of those included in the range of the 25-75%iles. The mean is calculated in this way from the observed climate (30 years) and the model climate (30*15 = 450 model years), and the scaling factor is the ratio of the two means. This approach is taken to avoid the means being unduly influenced by outliers in the distribution of tropical storm numbers. To estimate the standard deviation of the predicted number of tropical storms, we scale the standard deviation of the original model forecast ensemble by the square root of the scaling factor applied to the mean number of storms. (If we were to model the number of storms in a season as a Poisson process, this square root scaling would be exactly the right way to scale the spread of the ensemble. We are not claiming that this 'model' is precisely the correct one to use, but we consider the derived scaling to be both reasonable and robust).

The plots of tropical storm frequency show both the expected number of storms in each relevant basin for the coming six months, and the model-estimated standard deviation. Also shown is the result of a w-test on whether the 51 member ensemble forecast is or is not significantly shifted relative to the model climatological distribution.

Forecasts of Hurricane/Typhoon numbers are shown in the same format, and using the same methodology, but with a higher threshold to distinguish the stronger storms.

Forecasts are also provided of ACE, or Accumulated Cyclone Energy. This is calculated by accumulating the kinetic energy of each storm across its area of influence and its lifetime - so large, intense, long-lived storms will contribute much more than small scale or short lived storms. These are shown as scores normalized by the mean climatological value - greater than 1 means more energy than usual, less than 1 means less.

3.5.2 Track density

There are important interannual variations in the tracks of tropical storms, driven by large scale SST anomalies and other predictable factors. These changes in tracks are important for assessing the risk of landfall in various regions, and are just as relevant as variations in the total number of storms. Earlier forecast systems were not able to produce very good track information, in particular because the low resolution meant that the lifetime of model storms was too short. With the improved resolution of S4, it is now possible to start providing track density information.

The tropical storm density anomaly map shows the anomaly in the number of tropical storms expected to pass within about 300 km of a point. The numbers have not been calibrated, and so the map should be considered as a qualitative indicator of the expected anomaly. This plot naturally emphasizes areas where large numbers of storms usually occur - it is these areas that have large absolute variations.

The standardized tropical storm density standardizes the predicted anomaly against the usual model variability in the ensemble mean. This removes the "bias" to high storm areas, and shows more clearly where there are meaningful signals in moderate and lower storm density areas. To avoid unnecessary alarm or over-interpretation of results, we only use three simple categories: "reduced expected value", where the anomaly in the ensemble mean is below minus one standard deviation of the ensemble mean, "enhanced expected value", where the anomaly is above one standard deviation, and "usual expected value" for values in between. Areas where the model prediction and climate are both for a track density of less than 0.5 are blanked out in white. Note that this is not a tercile probability map - it takes no account of how much variation there is between ensemble members for a given year, but only standardizes using the variability of the ensemble mean.

3.5.3 Verification

For verification purposes, plots are provided showing the time-series of re-forecasts and forecasts of tropical storm numbers for each basin from the relevant start month. These plots also show the observed tropical storm numbers, as given by the National Hurricane Center in Miami and the Joint Typhoon Warning Center in Guam. The plots also include some verification statistics: the ACC between the predicted ensemble mean and the observed number of storms, and the cross-validated r.m.s. error in the tropical storm number. These time-series plots are generated dynamically, and will include progressively more years of data during the lifetime of System 4.

Time series plots and verification information are provided separately for forecasts of tropical storm number, hurricane/typhoon number and ACE.

For the track density products, maps are shown of ACC between ensemble mean predicted track density and the observed values, for the 30 years of the hindcast set.

3.6 Annual range forecasts

Four times a year, annual range forecasts are run as an extension of the usual 7-month long seasonal forecasts. The primary purpose of the annual-range forecasts is to give an outlook for El Niño. Note that these forecasts are still considered experimental, and the products from them are subject to change.

- Niño plumes: The Niño plume plots are produced as per the seasonal forecast Niño plumes. Verification information is also provided.
- Spatial maps: The spatial maps are being reviewed, and are not presently available for System 4.
- Climagrams: The climagrams are not produced.
- Tropical storm forecasts: The tropical storm forecasts are not produced.

4. Seasonal forecast data products

Data products consist both of directly output data from the model forecasts, and various derived fields which are calculated and then encoded in GRIB for archiving and distribution. The data products give access to quantitative forecast values, and allow the creation of an almost unlimited range of user-specific forecast products. ECMWF member states can access all of the data directly from the archive system, and/or obtain real-time atmosphere forecast data via dissemination. Only a subset of the real-time data, as defined in the Catalogue of ECMWF Real-Time Products, is designated for commercial use and available for dissemination to commercial customers. For the full set of archive data, including for the re-forecast period, see Accessing forecasts.

4.1 Atmosphere model output

The atmosphere model outputs many fields, which are then archived. The following tables detail the direct output of the atmosphere model.

Additionally, monthly means of the fields in each model integration are automatically calculated and archived once the integration is complete. For surface fields, as well as the monthly MEAN field, fields consisting of the MIN and MAX values occurring during the month at each grid point are formed. The MEAN, MIN and MAX values are calculated from the set of instantaneous output fields, with a 6h or 24h sampling interval, and so do not sample variations occurring on time-scales shorter than this. (For a few fields such as 2 metre temperature, separate parameters are used to monitor minimum and maximum values on a time step by time step basis within the model itself). The standard deviation (SD) of the values used to calculate the monthly MEAN is also calculated. Since the available input data can be either 6h or 24h, for some fields the standard deviation includes the diurnal cycle, while for others it does not. For accumulated fields, the monthly mean rate of accumulation is calculated. Since archived data are generally in SI units, monthly mean fluxes have convenient units, W m^2. For rainfall, data is archived using the SI unit of m s^{-1}, so rainfall must be scaled by the user to a unit such as mm day^{-1}. Finally, for precipitation related fields MIN values are not calculated since in reality they are generally zero.

4.1.1 Upper air fields

The following upper-air fields are archived every 12 hours on the indicated pressure levels. Monthly mean values are also calculated for all these fields, but including a wider set of pressure levels, namely: 1000/925/850/700/500/400/300/250/200/150/100/70/50/30/20/10/5/2/1 hPa. Only the monthly MEAN is calculated for upper air fields, not MIN, MAX or SD.

Parameter number	Parameter name	Pressure levels (hPa)
129	Geopotential	1000/925/850/700/500/400/300/200/100/70/50/30/10
130	Temperature	1000/925/850/700/500/400/300/200/100/70/50/30/10
138	Vorticity (relative)	1000/925/850/700/500/400/300/200/100/70/50/30/10
155	Divergence	1000/925/850/700/500/400/300/200/100/70/50/30/10
133	Specific humidity	1000/925/850/700/500/400/300/200/100/70/50/30/10
203	Ozone (grid point	1000/925/850/700/500/400/300/200/100/70/50/30/10

The annual-range forecasts have a much reduced archive of pressure level data, since they are primarily designed for an ENSO outlook. 12 hourly data is only available for the first 4 fields, and for pressure levels 850, 700, 500, 400, 300 and 200 hPa. This data (together with corresponding surface fields) allows tropical cyclone tracks to be calculated. Monthly mean data for annual range forecasts are in principle available for all parameters at all levels, although for 1991-2008, for ensemble members 5-14 only, the annual range forecasts only have data from the pressure levels listed in the table above, and not the additional 6 levels which are usually available.

This field is produced every 12 hours as grid point data on selected isentropic surfaces:

Parameter number	Parameter name	Isentropic levels	Monthly mean output
60	Potential vorticity	315K/330K	MEAN

This field is produced every 12 hours as grid point data on a constant PV surface:

Parameter number	Parameter name	PV level	Monthly mean output
3	Potential temperature	2000 (ie PV = 2)	MEAN

4.1.2 Surface fields

The following surface fields are output and archived at step 0 only (2 fields):

Parameter number	Parameter name	Output frequency	Monthly mean output
129	Surface geopotential	step 0 only	
172	Land-sea mask	step 0 only	

The following surface fields are output and archived every 6 hours (6 fields):

Parameter number	Parameter name	Output frequency	Monthly mean output
164	Total cloud cover	6h	MEAN/MAX/MIN/SD
165	10m u wind component	6h	MEAN/MAX/MIN/SD
166	10m v wind component	6h	MEAN/MAX/MIN/SD
167	2m temperature	6h	MEAN/MAX/MIN/SD
168	2m dewpoint temperature	6h	MEAN/MAX/MIN/SD
151	Mean sea level pressure	6h	MEAN/MAX/MIN/SD

The following surface fields are output and archived every 24 hours (39 fields). Fields marked with an asterisk are accumulated fields - the archive contains the accumulated value of the field from the start of the forecast. Parameters 49,51,52 and 55 are not archived at step 0.

Parameter number	Parameter name	Output frequency	Monthly mean output
31	Sea ice cover	24h	MEAN/MAX/MIN/SD
33	Snow density	24h	MEAN/MAX/MIN/SD
34	Sea surface temperature	24h	MEAN/MAX/MIN/SD
39	Volumetric soil water layer 1	24h	MEAN/MAX/MIN/SD
40	Volumetric soil water layer 2	24h	MEAN/MAX/MIN/SD
41	Volumetric soil water layer 3	24h	MEAN/MAX/MIN/SD
42	Volumetric soil water layer 4	24h	MEAN/MAX/MIN/SD
49	Maximum 10m wind gust	24h	MEAN/MAX/MIN/SD
51	Max 2m temperature in previous 24h	24h	MEAN/MAX/MIN/SD

52	Min 2m temperature in previous 24h	24h	MEAN/MAX/MIN/SD
55	Mean 2m temperature in previous 24h	24h	MEAN/MAX/MIN/SD
78	Total column liquid water	24h	MEAN/MAX/MIN/SD
79	Total column ice water	24h	MEAN/MAX/MIN/SD
137	Total column water vapour	24h	MEAN/MAX/MIN/SD
139	Soil temp level 1	24h	MEAN/MAX/MIN/SD
141	Snow depth	24h	MEAN/MAX/MIN/SD
142	Large scale precipitation*	24h	MEAN/MAX/SD
143	Convective precipitation*	24h	MEAN/MAX/SD
144	Snow fall*	24h	MEAN/MAX/SD
146	Surface sensible heat flux*	24h	MEAN/MAX/MIN/SD
147	Surface latent heat flux*	24h	MEAN/MAX/MIN/SD
169	Surface solar radiation downwards*	24h	MEAN/MAX/MIN/SD
170	Soil temp level 2	24h	MEAN/MAX/MIN/SD
175	Surface thermal radiation downwards*	24h	MEAN/MAX/MIN/SD
176	Surface solar radiation*	24h	MEAN/MAX/MIN/SD
177	Surface thermal radiation*	24h	MEAN/MAX/MIN/SD
178	Top solar radiation*	24h	MEAN/MAX/MIN/SD
179	Top thermal radiation*	24h	MEAN/MAX/MIN/SD
180	East/West surface stress*	24h	MEAN/MAX/MIN/SD
181	North/South surface stress*	24h	MEAN/MAX/MIN/SD
182	Evaporation*	24h	MEAN/MAX/MIN/SD
183	Soil temperature level 3	24h	MEAN/MAX/MIN/SD

186	Low cloud cover	24h	MEAN/MAX/MIN/SD
189	Sunshine duration*	24h	MEAN/MAX/MIN/SD
205	Runoff*	24h	MEAN/MAX/MIN/SD
206	Total column ozone	24h	MEAN/MAX/MIN/SD
228	Total precipitation	24h	MEAN/MAX/SD
236	Soil temperature level 4	24h	MEAN/MAX/MIN/SD
243	Forecast albedo	24h	MEAN/MAX/MIN/SD

The following fields are archived for initial dates from 1981-1990 and from 2009 to present:

8	Surface runoff*	24h	MEAN/MAX/MIN/SD
9	Sub-surface runoff*	24h	MEAN/MAX/MIN/SD

The following derived field is not archived at daily resolution, but the monthly statistics are calculated and archived:

207	10m scalar wind speed	MEAN/MAX/MIN/SD

The annual range forecasts have a much reduced archive of daily values of surface fields, being SST, total precipitation, OLR, daily Tmax and Tmin, and 12 hourly values of MSLP and 10m winds. Monthly means are however available for the full set of parameters, just as for the seasonal forecasts.

4.1.3 Model level fields

A limited number of integrations produce and archive model level data. This is primarily intended to allow use in dynamical downscaling. Because both the production and archiving of model level data can be very expensive, only a limited set of data is available. For the re-forecasts, only the first 5 ensemble members have model level data. For the forecasts, the first 11 members have model level data. In all cases, model level data is produced at 12h intervals from step 0 to 3672 (i.e. the first 5 months only). Only every second level is archived (1,3,5,7,....,89,91), a total of 46 levels. Two "surface" fields that are needed to reconstruct the model state are archived as level=1, following ECMWF convention.

Parameter number	Parameter name	Output frequency	Levels
129	Surface geopotential	12h	1
152	Log surface pressure	12h	1
130	Temperature	12h	1/TO/91/BY/2
133	Specific humidity	12h	1/TO/91/BY/2
138	Vorticity (relative)	12h	1/TO/91/BY/2
155	Divergence	12h	1/TO/91/BY/2
246	Cloud liquid water content	12h	1/TO/91/BY/2
247	Cloud ice water content	12h	1/TO/91/BY/2

For the ensemble members and time step range for which model level data is output, 4 surface fields are output every 12 hours instead of every 24 hours, namely parameters 139, 170, 183 and 236 containing soil temperature.

4.2 Wave forecast data

The wave model archives the following fields every 24 hours (6 fields):

229	Significant wave height	MEAN/MAX/MIN/SD
230	Mean wave direction	-
231	Peak period of 1d spectra	MEAN/MAX/MIN/SD
232	Mean wave period	MEAN/MAX/MIN/SD
233	Coefficient of drag with waves	MEAN/MAX/MIN/SD
245	10m wind speed	MEAN/MAX/MIN/SD

Wave direction cannot be meaningfully averaged as a scalar quantity (e.g. the average of 359° and 1° is 180°!), so no monthly means are formed.

4.3 Ocean forecast data

Ocean data from the NEMO model is produced in netCDF format, with a complex native grid. The data are not available from MARS. Some raw data from the forecasts is archived to tape, but there are no immediate plans to make ocean data from the seasonal forecasts available to users. This service may be provided at a later date if there is sufficient demand.

4.4 Derived forecast products

For the real-time forecasts only (from May 2011 onwards for System 4), we calculate a number of derived forecast products.

4.4.1 Monthly mean anomalies

Forecast monthly mean anomalies are calculated relative to a climate mean formed from the appropriate 1981-2010 re-forecasts. The anomalies are calculated for each ensemble member and for all of the monthly mean fields. For S4, these data are archived in the MMSA stream as type FCMEAN.

4.4.2 Ensemble means

Ensemble means of the 51 member forecast ensemble are calculated for all of the monthly mean fields, and archived in the MSMM stream as type EM. The ensemble mean of the anomalies is also calculated for each monthly mean field, and archived in the MMSA stream as type EM.

4.4.3 Hindcast climate means

This is a new forecast product introduced with System 4. For each forecast monthly mean field, for a given start date and lead time, the climate mean of the corresponding 1981-2010 reforecasts is calculated. The climate means are archived as a new type HCMEAN in the MSMM stream. The date of the HCMEAN data is the date of the real-time forecast with which they are associated.

4.4.4 Other derived products

For S4, it is planned to introduce additional derived products, in particular containing information on probabilities, e.g. of exceeding tercile boundaries. These additional derived probabilities are not yet available, but will be added to the system in the near future, some time after the start of System 4.

4.5 Mars retrievals

Data are archived in the multi-model streams, with ORIGIN=ECMF:

- STREAM=MMSF for the direct model output;

- STREAM=MSMM for the monthly means.

Selected monthly mean anomalies, which are only produced for the real-time forecasts, will be archived with appropriate parameter definitions in STREAM=MSMM, instead of STREAM=MMSA. Wave model output will continue to be archived as for System 3 in the WASF and SWMM streams.

All production data (re-forecasts, pre-operational and operational real-time forecasts) are archived as EXPVER=0001, SYSTEM=4. If atmosphere data is retrieved on the archived grid, then the resolution will differ from that of System 3.

An example showing how to modify a MARS request for System 3 data according to the details given above in order to retrieve the equivalent System 4 data is given in the table:

Mars request - System 3	MARS request - System 4
RETRIEVE, STREAM = MSMM, ORIGIN = ECMF, SYSTEM = 3, METHOD = 1, NUMBER = 0/TO/40, CLASS = OD, EXPVER = 1, DATE = 20110501, TIME = 00, TYPE = FCMEAN, LEVTYPE = SFC, PARAM = 51, FCMONTH = 1/2/3/4/5/6/7, TARGET = 2m_tmax_monthly RETRIEVE, NUMBER = 0/TO/10, DATE = 19810501/19820501/ 19830501/19840501/ 19850501/19860501/ 19870501/19880501/ 19890501/19900501/ 19910501/19920501/ 19930501/19940501/ 19950501/19960501/ 19970501/19980501/ 19990501/20000501/ 20010501/20020501/ 20030501/20040501/ 20050501, TARGET = 2m_tmax_monthly_climate	RETRIEVE, STREAM = MSMM, ORIGIN = ECMF, SYSTEM = **4**, METHOD = 1, NUMBER = 0/TO/**50**, CLASS = OD, EXPVER = 1, DATE = 20110501, TIME = 00, TYPE = FCMEAN, LEVTYPE = SFC, PARAM = 51, FCMONTH = 1/2/3/4/5/6/7, TARGET = 2m_tmax_monthly RETRIEVE, NUMBER = 0/TO/**14**, DATE = 19810501/19820501/ 19830501/19840501/ 19850501/19860501/ 19870501/19880501/ 19890501/19900501/ 19910501/19920501/ 19930501/19940501/ 19950501/19960501/ 19970501/19980501/ 19990501/20000501/ 20010501/20020501/ 20030501/20040501/ 20050501/**20060501/** **20070501/20080501/** **20090501/20100501**, TARGET = 2m_tmax_monthly_climate

The longer range integrations (out to 13 months) are archived separately from the 7 month integrations, and are accessed by specifying METHOD=3 instead of METHOD=1. The first 7 months of METHOD=3 data for each extended integration is a simple copy of the corresponding METHOD=1 data.

Note that all seasonal forecasts start at 00 UTC. Several types are available in the monthly mean stream, including fcmean (forecast mean), fcmax (the maximum value of the field occurring during the month) and fcmin. The month(s) to be retrieved are specified in terms of time into the forecast with fcmonth.

Wave model data are in stream=wasf and stream=swmm for daily and monthly mean values respectively.

5. Product interpretation

The ECMWF seasonal forecast system is conceived of as a numerical system. The products, at least for now, are simply a statement of how the numerical calculations behave. Such numerical products contain information on what is to be expected on seasonal timescales, but they also contain errors. Unthinking use of the raw numerical forecast products is not recommended. Actual forecasts for users should be carefully prepared, perhaps combining data from several empirical and/or numerical sources. The creation and issuing of properly prepared forecast statements is not a task undertaken by ECMWF, but is left to others, such as National Meteorological Services or appropriate international organizations. The probability maps on the web pages are **uncalibrated** - that is, they directly represent model output, and no adjustments to the probabilities have been made to account for model errors.

A correct interpretation of seasonal forecast products depends both on understanding the plot, and on understanding the characteristics of the forecasting system as a whole. In particular it is essential to use information about the past performance of the seasonal forecast, the spatial distribution of the forecast skill and the forecast reliability. Remember that the number of past cases is limited, and although we can gain some indication of model performance, sampling errors mean that it is easy to either over- or under-estimate model skill.

The significance tests shown on the spatial maps (see Section 3), measure how confident we are that the model forecast distribution is shifted with respect to the model climate. The significance test is not a measure of confidence in the skill of the forecast, because it takes no account of what has been observed to happen in the real world. Skill estimates are given for all forecast products, based on past performance. In most cases, however, there is a large sampling error on this estimate - in some cases skill will be underestimated, and in some cases it will be overestimated.

The ECMWF seasonal forecast model is global, has a surface grid with an 80 km spacing, and at its best can only hope to represent large scale weather patterns. Local weather and climate is much influenced by features too small to be included in the relatively low resolution model (hills, coastlines, land surface properties). Simply trying to read off local values from the maps could be very misleading. There are various objective methods which might in principle be useful for transforming the global-scale numerical model output into improved regional or local scale products. Study of patterns of variability (EOFs, covariance statistics etc.) may enable erroneous shifts in model variability to be corrected - if data records are long enough to be confident of the statistical robustness of any purported shifts. Simple adaptive filters may be useful for obtaining local values from direct model output.

Interpreting rainfall anomalies in low rainfall regimes needs to be done carefully. There are substantial desert areas where the median rainfall in a month or season is zero, but heavy rainfall does sometimes occur. In this case the mean rainfall is always positive, the model mean and observed mean are almost certainly different, the spread of rainfall is usually small (i.e. most ensemble values having the same value, zero), the meaning of "rainfall anomaly" may not be what it seems, and distributions are very non-normal. There are different ways of dealing with rainfall variability in such regions, and different ways of representing the results graphically. No simple treatment resolves all the difficulties of interpretation, and the general processing which is used to cover all rainfall regimes is certainly inadequate.

6. Operational history

System 4 became operational in November 2011, when it replaced System 3. It was run in pre-operational mode from May 2011, and a full 51 member ensemble exists from this date on. System 3 continued to run in non-operational mode for a few months after the end of its operational phase. Similar overlaps occurred with earlier transitions between systems, although there is only ever one system which is operational at a given time.

System 3 was operational from March 2007 to October 2011

System 2 was operational from to January 2002 to February 2007

In the early years of seasonal forecasting at ECMWF, the systems were run as experimental real-time forecast systems, and were not fully operational. Full operational status was achieved during the lifetime of System 2.

System 1 was "operational" from January 1997 to December 2001

7. References and further literature

System 4 has been documented in an ECMWF technical memorandum:

https://www.ecmwf.int/en/elibrary/11209-new-ecmwf-seasonal-forecast-system-system-4

Many parts of the System 4 design have been inherited from earlier ECWMF seasonal forecast systems, and are discussed in the following references:

7.1 Further literature on the ECMWF seasonal forecasting system:

- Alves O., M Balmaseda, D. Anderson and T Stockdale 2004 Sensitivity of dynamical seasonal forecasts to ocean initial conditions. Quarterly Journal Royal Meteoroloical Society, 130, Jan 2004, 647-668.
- Segschneider, J., D.L.T. Anderson, J. Vialard, M. Balmaseda, T.N. Stockdale, A.Troccoli, and K. Haines, 2001. Initialization of seasonal forecasts assimilating sea level and temperature observations.J. Climate, 14, 4292-4307.
- Stockdale, T N, 1997: Coupled ocean atmosphere forecasts in the presence of climate drift. Mon. Wea. Rev., 125, 809-818.
- Stockdale T. N., D. L. T. Anderson, J. O. S Alves, and M. A. Balmaseda, 1998. Global seasonal rainfall forecasts using a coupled ocean-atmosphere model. Nature, 392, 370-373.
- Stockdale, T. N., D. L. T. Anderson, M. A. Balmaseda, F. Doblas-Reyes, L. Ferranti, K. Mogensen, T. N. Palmer, F. Molteni and F. Vitart, 2011: ECMWF seasonal forecast system 3 and its prediction of sea surface temperature. Clim. Dyn. doi 10.1007/s00382-010-0947-3
- Vialard, J., F. Vitart, M.A. Balmaseda, T.N. Stockdale and D.L.T.Anderson, 2005. An ensemble generation method for seasonal forecasting with an ocean-atmosphere coupled model. Mon. Wea. Rev. 133, 441-453.
- Vitart, F., and T.N. Stockdale, 2001: Seasonal forecasting of tropical storms using coupled GCM integrations. Mon. Wea. Rev, 129, 2521-2527.

7.2 Selected ECMWF publications:

- Segschneider, D.L.T. Anderson, J. Vialard, M. Balmaseda, T.N. Stockdale, A. Troccoli and K.HainesInitialization of seasonal forecasts assimilating sea level and temperature observations December 2000 technical Mem. 320

- Weaver, J. Vialard, D. L. T. Anderson and P. Delecluse Three- and four-dimensional variational assimilation using a general circulation model of the tropical Pacific Ocean March 2002 Tech.Mem.365
- Vitart, D Anderson and T Stockdale Seasonal forecasting of tropical cyclone landfall over Mozambique using coupled GCM integrationsOctober 2002 Tech. Mem. 387
- Anderson, D., T. Stockdale, M. Balmaseda, L. Ferranti, F. Vitart, P. Doblas-Reyes, R. Hagedorn, T. Jung, A. Vidard, A. Troccoli and T. Palmer Comparison of the ECMWF seasonal forecast Systems 1 and 2, including the relative performance for the 1997/8 El Niño (SAC Report Sept 2002) April 2003 Tech. Mem. 404
- ECMWF Newsletter No.98 -Summer 2003: The ECMWF seasonal forecasting system p.17-25
- ECMWF Newsletter No.99 -Autumn/Winter 2003: DEMETER: Development of a European multi-model ensemble system for seasonal to interannual prediction p.8-17
  Vialard, J., F. Vitart, M.A. Balmaseda, T.N. Palmer and D.L.T.Anderson 2003: An ensemble generation method for seasonal forecasting with an ocean-atmosphere coupled model Technical Memorandum 417
  Coelho, C.A.S., S. Pezzulli, M. Balmaseda, F.J. Doblas-Reyes and D.B. Stephenson 2003: Skill of coupled model seasonal forecasts: A Bayesian assessment of ECMWF ENSO forecasts. Technical Memorandum 426
  Anderson, D.L.T., T. Stockdale, M. Balmaseda, L. Ferranti, F. Vitart, F. Molteni, F. Doblas-Reyes, K. Mogenson and A. Vidard, 2007: Development of the ECMWF seasonal forecast System 3. Technical Memorandum 503.
- Anderson, D.L.T., T. Stockdale, M. Balmaseda, L. Ferranti, F. Vitart, F. Molteni, F. Doblas-Reyes, K. Mogenson and A. Vidard, 2007: Seasonal forecast System 3. ECMWF Newsletter N.110 p.19.
- Balmaseda, M., A. Vidard and D. Anderson, 2007: The ECMWF System 3 ocean analysis system. Technical Memorandum 508.
- Vitart Frederic, Tim Stockdale, Laura Ferranti 2007: Seasonal Forecasting of Tropical storms frequency. ECMWF Newsletter N.112 p.16
- Molteni Franco, Laura Ferranti, Magdalena Balmaseda, Tim Stockdale, Frederic Vitart 2007: New web products for the ECMWF seasonal forecast system-3. ECMWF Newsletter N.111 p.28