Workshop
"Sub-seasonal to Seasonal Prediction"
Met Office, Exeter –1 to 3 December 2010

Current capabilities in Sub-seasonal to Seasonal Prediction

David Anderson

Executive summary

The main goals of this workshop are to establish current capabilities in sub-seasonal to seasonal prediction, to identify high-priority research topics and demonstration projects and to develop recommendations for the establishment of an international research project.

Considerable progress has been made in improving the skill of medium range weather forecasts and in developing operational seasonal forecasting. Forecasting in the intermediate range between medium range and seasonal is difficult as the importance of the initial conditions wanes, and the importance of slower boundary conditions such as sea surface temperature increases but has only a modest influence on the weather and climate, especially away from the tropical regions. Tropical sea surface temperatures play an important role not only in controlling the weather/climate in the tropics but in the extra-tropics also, through various tele-connections. The El Nino Southern Oscillation (ENSO) is the best-known long-lived source of predictability in the tropics but changes in SST in the Indian Ocean are also significant, though the forecast horizon is likely to be shorter than for ENSO. Predicting changes in the equatorial Atlantic SSTs has been less successful. Despite the difficulties in forecasting for the extended range, there is none the less considerable potential and it is worthwhile developing a research strategy to explore and exploit this potential.

TIGGE has been successful in establishing an extensive database of medium-range forecasts from 10 different models, most of which produce forecasts to at least 9 days. In contrast to seasonal forecasting, systematic model errors are much less dominant for short- and medium-range forecasts, so it was not seen as a priority to run back integrations (or hindcasts or reforecasts) to correct model errors at this range. Consequently, hindcast data were not included in the TIGGE data set. Two of the main priorities of TIGGE are to foster research on methods to combine ensembles and to correct systematic errors. Several groups have focused on the combination of ensemble forecasts from different models; a multi-model (MM) ensemble from four centres shows improved skill over any single model at the two-week time range. In the four-member MM the most important contributing model was shown to have significant model error, implying that model error must be significant in the other models also and suggesting that procedures should be put in place for handling this, not just in the medium range but even more importantly for the extended range considered here. Additionally, an extensive series of hindcasts was generated by one participating group which allowed the forecasts to be calibrated by partly correcting for model error.

On the seasonal range, model error has been recognised as a major problem ab initio as the model drift is as large as the signal being forecast. Calibration can be made by removing drift, and adjusting the spread of the ensemble. The experience in the TIGGE set was that a calibrated forecast from a single model was as skilful as a multi-model of un-calibrated forecasts. Since there were hindcasts for only one model a MM forecast of calibrated models was not possible but should be considered in future, especially for the intra-seasonal range considered here (for which an extensive set of hindcasts spanning many years will certainly be needed). Calibration might be particularly useful in applications. An alternative approach is to use a shorter forecast range for which more data and realisations are available to calibrate the forecasts at a longer range, but this has not been tested.

One of the long-lived atmospheric phenomena is the Madden Julian Oscillation (MJO) which exerts a strong influence on tropical weather and climate as well as influencing extra-tropical weather through interactions with phenomena such as the North Atlantic Oscillation (NAO) and other tele-connections. Unfortunately, representing the MJO well in models for intra-seasonal and seasonal timescales has not been easy. Considerable progress has been made recently in at least one operational centre in improving the representation of the MJO leading to improved predictability on the intra-seasonal timescale. However, even in this case the representation still has deficiencies: propagation speeds are too slow and some of the tele-connections are too weak. The improvement noted came primarily from improvements in convective parameterization but it is likely that further parameterization improvements are needed. Increasing the model resolution to ~10 km did not improve the representation of the MJO, nor did it improve the representation of the Indian monsoon.

Using their monthly forecast system, Vitart and Molteni (2010) have shown an increase in skill in the monthly timescale relative to an earlier version of the model and have recently experimented with increasing the forecast range from 31 to 45 days. They find that the there is little skill at the 20-25 day range if there is no MJO present in the initial conditions but if there is an MJO present then the skill is increased. It has been observed that there are phase-lags between the MJO and the extra-tropical response (Lin et al 2010a, Cassou 2008). The model seems to capture many of these interactions, not just in the North American region but even in the European sector. As the time range increases the temporal averaging period needed to smooth out chaotic processes increases from daily to weekly to monthly to seasonal and spatial averages can be increased as well depending on the application in mind. Consistent with Vitart and Molteni (2010), Lin et al. (2010a), using the Canadian operational model, have shown that the MJO signal in the initial condition helps to increase the forecast skill of the North Atlantic Oscillation (NAO) at the 10-30 day range.

Improvements in the parameterization of convection can lead not only to improvements in the MJO but also in the frequency of blocking, two areas in which models have consistently done badly, though here too there have been encouraging improvements recently (for example Scaife et al 2011). Parameterization is a very wide topic but the intention is not to cover every possible angle on this here. Jung et al 2010 gives a flavour of the impact of various improvements in parameterization over the last few years. Notwithstanding the very significant improvements that have been made, it is clear that errors in the representation of fast physics processes remain a key limiting factor in the skill of our models across all timescales from NWP to sub-seasonal to seasonal. Efforts to develop these representations are therefore of crucial importance for many applications, and full advantage should be taken of testing across timescales for example NWP or TRANSPOSE-AMIP short-range testing may provide insights into the sources of error for seasonal prediction which are hard to glean directly from study of the seasonal results. A key challenge for model analysis and diagnosis on any timescale is to provide insights which are detailed enough to help inform which parts of the physical parametrization are in error and thus aid model developers (rather than, for example, just saying that some process such as the MJO or ENSO is poorly represented). Studies such as those showing evidence for an unrealistic link between humidity and rainfall in many simulations of the MJO are good examples of this being done in practice. There is also an argument that parametrization is a stochastic process and that the approach has to be broadened, to include various forms of stochastic forcing, to be less local and to represent uncertainty in imprecisely known process parameters.

Substantial progress has been made with convective parameterization although a conundrum still remains: we do not yet have the computational capacity to fully resolve moist convection in global models (~100m mesh required) and convective parameterizations have not been improved to the level where they can reliably replicate convective organization. Whilst not represented by parameterizations, convective organization is intrinsic to important tropical phenomena (for example the MJO, monsoons, inter-tropical convergence zone), as well as propagating precipitation systems over mid-latitude continents. Representing convective organization in sub-seasonal to seasonal prediction models involves either traditional parameterization of cumulus and explicit representation of convective organization, or, preferably, its parameterization. Convective organization is a central element of the Year of Tropical Convection (YOTC). Coordinated jointly by the WCRP and WWRP/THORPEX, YOTC aims to advance the characterization, diagnosis, modelling, parameterization of multi-scale tropical convection and its large-scale effects (Moncrieff et al 2007; Waliser and Moncrieff 2008; www.ucar.edu/yotc). This is relevant to intra-seasonal to seasonal prediction.

Land processes can add to forecast skill and if the land is not well represented degradation in skill results. Methods for initializing the land component not just in real-time but also in extensive hindcasts are needed. This could involve stand-alone land analyses or alternatively schemes such as ALI (Hudson and Alves 2007) which allow the model to develop its own soil moisture etc. when nudged towards an atmospheric reanalysis or analysis. Ensemble generation usually includes the representation of uncertainty in the atmospheric initial conditions and ocean initial conditions but not in land conditions. Insufficient spread in near surface temperature suggests that there should be an ensemble of land states also. Snow and ice cover can contribute to predictability on the intra-seasonal to seasonal range but only 2 centres (Environment Canada and UKMO) currently initialise and dynamically evolve sea-ice in operational forecasts. An international research project under the CLIVAR Working Group for Seasonal to Inter-annual Prediction to advance sea-ice prediction has been started recently. The importance of the stratosphere has not been fully assessed but many individual case studies now show a likely role for its influence on the extra-tropics. While the influence of the stratosphere on year round averaged skill scores may be modest, there is a good case for an impact on the NAO and the southern annular mode, especially during a sudden stratospheric warming and other times when the polar vortex is active. An international CLIVAR project run under the CLIVAR Working Group on Seasonal to Inter-annual Prediction is now in progress to quantify the improvements in forecast skill resulting from proper inclusion of the stratosphere. Some centres (for example the UKMO) already run with a well-resolved stratosphere including such effects as the low frequency QBO and others plan to introduce similar improvements in the coming year.

Until recently the atmosphere and ocean have been analysed separately and in many centres this is still the case. There may be benefits, such as reduced initialization shock if the two media were analysed together. There could also be some advantage in being able to use cross covariances between the two media during assimilation to better deal with lack of observations. There is some progress towards coupled atmosphere-ocean data assimilation but it is mainly in the sense of weak coupling such as using the coupled model to give the first guess at 6 hour intervals. The atmosphere and ocean are then analysed separately. Stronger coupling requires a method of addressing the different timescales in the atmosphere and ocean. For long windows this requires temporal smoothing or damping on the atmospheric analysis. Ocean model analysis has mainly been using OI but there is a move towards 3dvar or ensemble Kalman filtering of one sort or another.

Current capabilities in Sub-seasonal to Seasonal Prediction

1. Introduction

The main goals of this workshop are to establish current capabilities in sub-seasonal to seasonal prediction, to identify high-priority research topics and demonstration projects and to develop recommendations for the establishment of an international research project. There is a further assumption that it is necessary to develop a seamless approach to weather and seasonal prediction in order to improve predictions at these ranges. The latter point may not necessarily be true at this stage for providing forecasts for different time-scales with best possible skill but exploring the pros and cons of having a unified approach is desirable.

Forecasting at the synoptic to medium range is primarily an atmospheric initial condition problem, although there can be an influence from land and ocean temperatures. Forecasting at the seasonal range depends strongly on the slowly evolving boundary conditions, primarily SST but land, snow cover and sea-ice can all influence the solution. The atmospheric circulation after a season may not be entirely a boundary forced problem, however, and atmospheric initial conditions could be important in preconditioning the evolution of SST. In any event, it is advisable to initialize the various components of the model listed above as accurately as is reasonably possible.

In the medium range it is feasible to predict the evolution of synoptic systems, but as the forecast range gets longer our ability to see the future becomes fuzzier. The details of synoptic variability are not predictable as the influence of chaos takes over. However, this can be offset to some degree by averaging either in time, for example weekly means, monthly means, seasonal means, or in space, using longer time and larger space averages at longer time horizons.

There is clear evidence that tropical SSTs influence not just the tropical weather but also, by tele-connections, various regions of the extra-tropics. In the mid-latitudes the signal to noise is lower than in the tropics because of the greater synoptic variability not directly linked to SST variability. There are regions or patterns where the variability is linked, however such as the PNA (Pacific North American pattern) linked to ENSO SST variability. Much of the potential predictability on forecast horizons of weeks to months is linked to ENSO in the Pacific and SST variability in the Indian Ocean. One might expect that the equatorial Atlantic should also induce predictability but this has been harder to realize, possibly because of model error there, although it is very likely that even if model error were reduced, the predictability horizon would still be less than in the Pacific.

Over the years there has been steady progress in improving the skill of forecasts for the medium range as well as at the seasonal range. The objective of the current initiative is to increase the forecast range from the medium range but not as far as to the seasonal range. ENSO and the variability in the tropical Indian Ocean, sometimes loosely called the IOD or Indian Ocean Dipole (Saji et al. 1999, Webster et al 1999) are important sources of predictability on the timescales of interest here. Given the objectives of the meeting and the strong thrust within WMO to develop seamless prediction systems, one way would be to start from the NWP end and consider what needs to be done to extend the range. The alternative would be to start at the seasonal end and consider how to adapt these to shorter range. The resolution tends to be considerably lower for the seasonal forecast systems and these are less likely to do well in the intra-seasonal range. On the plus side, however, they all have the atmosphere coupled to an active ocean. Perhaps the development of systems with variable resolution is a way of getting the best of both worlds, putting the extra resolution in the short to medium range where it most likely to deliver benefit and then reduce the resolution, but perhaps increasing the ensemble size at the longer forecast range. There are several components to a useful forecast system on the range 1 to 100 days. Issues relating to model development, to analysing and initialising the atmospheric, oceanic and land states, and to post-processing the output will be discussed. The intra-seasonal to inter-annual prediction problem is also discussed in Toth et al (2007), Weller et al (2010), Brunet et al (2010), Shapiro et al (2010). The National Academy of Sciences has just recently published a book on seasonal forecasting, Assessment of Intra-seasonal to inter-annual climate prediction and predictability, available for free download from www.nap.edu.

Predictions are necessarily uncertain as the models used to make them are flawed, and the initial conditions from which the forecasts are made are flawed. Model error is one of the most pressing problems in extended range forecasting. One way of dealing with the various uncertainties is to forecast a PDF (Probability Distribution Function) of some variable and compare it with some reference PDF, ideally representing the true climate, but often a suitable model climate is used as a surrogate. By comparing the forecast PDF with the reference PDF one gets a sense of predictability, but it is not until one has an application in mind that one knows if the forecast is really useful. The most usual situation is where the mean of the PDF of the forecast is shifted with respect to the mean of reference pdf. However, it is perfectly possible to get a PDF whose mean is not shifted with respect to the reference PDF but for the forecast still to be useful. Likewise it is possible to get a PDF whose mean is shifted but for the forecast not to contain any useful information for a particular application. See also Palmer 2006, Kumar 2009, 2007.

From an applications point of view, it is assumed that improved weather forecasts will lead to improved socio-economic forecasts and applications. It is likely that this will involve the development of an application model to be run over the meteorological output, which is a significant task in its own right. For example, Thomson et al (2006) successfully applied a malaria prediction model to output of the DEMETER multi-model seasonal forecast system, but one lesson from this exercise and from using the meteorological output to drive crop models is that adapting the application model to run off the meteorological model required considerable effort. However, in this paper we will concentrate more on issues related to improving the meteorological forecasts. This undoubtedly involves model improvement through increased resolution and improved parameterisation but also developments in data assimilation, and post-processing of the meteorological output to correct for model error.

2. Improvement in model forecasts by post processing

2.1 Dealing with model error through post processing and the need for reforecasts

Many national Met services use Model Output Statistics (MOS) to correct systematic errors in the model predictions, particularly for site-specific forecasting. The MOS approach entails correcting current forecasts based on a training data set of errors in previous forecasts, ideally based on up to date version of the NWP model. See Wilks and Hamill (2007) for a discussion of some methods for MOS. Often the training data set is short, or, if long, it is cobbled together from various model cycles and therefore not homogeneous. One approach that is increasingly used for correcting shorter range forecasts is to use a Kalman filter MOS (KFMOS), in which the error corrections are continually updated to track changes in the model error characteristics, and flow-dependent variations in model error. However, that approach is not readily applicable to the longer range; the longer the forecast range, the more out of date is any estimate of model bias based on recent observation minus forecast differences, but see Delsole and Shukla (2010). For longer ranges, any training data needs to be based on historical training data; a good hindcast data set would be a firm basis for the application of MOS techniques to correct systematic model errors.

TIGGE is a major experiment to accelerate the improvements in the accuracy of 1-day to 2 week high-impact weather forecasts for the benefit of society (Bougeault et al 2010). A data base of medium range ensemble predictions from ten of the leading global NWP centres was created with the aim of supporting research on ensemble forecasting focused on improving the prediction of high-impact weather for the one-day to two-week time-range. With the focus on those time ranges, the forecast skill is not dominated by systematic model errors, so the emphasis of TIGGE was on collecting the real-time ensemble forecasts and not retrospective forecasts. That said, the TIGGE project recognised that the calibration of ensemble forecasts, to correct for model biases or for downscaling, is a key topic for research that could be carried out using the TIGGE data set. Another active research area facilitated by the TIGGE data is the combination of ensemble forecasts from multiple models to construct a multi-model ensemble (MME). While some cancellation of errors may be achieved by simple combination of forecasts from different models, a more informed approach is needed - especially when dealing with more extreme events. Johnson and Swinbank (2009) used a simple bias correction scheme using recent analysis minus forecast differences in three ensemble forecasts, and looked for the additional benefit of combining the forecasts in an MME. This study showed the benefit of MME for forecasts of 2m temperature, but only marginal benefit for mean sea level pressure or 500 hPa geopotential height. Matsueda and Tanaka (2009) showed that there was some benefit from MME for 500hPa geopotential heights, but they did not apply any bias correction.

One participating centre additionally carried out a large set of reforecasts enabling calibration based on those data to be compared to the benefits of MME; results demonstrated similar benefits are achieved using both MME and reforecast bias correction. In principle, one could construct a MME of reforecast-calibrated forecasts, but this seems not to have been done to any great extent in the context of NWP, although there has been some work in this line in seasonal forecasting (Coelho et al 2006, Stephenson et al 2005). It has not been done so far in the context of TIGGE, as there was only one model with an extensive set of reforecasts, though this could change with the hindcasts from the NCEP reanalysis. The cost of producing reforecast data sets has acted as a strong disincentive. Another disincentive is that most NWP centres do not have reanalysis data that could be used to initialise retrospective forecasts consistently with the real-time forecasts; systematic differences in initial conditions (including surface as well as atmospheric fields) could be misinterpreted as systematic model errors, especially in the initial stages of the forecast. Nevertheless, if NWP and seasonal prediction systems are brought into line, there is a good opportunity for medium-range forecasts to benefit from the reforecasts and calibration methods that are being carried out for longer-range prediction.

In seasonal forecasting, it was recognised that model error, especially drift was a serious problem right from the earliest days of using comprehensive general circulation models for prediction, and further that model error seriously reduced potential long range skill, since the size of model drift was as large as the signal being predicted. One way of dealing with model error was to correct it a posteriori.

By making forecasts not just for the present and recent past, but for as long in the past as is practical, creates a data set from which corrections to the model output could be made. The back integrations, also sometimes called hindcasts or reforecasts typically go back to 1982 for operational systems. In non-operational mode, reforecasts have been carried out at 3-monthly intervals for the much longer period from 1958, essentially for the period for which atmospheric reanalyses are available. The ENSEMBLES project (Doblas Reyes et al 2009, Weisheimer et al 2009) is a good example of this extended period. CliPAS (Wang et al 2009) extended the DEMETER range of models with another 7; the experience from ENSEMBLES, however, was that adding the DEMETER models to the newer and better ENSEMBLES models did not bring increased skill (Weisheimer et al 2009).

The a posteriori correction, commonly used in seasonal and monthly forecasting, is essentially a linear correction and one might question its validity in a complex nonlinear system. However, it seems to have performed rather well, though, of course, it cannot correct all error (Stockdale et al 1998). An alternative strategy is to try to correct model drift through a priori correction by Flux Correction (FC). This should most likely be applied to surface winds since equatorial wind errors strongly influence the tropical ocean, but corrections to surface heating and fresh water are also possible. JMA uses flux correction in their seasonal forecast system (Takahashi, ECMWF seminar series, Sept 2010). It has also been tried in the Australian seasonal forecast system POAMA24.b using both heat flux and stress corrections. Some things such as the ENSO tele-connection to Australian rainfall are improved, but only slightly. The fundamental problem in the Australian case is that FC does not overcome the intrinsic atmospheric model error which is a tendency to make a double ITCZ, with too little convection in the eastern Indian Ocean. Flux correction has also been tried at ECMWF but never implemented (Stockdale private communication).

The reason for building up a large set of back integrations is twofold. It is partly to define the model climatology as a function of forecast lead and start date. Forecasts can then be referenced to this model climatology and anomalies calculated. Removal of model error is not the sole consideration, however. A skill assessment is essential. Seasonal forecasting traditionally concentrated on predicting El Nino and La Nina, the warm and cold states of the tropical Pacific, and of the global tele-connections emanating from anomalous behaviour in the equatorial Pacific. El Nina/La Nina events typically occur every few years, with big events being rather rare; the1982/3 and 1997/8 El Ninos are the only two major warm events in the last 30 years. To judge if a forecast system is reliable requires forecasts of as many realisations as possible and confidence that it will not forecast events falsely. Although the timescale envisaged in this report is less than the seasonal range (6 – 12 months), it is not markedly so and therefore a significant range of hindcasts is required to validate the system.

2.2 Outline of the ECMWF seasonal and monthly forecast systems as examples of possible operational systems

To focus the mind, it might be useful to consider a specific system. To that end, a brief description of the ECMWF seasonal forecast S3 system is given. (S3 is still the operational system, though it was developed in 2005 and implemented in 2006, and is therefore somewhat dated and so unlikely to be top of the range. Stockdale et al 2011, Anderson et al 2007) An ensemble of 11 members for every month for the 25 years 1982-2006 was generated. Each reforecast extends to 7 months, with some extended to 13 months. The back integrations or reforecasts are not a small component of a forecast system. The EC S3 needs nearly 160 years of coupled integration for reforecasts for every month of the year (i.e. nearly 2000 years to cover all 12 months). Consequently the atmospheric model resolution is considerably lower than that used in the EPS -- T159L62 in seasonal compared to T399L62 as used in the EPS at the time of implementation of S3 and T639 (32km) now. The ocean resolution is one degree but with a refinement in the equatorial region to 1/3rd of a degree in order to resolve the important equatorially trapped Kelvin and Rossby waves.

The real-time cost of the forecasts system is considerably less than that of the reforecasts: for real-time forecasts the ensemble size is 41, so for a 6-month forecast 20 years of coupled integration are required, compared to 160 years for the hindcasts i.e. the reforecasts cost 8 times as much as the operational forecast. Because the back integrations make the system so expensive, seasonal forecast systems have tended to be upgraded much less frequently than NWP systems; for example at ECMWF it is typically 4-5 years between upgrades (and at NCEP it is more than 6 years since their last upgrade). This means the reforecasts only need to be done once every 4 to 5 years. The cost of the back integrations can be reduced by reducing the hindcast period or the ensemble size as is done at JMA and UKMO. By running most of the hindcasts in near real time the Met Office now upgrades their system much more frequently (Arribas et al, 2011). However, as the need for calibration of NWP systems grows, it is likely that upgrades to NWP systems will also become less frequent and so the gap between NWP procedures and seasonal procedures may narrow somewhat.

When the ECMWF monthly forecast system was first developed and introduced operationally in 2003/4, it sat somewhere between EPS (10 day) forecasts and seasonal but was not operationally linked to either. The coupled model structure was the same as used in the seasonal system though the atmospheric model was not. The ocean model was the same and the ocean analysis used to provide the ocean initial conditions was essentially the same as for the seasonal forecast system though slightly modified to allow more timely ocean analyses. There were two differences, however: firstly, the diagnostics produced by the model were partly adapted from the EPS and partly from the seasonal suite, secondly the monthly forecast system was updated every time there was an update to the EPS system. This requires that the reforecast set had to be performed more often. Some reduction in cost was made by reducing the period of the reforecasts to 12 years. The reforecasts were made out to 32 days, the ensemble size was only 5 members, and forecasts were made once per fortnight requiring 10 coupled model years for each month, much less than that of the seasonal system. However, the system was upgraded more frequently, the resolution was higher and the frequency of monthly forecasts was soon increased to weekly making the system no less costly. In principle this approach can be used for seasonal forecasting – as, for example, is now done at UKMO, allowing a seamless NWP-monthly-seasonal forecasting system. The cost can be substantial; however, unless a reduced set of reforecasts is produced in which case the assessment of skill can be compromised.

2.3 Calibrated forecasts and multi-model forecasts for the extended range

Since reforecasts are essential for monthly forecasting, they can be used for NWP also. This has already been pointed out in sectionn 2.1 in relation to MOS. Hamill, et al (2006) have argued that reforecasting can be useful not just for removing first order model drift but also important for increasing model spread. Interestingly they found greatest advantage for the latter correction when applied to near surface variables such as T_{2m}. This may be in part because the reforecasting when compared to observations can act to downscale the forecast for terrain effects. The improvement in skill was mainly due to improved resolution rather than reliability, though this may be model dependent. {Resolution measures the ability of the forecasts to distinguish between situations with different observed event frequencies whereas reliability indicates how closely the observed event frequency matches the forecast probability namely forecasts with an assigned probability of occurrence of 20% should occur 20% of the time etc.). Although the model used in this study was far from state of the art, either in model cycle, being of 1998 vintage i.e. 6 years behind the then-current NCEP model version, or in resolution, Hamill et al 2006 show that their 2-week forecasts were more accurate than the NCEP operational forecasts. Reforecasts have become feasible of recent times since there are extensive reanalyses from which the reforecasts can be started. Ideally one would like the model used for the reforecasts to match that used for the reanalysis but that is unlikely to happen for a long time. Hamill et al (2006) make the point that generating reforecasts is an eminently parallel process. For the model they used, a cluster of personal computers was used. This idea could be extended to the widely used CPDN (Climate Prediction Dot Net) network, (Allen and Stainsforth (2002), www.climateprediction.net) provided the CPDN network can be expanded to deal with higher resolution models than currently used.

It is likely that correcting the model PDF will have advantages for forecasting extreme weather also, since the model PDF is unlikely to be as broad as that of nature. However, by referencing model events relative to the model PDF it should be possible to correct for this effect. Zsoter (2006), Gober et al (2008) give examples of extreme event forecasting.

Whitaker et al 2006 considered a multi-model combination of two calibrated forecast systems, that of NCEP and ECMWF. As in Hamill et al 2006, the former was a low resolution, rather old model but with reforecasts every day from 1979. The latter was a higher resolution model but with only twelve years of reforecasts, and then only started every two weeks. (This latter was the reforecast set of integrations originally developed for monthly forecasts. The ensemble set consisted of only 5 members compared with 15 for the NCEP-model set of Hamill et al 2006). The length of the ECMWF forecasts was 32 days but since the NCEP model set only extended to 15 days, comparison could only be made of this shorter period. Since there was a mismatch in model quality, in the length of reforecasts, in ensemble size and in forecast start dates, this was hardly a compelling test of multi-model calibrated forecasts. Nonetheless they considered a common set of 5 members each, for 7 start dates spanning winter (Dec-Feb) for 12 years. The forecasts were for T850, which were compared with the ECMWF ERA-40 reanalyses and also the NCEP-NCAR reanalyses. It made little difference which analysis set was used for verification. [This is not always the case. Park et al 2008 show an example where verifying short-range forecasts against its own analysis is beneficial, but the variable chosen, T 850 in the tropics, had substantial bias between different analysis systems, presumably because there were insufficient data to adequately restrain the analysis]. The un-calibrated forecasts had poor reliability. The skill of both systems was improved by calibration. Although the ECMWF system was considerably more accurate than the NCEP system, it still benefitted substantially from calibration. The multi-model forecast based on a combination of the two calibrated systems was a further improvement on the ECMWF calibrated system. Thus, although there was a considerable mismatch in the quality of the two systems, a calibrated multi-model product was better than either system alone. These results were based solely on T850 averaged over forecast days 8-14. Would a similar improvement hold for a surface variable such as T_{2m}, or for a more chaotic variable such as precipitation?

The skill advantage in using reforecast runs for T_{2m} has recently been considered by Hamill and Whitaker (2007) though only for a single model, and not in a multi-model context. In addition to T_{2m}, they considered Z500, and T850. All three variables are approximately normally distributed and so several methods for calibration might work. They chose to test two corrections, one that eliminates bias and a second which corrects for spread deficiencies as well as bias. Since most systems do not account for model error, the spread is too small and the forecasts overconfident. Correcting for spread should mitigate this over confidence. They found that Z500 was the variable least improved by calibration and most of the improvement came from bias correction. On the other hand forecasts from the raw ensemble were least skilful for T_{2m}, application of bias correction substantially increased the skill but the application of correction for spread produced the largest improvement relative to the bias correction. A possible conclusion is that surface variables might be those that benefit most from calibration.

Hagedorn (2008) has considered calibration in the context of the TIGGE forecasts. An extensive set of reforecasts, consisting of 15 members out to 15 days every week at T255 resolution, was carried out for the years 1982-2001. Note that the version of the model used predates that which has an improved MJO, a significant development which will be considered later. The real-time TIGGE ensemble consists of 51 members. The initial conditions come from ERA-40. Two calibration methods were tried: simple bias correction, adjusting just the mean of the PDF and a nonhomogeneous Gaussian regression (NGR) which additionally adjusts the spread of the ensemble. The only field tested was T_{2m} but Hamill and Whitaker (2007) found that surface fields benefitted most, though this conclusion was based on a model of relatively low skill. Hagedorn was able to subsample the 15 member ensemble to look at the sensitivity of results to ensemble size. For short-range forecasts, an ensemble of 5 seemed adequate as neighbouring weeks were used to augment the ensemble, but as the forecast range increased it was necessary to use a larger ensemble. Because there was a reforecast ensemble, it is possible to assess the relative merits of calibration compared to a multi-model ensemble, which also acts to increase the ensemble spread and prevent the forecasts from being over confident.

Ten models contributed to the TIGGE multi-model ensemble data set but only 9 produced forecasts out to 9 days, of which four extended to 15 days. Hagedorn (2010) has shown that the Multi Model procedure was no more skilful than the most skilful single model forecast system of all the 9 systems considered (those which were producing forecasts of 9 days or longer). However, if the multi-model was restricted to using just the four best models then some advantage was gained over using only one model. Thus the multi-model is a potentially useful method for extended range forecasting. The four best models in fact produced forecasts to 15 days and so a comparison of the multi-model with the calibrated model forecasts out to 15 days was possible. The four-component multi-model gave significantly better results than the 9-component and was an improvement on the best un-calibrated single model. However, calibration achieved skill comparable to that of the multi-model. No calibrated multi-model was possible as no other model had a reforecast set on which to base calibration. An advantage of the calibration method is that it can partly take downscaling into account which is advantageous especially in regions of complex topography. A disadvantage is that the fields are not spatially synoptically coherent. The approach has not been tested on TIGGE data for fields such as precipitation or surface wind but should be.

An alternative approach to calibration is given by Palmer et al (2008). The underlying idea is that shorter range processes can be validated better than longer range as there are more realisations of them. If model a does a better job of representing shorter range processes than model b, then it could be better believed for longer-range forecasts. As shown above, the multi-model often is more accurate than any single model, especially at longer time horizons. Further disagreement between models would be interpreted as low confidence in the prediction and agreement between models is often used to increase confidence in their predictions. However, if the models had systematic errors then the latter might not be true; consensus is not necessarily an indication of correctness. Palmer et al (2008) use the seamless idea to illustrate how confidence in model behaviour at one time scale can be used to qualify the confidence at another. [This idea is not universally accepted. The reader might also want to read the comments by Scaife et al (2009) as well as the reply by Palmer et al.]. The particular example they chose was to use seasonal forecasts to qualify the predictions of anthropogenic climate change (ACC). Ideally the same models should be used for the different timescales; this was not true in their case, so their results are illustrative rather than definitive. They showed how reliability diagrams from multi-model ensemble of seasonal climate forecasts of regional precipitation can be used to correct the ACC predictions for various regions. The ACC models suggest that the occurrence of wet June July August will substantially increase, but if one takes into account the poor representation of the monsoon in seasonal forecast models and by implication in the same models when used for ACC forecasts, then the signal is considerably down-weighted. There are various caveats not listed here but the idea based on seamless systems is worth exploring further. One way to do that would be to use say an analysis of the reliability of monthly forecasts to calibrate the forecasts skill for seasonal forecasts, or, perhaps of greater relevance to the objectives of this conference, to use the medium range to calibrate monthly.

3. The intermediate timescale between medium-range and seasonal

3.1 The tropical influence

Considerable attention has been given to the weather timescale, say out to 10 days as well as to the longer El Nino timescale, say 6-12 months. For the former the main source of information is in the atmospheric initial conditions, for the latter much resides in the oceanic initial conditions. For the former, interaction with the ocean is not normally dominant, though it might be important in certain situations, for example slow-moving tropical cyclones. For the latter it is essential to have an active dynamical ocean. El Nino prediction is predicated on the assumption that slower oceanic processes control the evolution of tropical convection, shifting it from the maritime continental area towards the central Pacific. This involves oceanic Kelvin and planetary waves with timescales of weeks to months. It does not preclude the importance of intermediate time scales such as those associated with the MJO. In fact Shi et al 2010 show that even at three month lead time there is a memory in the evolution of ENSO stemming from the initial atmospheric state. (See also the references therein.)

Is there predictability in the intermediate time scales? One of the potential sources of predictability comes from the MJO. This has a natural timescale somewhere in the range 30-70 days. It is associated with regions of enhanced or reduced precipitation, and propagates eastwards, with speeds of ~5m/s, depending on its longitude, somewhat faster than the equatorial oceanic Kelvin wave. Interaction with the ocean may play some role in its propagation and development but does not appear to be crucial to its existence (Woolnough et al 2007, Takaya et al 2010). The way convection is represented in numerical models seems to influence the characteristics of the MJO quite strongly, however. Until recently it was poorly represented in most models. There are now some models which have something resembling an MJO.

(Vitart and Molteni 2009, Waliser et al 2009, Shi et al 2010, Wang et al 2010 and Gottschalck et al 2010).

The MJO clearly influences precipitation in the tropics. It influences tropical cyclone activity in the western and eastern north Pacific, the Gulf of Mexico, southern Indian Ocean and Australia. See Vitart 2009 for a list of references. It also influences the Asian and Australian monsoon onset and breaks and is associated with northward moving events in the Bay of Bengal (Lawrence and Webster 2002). The prolonged dry period in the Indian summer monsoon of 2002 might be associated with an absence of northward propagation of MJO activity (Saith and Slingo (2006)). Not only is the MJO important in the tropics, there is growing evidence that it has an important influence on northern hemisphere weather in the PNA (Pacific North American pattern) and even in the Atlantic and European sectors. Lin Brunet and Derome (2009) have studied the link to modes of the northern hemisphere including the North Atlantic Oscillation. They find a lagged response with the MJO leading the NAO, though they also find equator-ward propagation in the north Atlantic upper level (200mb) winds. Cassou (2008) has also related the MJO to the NAO and estimated the phase lags. The MJO has also been found to influence the extra-tropical weather in various locations. For example, Higgins et al. (2000) and Mo and Higgins (1998) investigated the relationships between tropical convection associated with the MJO and U.S. West Coast precipitation. Vecchi and Bond (2004) found that the phase of the MJO has a substantial systematic and spatially coherent effect on sub-seasonal variability in wintertime surface air temperature in the Arctic region. Wheeler et al. (2009) documented the MJO impact on Australian Rainfall and circulation. Lin and Brunet (2009) and Lin et al. (2010b) found significant lag connection between the MJO and the intra-seasonal variability of temperature and precipitation in Canada. The importance of the tropics in extra-tropical weather forecasting has been illustrated by several authors (see Jung et al 2010). Early results from Ferranti et al (1990) indicated that better representation of the MJO lead to better mid-latitude forecasts in the northern hemisphere. This has been recently revisited by Jung et al (2010) using a much more recent cycle of the atmospheric model and at considerably higher resolution. The benefit of the connection of the MJO and NAO in intra-seasonal forecasting was demonstrated in Lin et al. (2010a), who analysed the output of the intra-seasonal hindcast experiment conducted with the GEM global atmospheric model during 24 extended winters. It is found that with a lead time up to about one month the NAO forecast skill is significantly influenced by the existence of the MJO signal in the initial condition. A strong MJO leads to a better NAO forecast skill than a weak MJO. These results indicate that it is possible to increase the forecasts kill of the NAO and the extra-tropical surface air temperature with an improved tropical initialization, a better prediction of the tropical MJO and a better representation of the tropical-extra-tropical interaction in dynamical models.

Rashid et al. (2010) estimated the potential predictability of the MJO in the POAMA model to be ~40 days (or longer), by measuring how well the POAMA model can forecast itself, subject to small initial perturbations (i.e. using a perfect model assumption whereby one forecast member is considered truth and an ensemble mean forecast is formed with the remaining members). The POAMA model seems to do a modest job of representing the MJO and these predictability estimates might therefore be quite reasonable.

All the above suggests that there might be useful predictability on the MJO timescale, not just in the tropics but in the extra-tropics also. Has this potential predictability been confirmed? Until recently the answer would have been no, as the MJO has been rather difficult to simulate in models. One might think this could be resolution dependent but it is not obviously so. As one outcome of the ATHENA project where a model has been run at various resolutions (125km, 39km,16km, and 10km), Jung (2010) finds that the MJO is not substantially improved at the highest resolution, suggesting that parameterisation remains an important issue. Indeed the signs for this have been there for some time; it has been found that by imposing a CAPE threshold on parameterisation, more MJO-like events develop in models which do not have a particularly good representation of the MJO otherwise (Vitart, Palmer personal communication). It can also influence the speed of eastward propagation. The CAPE threshold might induce other less desirable features, however, and so is not a solution to the MJO problem but does suggest parameterisation is an important issue.

Vitart and Molteni (2009) have studied the prediction of the MJO and the extra-tropical tele-connections. This involves an extensive set of reforecasts with a coupled atmosphere-ocean model. Most importantly it is done with a version of the model with a much improved representation of the MJO. In particular, deep convection is made sensitive to environmental moisture instead of being controlled by large-scale moisture convergence (Bechtold et al 2008). A series of 46-day reforecasts from the 15^{th} of each month from 1989 to 2008 has been completed. The ensemble size is 15. The model resolution matches that of the then monthly forecast system viz.TL399 for the first 10 days and T255 thereafter out to day 46. They find that the amplitude of the MJO is maintained throughout the integration period; in fact it is a little too strong, though this has been improved in later versions of the model (not used in this study). This is contrary to earlier versions of the model when the amplitude decreased by 50% in the first 10 days. Using the Wheeler and Hendon (2004) MJO index they find that they can predict the evolution of the MJO for about 20 days. This is still quite a bit short of the predictability estimate of 40 days by Rashid et al 2010 but given the difficulty in representing the MJO, and the possibility of future improvements from better parameterisation of convection, it is perhaps not too unreasonable to expect some skill improvements at these longer lead times. They find most of the MJO tropical precipitation links such as precipitation over South America and equatorial Africa are well represented. In the extra-tropics the delayed links to the NAO are consistent with those observed by Lin et al 2009 and Cassou 2008 though they seem to be too weak, but overall the tele-connections are consistent with those found in ERA-interim. There are still deficiencies in the representation of the MJO, however. The propagation speed tends to be smaller than observed, on average and there is a tendency for MJO events to stall over the maritime continent. See also Jung et al 2010.

From the above set of reforecasts they also assessed the extra-tropical forecast skill. If there is an MJO in the initial conditions, then the skill in days 12-18, 19-25 is enhanced. When there is no MJO in the initial conditions there is no skill in the 19-25 day range whereas the forecasts are reliable if there is an MJO. The implication is that improving the MJO will lead to increased skill in this time range in middle` latitudes. Although the atmosphere is coupled to a dynamical ocean (after day 10), Woolnough et al 2007, have shown that coupling to a mixed layer with a thin top layer (1m compared to 10m in the OGCM) may lead to faster MJO propagation. This may be significant, as Vitart and Molteni (2009) argue that the greatest weakness in the model is the slow propagation speed and that this leads to too weak a tele-connection in the Atlantic European sector. The importance of the MJO to extra-tropical prediction is consistent with the results of Jung et al (2010) who found through relaxation experiments that part of the medium and extended range prediction skill originates in the tropics in association with the MJO.

Weigel et al (2008) have considered the verification of the ECMWF monthly forecast system. They considered weekly averages of T_{2m} from the first version of the operational system, using 2006 as the forecast year, with 51 ensemble members. The hindcast has only 5 members and so the ensemble set consisting of the reforecasts and the real-time forecasts for 2006 is not of uniform size throughout. To overcome this they used a de-biased rank probability skill score as a measure of skill which they claim has the advantage of being insensitive to the unreliability arising from small ensemble sizes. They found that there was skill out to about 18 days, but there were occasions when the forecasts were useful to longer periods. Filtering out unpredictable high-frequency weather noise by using longer averaging periods also increased the skill. Interestingly they found that the forecasts are essentially reliable and argue that recalibration would not significantly improve the forecasts, (contrary to the results of Whitaker et al (2006), though different averaging period and variables were used).

Vitart and Molteni (2009b) have also looked at some specific extreme events using the same model runs as described above and compared them with other runs of the same model but lower resolution. They find better forecasts of the 2003 heat-wave over Europe and the wet July 2007 over the England, better representation of the frequency of hurricanes and better prediction of early monsoon rainfall, the improvement coming in part by increased resolution and in part by a more recent model cycle. Improvements in monthly mean 2m temperature skill for days 15-46 relative to the seasonal forecast system were also noted.

A further analysis of these reforecasts has been used to explore the link between tropical storm prediction and the MJO. (See the papers cited in Vitart 2009 concerning the observational basis for this link). How well this can be represented in forecast models has not been previously considered in any depth since extended range forecast models have not been able to represent the MJO well enough. Prior to cy32r3, too few tropical storms formed in the Gulf of Mexico, but not only does cy32r3 give a better representation of the MJO, it leads to a better distribution of tropical cyclones, though the frequency of intense tropical cyclones and ACE (accumulated cyclone energy) is under-predicted. The analysis is of forecasts over the last month of the 46-day forecasts. The anomalous tropical cyclone density moves eastward with an MJO but the relationship in the Gulf of Mexico and tropical Atlantic is weaker in the forecasts than in the observations. In the North Pacific the agreement is good while in the northern Indian Ocean the model overemphasises the relationship. Apart from the Gulf of Mexico, there is good agreement between observed landfall and the predicted risk of landfall with the phase of the MJO. Taking tropical cyclones as extreme events, then this result together with the results in Vitart and Molteni (2009) suggests that there could be useful skill in predicting extreme events, one of the objectives of this workshop. Hudson et al (2010) show that ENSO, the IOD and SAM are all “sources” of intra-seasonal predictability: if ENSO/IOD/SAM are in extreme phases, intra-seasonal prediction is extended. They argue that it is not predicting intra-seasonal variations in the tropics per se that matters, but that these slow variations tilt the PDF one way or the other and this tilt can be detected as short as 2 weeks into the forecast.

Although consideration has been given to the importance of the MJO, one should not forget the importance of SST anomalies, particularly those associated with ENSO and the IOD or at least the SST variability at the two sides of the Indian Ocean. Skill has improved in our ability to predict the SSTs though the tele-connections to mid-latitudes are only starting to be represented (for example Ineson and Scaife 2008).

3.1.1 A virtual field campaign to improve representation of the tropics

The YOTC project can be considered as a *virtual global field campaign,* where "virtual" means that the atmosphere and its variability are not sampled through specialized and typically costly field campaigns, but instead are represented by state-of-the-art resources available from global weather prediction models and multi-sensor global satellite measurements. Advantages of the virtual approach are cost effectiveness and flexibility. It is cost effective because the data exist in operational weather centres. It is flexible because the database can be improved as new data-assimilation procedures become available, as more data are incorporated in the assimilation procedures, as models progressively attain higher resolution, and as physical parameterizations improve.

The ECMWF has provided such a database, where the Year in this case encompasses May 1 2008 to April 30 2010. The YOTC database includes a full (4 times daily) analysis, forecasts and special process-related diagnostics from the T799 (25 km) Integrated Forecast System (T1279 or 15 km from January 2010). The weaker MJOs during the early part of the Year gave way to stronger events in October 2009 – March 2010, consistent with a change from La Niña to El Niño climate states. The Year sampled a full ENSO cycle, and an Arctic Oscillation. The US National Centres for Environmental Prediction (NCEP) and the NASA Global Modelling and Assimilation Office (GMAO) also provide databases from their prediction systems.

The above high-resolution deterministic databases provide information on parameterized processes and the associated scale-interactions. This complements the probabilistic TIGGE database. Consequently, the YOTC project and its research component are vital new resources for improving the representation of the tropics in global models as well as the influence of the tropics on global weather and climate.

3.2 The role of Land initialisation

The influence of the land surface on weather prediction through boundary layer evolution and precipitation has been considered by, inter alia, Betts and Viterbo (2005), Fennessy and Shukla (1999). One possible feedback is that low soil moisture leads to higher surface temperature and reduced evaporation leading to reduced convection; boundary layer clouds may also be important. NWP models have land models of varying complexity. Variables for initiating the land component are not easily observed. This is especially true if one needs a long record or reanalysis. Errors in P or E can give rise to errors in soil moisture. Land data assimilation systems adjust the land model's soil moisture reservoirs in response to observations of atmospheric temperature and humidity using Optimal Interpolation (OI). If simulated relative humidity is too low compared to observations, soil moisture is increased so that evaporation increases, thereby increasing the simulated humidity. While this approach for initialization has been used with success to improve weather forecasts, errors in simulated relative humidity and temperature need not stem from errors in soil moisture; they could stem from errors in parameterization, so that the modified soil moisture contents may not be accurate. Drusch and Viterbo (2007) note that soil moisture profiles obtained through the OI approach are inaccurate and Ferranti and Viterbo (2006) showed that this approach caused serious errors for the extended range forecasts of the 2003 European heat wave. Wang et al. (2010) document errors in the initial specification of soil moisture leading to substantial errors on predictions on seasonal time-scale.

Douville (2004) considered the importance of initialising soil moisture in models, finding soil moisture to have limited influence on atmospheric predictability, except over North America. He did not find a clear signal over Europe. On the other hand Ferranti and Viterbo (2006) considering the very dry hot summer of 2003 over Europe found that the ERA-40 reanalysis did show dry conditions and this was evident from as early as March 2003 although the analysis, because of the land-atmosphere coupling severely damped the seasonal cycle and the anomalies. The reason for the damping is that an analysis increment was added to the soil analysis in line with the above mentioned OI correction strategy. In fact Dirmeyer et al (2004) found that the annual cycle of soil moisture in ERA-40 to be the smallest in all the reanalyses and climatologies he considered. Because the soil does not dry out as much as it should in late spring and summer, this impacts the quality of the forecasts made from these analyses. From a series of sensitivity experiments adjusting the soil moisture in the root zone (1m) and the deep soil layer (3m) Ferranti and Viterbo (2006) found that the response is a strongly nonlinear function of the initial soil moisture. The timescale on which anomalies in soil moisture affect the atmosphere can be up to two months, depending on the depth to which the anomaly extends.

Vitart (2005) showed that probabilistic scores indicated some potentially useful skill for the periods days 12 to 18 and 19 to 32 and for a case study of the 2003 heat wave over Europe Vitart and Molteni (2010) find that the model successfully predicted a risk of significantly warm temperatures for the month of August for forecasts started on the 15th of July. The above studies and others like them, such as the 15 years' worth of forecasts of Koster et al (2004) suggest that realistic soil moisture initialization can provide some increase in the quality of precipitation and air temperature prediction out to a month or more. In an attempt to generate a multi-model "consensus" view of how realistic land initialization affects forecast quality, several modelling groups are involved in the Global Land-Atmosphere Coupling Experiment (GLACE-2 project) in which participants perform two parallel sets of forecasts: one in which land surface states, particularly soil moisture, are initialized realistically and one in which they are not. A comparison of the skill derived from these two sets allows a direct quantification of the impact of land initialization on forecast quality. First results show that across the models, land initialization does improve the correlation skill of temperature forecasts out to 60 days, but for precipitation, the consensus is less robust. The results do show, however, some small land-related increases in accuracy for precipitation out to at least 45 days, especially when conditioned on the size of the initial anomaly (Koster et al., 2010), and especially in the north central United States. For precipitation forecasts, contributions to skill are much weaker but are still significant out to 45 days in some locations. Skill levels increase markedly when calculations are conditioned on the magnitude of the initial soil moisture anomaly.

A major source of uncertainty in validating land models is the strength of land-atmosphere coupling, the degree to which soil moisture variations affect variations in precipitation and air temperature. Such coupling strength cannot be measured directly with instruments (it can only be inferred indirectly at best), and the estimates of coupling strength quantified with modelling systems vary widely (Koster et al., 2006), indicating a substantial uncertainty in our knowledge of how best to model the relevant underlying physical processes such as evaporation, the structure of the boundary layer, and moist convection. Evaluating these individual components is thus important, but it is currently hindered by data availability. It is unclear whether model-generated evaporation fluxes respond realistically to soil moisture variations. Such gaps imply a need for joint model development and observational analysis, focusing on all of the physical processes connecting soil moisture to atmospheric variables.

3.2.1 Problems with land initialization

For models which do not have a soil reanalysis going back twenty years or so, which matches the soil model in their forecast model, one option is to use one of the extensive atmospheric reanalyses for the back integrations. For the real-time forecasts land conditions should come from the operational land scheme and so match the coupled model. An inconsistency will occur at the transition from reanalysis to operational conditions which will disrupt the statistics to some (unknown) degree. A longer-term solution would be to use a reanalysis to force an offline soil reanalysis using the appropriate soil scheme incorporating observations of precipitation where possible.

An alternative strategy has been developed by Hudson and Alves (2007), Hudson et al (2010) using their Atmosphere-Land Initialization (ALI) scheme. The 3D atmospheric analyses are nudged to pre-existing analyses, in hind-cast-mode or the Australian Bureau of Numerical Weather Prediction (NWP) in real-time. The land surface is left to adjust to atmospheric forcing. This allows the same land and atmosphere model to be used for initialization and coupled forecasts and also allows consistency between real-time forecasts and hind-casts. When the soil moisture “analysis” is compared to independent analyses over Australia it agrees well (Hendon, private communication). A specific example of need for consistency in the initialization of land conditions between hindcasts and real-time forecasts was noted in the NCEP seasonal forecast system where drifts in the analysis that provided the initial states led to unrealistic forecast anomalies (see Wang et al., 2010, Figs. 11 and 12).

In GLOSEA3 (the UKMO seasonal forecasting system until recently) use is made of the ERA40 and ECMWF operational analyses in initialising the land conditions for seasonal forecasts. Initialising GLOSEA in this way is undoubtedly putting it at a disadvantage with respect to forecasting near surface air temperature. To avoid this, in GloSea4 an anomaly initialisation approach is followed for the soil moisture (Arribas et al, 2011).

3.3 Snow and ice cover

Another consideration in the development of a monthly and seasonal forecasting system is the initialisation of soil moisture, snow cover and sea ice. Shongwe et al (2007) have shown that land conditions in spring can give considerable skill in predicting near surface air temperature. The relative importance of snow cover and sea ice has also been investigated by Alexander et al 2004 and Kumar and Yang 2003 and references therein. The low ice cover in the Arctic was investigated by Balmaseda et al (2010) and Kumar et al. (2010) who show that the ice anomalies in the summers of 2007 and 2008 had a significant impact on the atmospheric circulation over the Euro-Atlantic Sector, characterized by a high pressure over the Arctic (Greenland) and low pressure centres over Western Europe and Northwest America. The impact is similar for the two consecutive years, and it is consistent with the observed atmospheric anomalies. Results also show that the impact of the ice is strongly dependent on the underlying sea surface temperature. Results from partial coupling experiments indicate that the sea surface temperature over the Northwest Atlantic strongly affects the mean state of atmospheric circulation over the Euro-Atlantic sector and conditions the response of the atmosphere to a given ice anomaly. Grant (personal communication) performed a parallel study with the UKMO ocean model forced with atmospheric anomalies and recovered much of the ice variability. Neither considered the coupled problem, since well-validated and initialised sea ice models are not yet part of seasonal forecast systems. GLOSEA4, the UKMO system does have initialised and active sea-ice, as does the seasonal forecast system at Environment Canada, and initialised ice area (but not thickness) gives some seasonal predictability of sea-ice. Analysis of the effects of this in the coupled forecast context is being carried out with other centres in an initiative recently started under CLIVAR. See http://www.clivar.org/organization/wgsip/chfp/chfp_projects.php for further details.

3.4 Stratospheric Processes

Baldwin and Dunkerton (2001) showed strong apparent downward propagation of easterly and westerly anomalies from the stratosphere to the troposphere on monthly timescales. Importantly, this tends to be followed by easterly (negative NAO/AO) conditions in the troposphere. Perturbation experiments also reproduce negative NAO/AO in response to weakened stratospheric winds on both seasonal and longer timescales (for example Boville 1984, Norton et al 2003, Scaife et al 2005, Scaife and Knight 2008). Jung et al (2010) find that relaxation of the extra-tropical stratosphere to the observed state leads to forecast error reduction in the high latitude and European troposphere, but that the tropical stratosphere has no such impact. They caution the interpretation of these results, however, as the troposphere strongly influences the NH stratosphere and other studies suggest a role for the tropical QBO on the extra-tropical surface climate (Boer and Hamilton 2008, Marshall and Scaife 2009).

Scaife and Knight 2008 suggest that the stratospheric sudden warming in Jan 2006 contributed to the cold winter of 2005/6 in the NH and reproduced stronger surface NAO and cold European signals in simulations where stratospheric variability was imposed according to observations. The QBO was in a negative phase which could also have contributed. On the other hand, Jung et al (2010) suggest that the origins were in the tropical troposphere. While relaxation experiments can be used to suggest remote origins of anomalies in extended range prediction and give an idea of how much forecast skill could be gained by reducing forecast error in various regions such as the tropics, they are not definitive. Recent results from the prototype ECMWF S4 indicate improved results in seasonal forecasting by using an active stratosphere and Hendon et al (in preparation) show a small reduction in RMSE some 15-20 days into the forecast over the polar cap by better resolving the stratosphere. This is a high latitude effect and limited to 5% reduction, leading the authors to question the need for an active stratosphere in the Australian monthly/seasonal forecast system.

Although the jury is still out on the exact level of improvement to be expected from including stratospheric processes, and the stratosphere is most likely to contribute in winter and under sudden stratospheric warming events, some modelling groups are starting to include the stratosphere in their extended range forecast models. The UKMO system now uses an 85 level model which includes a comprehensive representation of the stratosphere for seasonal forecasting, and ECMWF is considering 91 levels for their System-4, based on results from a stratosphere-resolving prototype system which shows enhanced skill on the seasonal range. Other current systems do not fully include the stratosphere.

In order to assess the impact of stratospheric processes on predictability and prediction, CLIVAR has launched the Stratosphere resolving Historical Forecast Project (SHFP), coordinated by Adam Scaife, Ben Kirtman and Tim Stockdale.
Specifically, its purpose is: to quantify improvements in *actual* predictability by initialising and resolving the stratosphere in seasonal forecast systems; to compare

with existing seasonal to inter-annual forecast skill and to provide a hindcast data set that may be used to demonstrate improvements in currently achievable season forecast skill for a range of variables and lead times; to understand improvements under particular scenarios such as El Nino and years with an active stratosphere; and to justify changes in operational seasonal forecast approaches and methods. For more details see http://www.clivar.org/organization/wgsip/chfp/chfp_projects.php .

4. Model Developments

4.1 Parameterisation

Few could doubt the importance of parameterisation, and while the effect of every change to every parameterisation is not assessed, it is possible to garner some information on the major changes to the physics which have taken place over the years. I imagine most Centres will have conducted studies such as that of Jung et al (2010) although I am not aware of any publications. Although it is specific to the work at one Centre, it clearly documents the changes which have made the greatest impact, and also records what everyone knows only too well that progress is seldom uniform and monotonic. The convection changes of Bechtold et al (2008) clearly made a big improvement to the medium range and extended range forecasts through the improvement to the MJO and to the simulation of blocking in both the Pacific and European sectors. However, this scheme lead to stronger equatorial winds and to degradation in the seasonal forecasts.

Notwithstanding the very significant improvements that have been made, it is clear that errors in the representation of fast physics processes remain a key limiting factor in the skill of our models across all timescales from NWP to sub-seasonal to seasonal. Efforts to develop these representations are therefore of crucial importance for many applications, and full advantage should be taken of testing across timescales for example NWP or TRANSPOSE-AMIP short-range testing may provide insights into the sources of error for seasonal prediction which are hard to glean directly from study of the seasonal results. A key challenge for model analysis and diagnosis on any timescale is to provide insights which are detailed enough to help inform which parts of the physical parametrization are in error and thus aid model developers (rather than, for example, just saying that some process such as the MJO or ENSO is poorly represented). Studies such as those showing evidence for an unrealistic link between humidity and rainfall in many simulations of the MJO are a good example of this being done in practice. Palmer has argued that any forecast system should represent the uncertainty in model parameterisations. In the sense that many sub-grid-scale realisations are possible for a given set of macro conditions, a stochastic physics element has been introduced to several forecast models. In many parameterisations there are parameters whose precise values are unknown. Some representation of this can be included by perturbing parameters in the forecast. Another form of stochastic parameterisation, representing missing processes in the model is the backscatter of energy, viz. the transfer of energy from the small scales to the larger. Variations of this idea have been tested at ECMWF, at MSC, Canada and at the UKMO. The latter scheme called SKEB2 (Stochastic Kinetic Energy Backscatter) has been used in the UKMO EPS system, and gives better ensemble spread as well as an improvement in forecast skill and an improved forecast blocking frequency. See for example Houtekamer et al 2007, Hou et al 2008 and Tennant et al 2010.

Despite the inclusion of various forms of stochastic physics in the ECMWF model, the T_{2m} remains under-dispersive, perhaps suggesting the inclusion of a stochastic element in the land surface component of the model. Palmer et al 2009 suggest that by for example, cellular automata, some representation of sub-grid-scale processes could advect from one grid box to another in a way that is not possible in conventional convective parameterisations and that this may be an important missing process in the development of the MJO. Stochastic parameterisations can reduce systematic biases through nonlinear rectification. This is discussed in a nice pedagogical way in Palmer and Weisheimer (2009). Although many results relating to the role of stochastic physics are positive, Stockdale, private communication, finds that the inclusion of stochastic physics in the prototype seasonal forecasting S4, is negative. The reason is unclear.

Much of the uncertainty involving weather and climate derives from a lack of understanding of how the meteorological scales at the intersection of weather and climate -- meso (10-100km), synoptic (100-1000km) and planetary scales -- interact. This uncertainty complicates attempts to predict multi-scale phenomena (for example the MJO, ITCZ and monsoons). These phenomena influence the weather and climate of the mid and high latitudes through the pole-ward migration of meteorological systems out of the tropics (for example tropical cyclones) and through the long-range effects of organized tropical convection that extend to, or strongly influence, the extra-tropics *via* planetary-wave tele-connections.

A fundamental hypothesis of the YOTC project (Waliser and Moncrieff 2008) is that convective organization in the 10-100 km (meso-scale) range is a building block of larger-scale convective organization. This centres on an upscale transport of energy and momentum conspicuously absent from traditional convective parameterizations. The spatial resolution of most climate models and seasonal models is presently too coarse to resolve most aspects of convective organization. Moreover, convective parameterizations represent just one type of convection (cumulus) whereas tropical organized convection is a *discrete spectrum* of scales (for example cumulonimbus, meso-scale convective systems, super-clusters) involving convection-wave interaction. The defining elements of YOTC from the parameterization perspective are: that the upscale effects of convective organization in the 10-100 km range are building blocks of larger-scale convective organization, for example the MJO; that the self-similarity of ~1000 km super-clusters and ~100 km meso-scale convective systems (Moncrieff and Klinker 1997) evinces scale-invariant dynamics; that interaction between moist convection and tropical waves is a key element; and that the way in which traditional cumulus parameterization interfaces with explicit convective organization is an emerging issue for sub-seasonal to seasonal prediction. For more information see Moncrieff et al. (2010).

4.2 Model initialisation, including coupled data assimilation

Many forecast groups, be they at operational centres or in the research community, have the facility to run coupled atmosphere ocean models but none has an operational coupled data assimilation and initialisation system that can deal with the different timescales of the two media. Groups are beginning to work on this, however .

Currently, in operational practice, atmospheric analyses come from the operational weather analysis system and the ocean initial conditions from the ocean analysis system. The two media are analysed separately meaning that both may be close to the observed state but are not necessarily in a consistent state and initialisation shock is possible, even likely, when they are coupled, though this could become less as the models become more realistic. Some initiatives to weakly couple the two media are already in place.

On the atmospheric side, data assimilation schemes are usually 4d var, weak or strong constraint, long or short window, Ensemble Kalman Filter (EnKF) or some variant. On the ocean side, many are still OI though this is slowly giving way to 3d-var or EnOI (Ensemble Optimal Interpolation) or some variant of EnKF. Several groups operationally perform not a single ocean analysis but a small ensemble of analyses, not usually to estimate covariance but to create a set of ocean initial conditions from which to start a forecast. Interestingly this idea is now being pursued in the atmospheric case at Centres using 4d-var; See for example Buizza et al (2010). Those using EnKF already have an ensemble of atmospheric initial conditions.

Just as EnKF, or some variant, is being developed in the atmospheric case, so too it is being developed for the ocean. Rienecker et al. (2010) used an EnKF to assess the impact of different observing systems in seasonal forecasts using an older version of NASA/GMAO's (Global Modelling and Assimilation Office) system. However, no systematic comparison has been made with the operational OI. Yin et al (2010) use a form of EnKF at considerably less cost than a full EnKF in a scheme called PEODAS (POAMA Ensemble Ocean Data Assimilation System). The analysis is used to calculate the background error covariance with dynamically balanced increments. The PEODAS analysis is judged to be considerably better than the previous ocean analysis used at the BoM Australia which was an OI scheme but was deficient in that S was not updated following T analysis, whereas these corrections are made in their EnOI scheme. There are plans to extend this scheme into the coupled domain.

Davey et al (2006) sought to create ocean initial conditions using OI, 3d-var, 4d-var and EnKF methods and to generate a substantial set of seasonal hindcasts from them. The idea was that by inter-comparing these results, decisions on a preferred ocean data assimilation system could be taken. In the event no 4d-var or EnKF system was sufficiently mature, at that time, to allow a solid comparison, but 4d-var will be developed within the NEMOVAR consortium. Currently a 3d-var system is being comprehensively developed for operational use at a variety of resolutions by the NEMOVAR consortium and this will be used at the UKMO and at ECMWF in the near future. However, as the ocean model is also being changed no comprehensive set of analyses and forecasts has been made to isolate the improvements coming from using 3d-var. In principle 3d-var and OI should be equivalent if all the side constraints are the same but this is unlikely to be the case.

A coupled data assimilation system is being developed in Japan by Sugiura et al 2008. They use an atmospheric GCM at a resolution of T42 coupled to an ocean GCM at a resolution of 1deg. The previous work on coupled data assimilation used atmospheric models which had no fast timescales. The timescales were those of the ocean and the assimilation window, typically 6 months. The introduction of an atmospheric GCM introduces fast timescales of hours to days compared to a window length of 6 months. Some mechanisms must be used in order to limit the growth of these fast modes. One method is to use temporal smoothing and add damping to the tangent linear and adjoint models. Ten day averages of the model fields are compared with 10-day averages of the data in a 9-month assimilation window. The approach used by Sugiura et al 2008 is a form of weak constraint 4d-var, in that in addition to determining the initial conditions of the ocean, model parameters are also adjusted. In particular, multiplicative values (alpha) for the drag coefficient and the transfer coefficients for the fluxes of heat and evaporation are included. Atmospheric initial conditions are not part of the control variables in the cost function, though atmospheric fields are part of the penalty in the cost function. The values of the alphas are optimised every 10 days.

They compare their analyses with those obtained using a simpler system, which was also used for the prior for each 9-month 4d-var analysis, and claim that they are superior. A limited set of forecasts is made for the 1996/98 period when there was a large El Nino and Indian Ocean Dipole. The forecast skill does indeed look quite impressive but there are questions. It is quite probable that the coupled model will drift over the 18 months of a forecast, yet no correction for this seems to be made. They argue that the data assimilation creates initial states which are in better balance with the model state and so reduces drift. However the first 9 months of the forecast coincide with the assimilation window. Having felt future observations, the first months are not a pure forecast. There is clear model drift after 18 months of forecast; it is not clear what would prevent a drift after the 9-month assimilation window. The limited period of the forecasts makes it difficult to know if the coupled data assimilation leads to generally superior initial conditions for forecasting purposes or not.

An alternative strategy for coupled assimilation has been developed at GFDL. Here an ensemble square root filter approach has been adopted. An early application to seasonal forecasting is given by Zhang et al (2005). To remove fast timescales, the atmospheric model was a slave to the ocean, based on a statistical regression of wind stress heat and fresh water fluxes onto tropical Pacific SSTs. The deterministic part of the atmospheric model is based on the statistical relationship described above, but an additional stochastic component is included derived from the differences between the NCEP analyses and that part determined by statistical regression. Thus the atmosphere can be considered as consisting of two parts, a slowly varying deterministic part that depends on SST and a highly chaotic or stochastic part that evolves independently of SST. For the 12-month forecasts of Nino-3.4 SST with the hybrid coupled model, conducted for 12 January cases and 11 July cases, the ensemble ocean data assimilation improves upon their 3D-var assimilation.

In a later study Zhang et al (2007) used a coupled GCM including land and ice components. The emphasis in this paper is on the analysis rather than forecasts, however, and no hindcasts from the analyses are performed. So we do not know if the EAKF has a beneficial impact on seasonal forecast skill. However multivariate analysis is assessed and indeed maintaining the T-S relationship in the ocean analyses is found to be important. (Such a correction was included in GLOSEA3 and in EC_S3 several years ago, and found to be important in PEODAS.) Atmospheric data are also used (U, V, T but not q). The data are monthly mean values from the NCEP reanalysis, so removing the fast timescale.

In a short report on their web site Zhang et al (2008) do show results from a recent extensive set of hindcasts. Initial conditions for the hindcasts come either from their standard 3d-var system or from their EAKF coupled system. These results show a vast improvement of the EAKF coupled system over the 3d-var system. How much this relates to having a coupled analysis is unclear. The 3d-var does not have cross covariances such as preserving the T-S relationship or analysing salinity whereas the EAKF does. The forecast skill based on the 3d-var was poor. Thus improvements they note from the EAKF might not result from having a coupled analysis system at all and could possibly be obtained simply from improving their 3d-var system. However, the authors do say that they believe much of the improvement is from having a coupled analysis system. This may be based on extrapolating results from an earlier system where they compared the EAKF without cross covariances in order to be closer to the 3d-var.

Another approach to consistent initialization of the coupled system is that of NCEP's new Climate Forecast System Reanalysis (CFSR, Saha et al., 2010). Although the assimilation itself is not coupled in the sense that observations from one medium contribute to corrections in the other (no other system really does this either), the first guess for both atmosphere and ocean come from a coupled forecast. The CFSR will be used to initialize NCEP's next generation of seasonal forecasts, but results are not yet available. A similar approach, but with a pre-computed atmospheric analysis, MERRA, is being followed by the GMAO.

Studies of the ensemble prediction skill of CGCMs by Yang et al. (2006, 2007) suggest that the seasonal forecast skill can be improved by including the uncertainties related to the coupled instabilities in the initial ensemble perturbations, for example the ENSO-associated coupled bred vectors (BVs). BVs, so called because they are bred from small perturbations introduced into a coupled model, are designed to capture the dominant growing errors in the atmosphere-ocean coupled system. Studies have demonstrated that the coupled breeding technique can isolate the instabilities of interest, within the dynamically complex coupled system. The GMAO has used coupled breeding within an older version of its coupled model, CGCMv1, although in research mode, not in the operational configuration. The study by Yang et al. (2007) showed that both the forecast errors and the BVs in the subsurface ocean are dominated by large-scale structures near the ocean thermocline, especially during the strong 1997-1998 El Niño. Yang et al. (2006) showed that the structures were robust in that they were replicated in the completely different coupled model used by NCEP. Hindcast experiments starting from January 1997 with one pair of BVs achieve a significant improvement compared to the control (unperturbed) hindcast by capturing many important features of this event, including the westerly wind burst in early 1997. In a more extensive study, Yang et al. (2009) confirmed the positive impact from the coupled BVs. They found that the impact was particularly significant for forecasts initialized from the cold phase of the annual cycle in tropical Pacific SST, attributing this to the lower bias in the coupled model at that time. Yang et al. (2009, 2010) show that the use of BVs is also useful for the state-dependent multivariate covariances used on ocean assimilation. They show improvements particularly in the salinity state estimates and density stratification using the BV covariance information and that these improvements have a positive impact on forecast skill, albeit for a single case of the 2006 warm event.

For sub-seasonal timescales, it is likely that the appropriate rescaling norm in the BV calculation will be different from the ocean-dominated norm used for seasonal time scales. In uncoupled experiments Chikamoto et al. (2007) showed that growing instabilities related to the tropical intra-seasonal oscillation (MJO) can be identified with a physically-meaningful rescaling amplitude to isolate the growth in tropical regions with the Japanese Meteorology Agency ensemble prediction system. At the BoM, there are plans to extend their PEODAS system to calculate BVs of the coupled system to initialize monthly forecasts.

One important caveat in these early investigations of (weakly) coupled assimilation/initialization is that the evaluations have been focused on tropical Pacific SST, and usually either Niño-3 or Niño-3.4 indices. Of course having a good prediction of tropical Pacific SST goes hand-in-hand with a good forecast of the atmospheric tele-connections and continental temperatures and precipitation, but the impacts on these other fields have not been evaluated.

In summary, one might imagine that 4D variational assimilation of the coupled ocean atmosphere system would be a good way to create initial conditions from which to launch extended range and seasonal forecasts with a coupled model. A 4D var approach requires the adjoint of the coupled system. The adjoint of the atmospheric model exists at several centres for example at ECMWF or the UKMO and centres such as BoM and KMA which use the UKMO atmospheric model, and adjoints for ocean models either exist or are being developed. No adjoint for the ocean model NEMO yet exists though work is in progress to develop it. The adjoint of an earlier version of the ocean model was constructed by Weaver and used for assimilation of ocean observations in the Pacific Ocean with encouraging results (Weaver et al 2003). However, it is not clear how to deal with rapidly growing atmospheric perturbations in a coupled system with an assimilation window set by slow coupled or oceanic timescales.

Coupled assimilation can mean different things, depending on the application. Currently SSTs are fixed during the atmospheric analyses. A simple extension to this would be to include SST in the control variables and in the cost function of the 4d-var so allowing the SSTs to adjust a bit and hopefully produce a slightly more consistent relationship with surface fluxes. The degree to which this would improve short range NWP forecasts is unknown. Persisting the SST anomaly into medium range (10 day) forecasts might give some benefit. A slightly more complex system would be to have the SST more thermodynamically and dynamically linked to the ocean through the use of a mixed layer model with some ocean variables such as temperature in the mixed layer included in the cost function and in the control variables. Further complexity would involve the use of an OGCM, but the timescales are still those of NWP rather than seasonal prediction. This approach, however, has the advantage that it is unlikely to adversely upset the medium range forecasts, currently initialised without coupling. At the other extreme, Smith et al (2007) uses a form of anomaly coupling to initialise decadal predictions. In their scheme, ocean analysis anomalies are first calculated and then added to a balanced state of the climate model. The precise state of the atmosphere at the start of a decadal forecast is not thought to be important for a decadal forecast but it is critical for medium to extended range forecasts.

4.3 Ocean atmosphere surface interactions

No comprehensive review of the literature is given here since there have been two recent workshops on this topic. See for example the workshop proceedings on Ocean-Atmosphere Interactions ECMWF November 2008 and a follow-up workshop at the UKMO in Dec 2009. Some NWP centres have a (3rd generation) wave model coupled to the atmospheric model to the extent that the Charnock parameter depends on the wave state and this information is passed to the atmosphere. Differential velocity between the atmosphere and ocean can be incorporated. An extension would be to include the waves in determining energy and momentum transfer. NCAR calculate the Langmuir coefficient from the wave model, allowing more representative upper-ocean mixing. These developments suggest that the advantages of having a wave model integrated into the NWP and Climate models should be considered.

Further work is needed to better our understanding of mixing process in the upper ocean. Consideration should be given to: Comparison of 1-D Ocean Single column models (OSCMs) with large eddy simulations (LES); Coordination of an inter-comparison of above involving multiple LES and OSCMs; the impacts of vertical resolution; parametrizations to deal with lack of vertical resolution; the need to represent well the diurnal cycle and the role of the skin (details of processes in the upper metre); Langmuir circulations; inertially resonant motions and their interaction with the meso-scale circulation and re-stratification by sub meso-scale eddies.

4.4 Inclusion of greenhouse gasses

The observed evolution of temperature and other climate variables is thought to be comprised of an anthropogenic externally forced trend due to greenhouse gases and aerosols (GHGA forcing) superimposed on the natural variability of the system. Until recently, most seasonal forecasts did not explicitly include the effects of anthropogenic GHGA forcing but assumed that the effect is small compared to that of the natural variability and that the global warming signal is, in any case, largely incorporated into the forecast in the observation-based initial conditions. It can be argued, however, that the forecasts would then lack radiative support for the warmer temperatures over land that constitute the trends there and that this adds avoidable error to the forecasts. In a sensitivity experiment in which a coupled atmosphere/ocean model is used to produce the forecasts, Doblas-Reyes et al. (2006) and Liniger et al. (2007) investigated the differences between seasonal forecasts with and without GHG forcing. They clearly showed that the temperature forecast skill increased due to a better representation of the regional temperature trend patterns. The effect was not just constrained to the land areas but improvements in skill also appeared over the extra-tropical oceans. The effect could be appreciated as quickly as a few weeks into the forecast. Boer (2009) suggested that those systems without representation of the anthropogenic forcing could instead use an a posteriori correction to increase the skill. As a consequence, the most recent versions of the operational seasonal forecast systems (for example Wang et al., 2010; Luo et al., 2011) include realistic GHGA in the simulations.

4.5 The importance of mid-latitude SSTs

No extensive assessment of the importance of mid latitude SSTs is included here but the reader might find the way the UKMO used forecasts and ocean analyses including the mid-latitude in their seasonal forecast system of interest. This is described in Graham et al (2006) and Folland et al (2006).

5. A summary of some operational or planned extended-range forecast systems

As mentioned earlier, the monthly forecast range is a range that some NMSs are beginning to look at with more interest. For example the JMA issues operationally, monthly forecasts every week (Takahashi, ECMWF seminar Sept 2010). These are integrated using initial conditions from their reanalysis project. The hindcast set is 5 members every 10 days from 1979. The real-time forecast ensemble size is 51. The characteristics of their system in this respect are very similar to the ECMWF monthly system, though the resolution is less (T159L60), compared to (T319L62). The real-time forecast size of 51 in both systems is adequate to distinguish moderate weather signals in the tropics and extra-tropics. However, the ensemble size of the reforecasts viz. 5 is inadequate for the same task. So a useful assessment of hindcast skill spanning 18 (ECMWF) to 25 (JMA) years is not generally possible. The hindcast size is useful, however, for identifying the model climate and allowing the model drift to be removed from the real-time forecasts, though maybe not much more, but recall the results of Weigel et al (2008) referred to earlier.

The Japan Meteorological Agency (JMA) started experimental tercile probability forecasts for 3-month-averaged surface temperature and precipitation over the globe in April 2005. The Model Output Statistics (MOS) technique based on the 30 years of hindcasts is used to generate the probability forecasts. An ordered probit model is used as the statistical tool of the MOS. The tercile boundaries are determined so that the climatological chance of occurrence for each category is 33.3 % for the hindcast period from 1979 to 2008. The specification of the numerical prediction model is described at http://ds.data.jma.go.jp/tcc/tcc/products/model/outline/index.html. Comparison of the reliability for surface temperature over 30 years derived from forecasts with and without MOS, and with MOS but only for the areas with Relative Operating Characteristics (ROC) scores greater than 0.5 indicates that MOS correction works effectively, especially for the latter case. Only probability forecasts for the areas with ROC score greater than 0.5 are released.

The Met Office is planning to start testing a monthly forecast system in spring 2011. It will be based on their new seasonal forecasting system Glosea-4 (Arribas et al 2011), with high vertical resolution in both atmosphere and ocean, extending well through the stratosphere. The atmospheric horizontal resolution is N96; the ocean resolution is 1 degree with equatorial refinement. It will be run weekly with 28 members. The seasonal forecast has also recently been upgraded to include initialisation and evolution of both the sea-ice and the stratospheric state.

The ECMWF monthly forecast system has been described earlier. There are plans to extend the forecast range to 45 days and to increase the frequency of forecast generation from once per week to twice per week. On the seasonal timescale, System-4 will be introduced into operational use in 2011. Inter alia this will include a new ocean model (NEMO) with variational assimilation (NEMOVAR). The horizontal and vertical resolutions in both atmosphere and ocean will likely be increased. Since it is five years since the last update, substantial changes in the atmospheric model have been included, including the changes to convective parameterisation which greatly improved the model simulation of the MJO. The monthly (or 45-day) system will also use NEMO and NEMOVAR.

Starting in January 2011, National Centres for Environmental Prediction (NCEP) will also be initiating a monthly prediction system. In real-time the NCEP monthly prediction system will consist of 16 runs/day and each forecast will be for 45-days. Forecasts will be made using a coupled system (i.e., the Coupled Forecast System v2, CFSv2) with a T126L62 atmospheric model. Real-time forecasts will be accompanied with a set of hindcasts for 1999-2010 with 4 runs each day. All the components of the forecast system will be initialized from the recently completed Climate Forecast System Reanalysis (CFSR, Saha et al. 2010).

Canadian plans for the monthly forecast system are to base the monthly forecast system on the Canadian global EPS. This system is currently providing 16-day operational forecasts twice daily. It is planned to extend the range to 35 days, three times per month to obtain monthly forecasts. This approach will be transferred to operations in two stages. The first one (2011) will use persisted SST anomalies obtained from the CMC SST analyses. In the second stage (2013-2014?), the GEM model will be coupled to an ocean-ice system based on the NEMO ocean model. It has been demonstrated that phase one will significantly improve on the current operational monthly forecast system by about 10 days.

The short term strategy for the Canadian (EC) seasonal forecasting system is to use a two-model ensemble coupled climate prediction system that makes use of data assimilation to initialize the atmosphere, ocean, sea-ice and land surface components. This system has been extensively evaluated in the context of the Coupled System Historical Forecast Project (CHFP2) produces forecasts out to 12 month lead that are superior to the current 2-tier, 4-model operational system, and compares favourably to other coupled systems being used elsewhere. Discussions are currently underway to begin operational implementation.

In Australia, the Bureau is currently experimenting with a monthly forecast system. The atmosphere and ocean are the same as those used for seasonal forecasting but the ocean analysis has been upgraded to an EnKF. In addition bred vectors are calculated and will be used in the generation of the ensemble. An ensemble of 30, ten from each of three versions of the model is used, and an extensive set of hindcasts is also planned from 1982.

6. Summary

Recent results suggest that there is some potentially useful predictability at timescales intermediate between NWP and seasonal and it is worth exploring this further. On the medium range, TIGGE has been successful in establishing a data base from which methods of post-processing model forecasts to improve skill can be tested. A multi-model (MM) approach has been tested, and if only the better models in TIGGE are used, then the MM forecasts are preferable to those from any single un-calibrated model. Model Output Statistics (MOS) and other methods to correct for model error have also been tested. While these latter work well for the shorter range they are less appropriate as the forecast lead time increases. The spread of a single forecast model is invariably too small since a single model does not represent all the uncertainty in a forecast, leading to forecasts which are overconfident. However, if a reforecast set spanning many years is available then model drift can be removed and the spread calibrated, such that the skill of a single model can match that of the MM, at a given location. Indeed calibration can partially take into account downscaling in regions of complex terrain. There is a loss of temporal and spatial cohesion, however, but calibration could be useful for applications. In a seamless system, in principle the verisimilitude of models at one time range can be used to calibrate them at another. No definitive study has been performed but the idea could be exploited further in the extended range.

Whereas for medium range forecasting, model error is usually not dominant and a reforecast set for bias and skill evaluation is not performed, this is less acceptable the longer the forecast range. It is necessary for extreme weather forecasting, even in the medium-range, however, since the model PDF is not in general the same as nature making it necessary to reference any forecast to the model PDF. By monthly timescales, reforecasts are essential as model error not be ignored at this range.

Prediction of tropical cyclone frequency and landfall seems to follow the observed relationship with ENSO and the MJO giving some support to the idea that some aspects of extreme weather related to tropical cyclones might be predictable. Further studies show that other extreme events, particularly for temperature show similar levels of skill to forecasts of mean climate (Hamilton et al 2011).

For seasonal forecasting, as long a reforecast set with as many ensemble members as possible is preferable, typically 25 years with 10-15 ensemble members. This allows drift to be evaluated as a function of lead time and starting date and some evaluation of skill although there are issues with skill changing over time so that hindcasts may not always give good indication of forecast skill.

Extending the medium range to say 30 days, could be done without coupling to an ocean. Although atmosphere ocean interaction might influence the MJO to some extent, having an active ocean model as part of the forecast system is probably not essential. This is probably true for the ENSO/IOD variability also, another major source of predictability on the monthly timescale. If an ocean is included then good resolution near the top surface of the ocean is desirable. For forecast range beyond 30 days, however, it is important to have an ocean module. It is feasible, though not desirable, to use a two-tier system whereby the ocean SSTs are obtained from another forecast system, but a fully-coupled atmosphere-ocean model is the preferred option.

An important source of predictability on the extended range comes from representing the MJO well as well as getting the ENSO, IOD, NAM and SAM and their tele-connections correct. A poor representation of the MJO is likely to degrade forecasts not only in the tropics but also in mid-latitudes. No model has a completely acceptable representation of the MJO and its tele-connections though there has been considerable progress over the last few years at some centres and improvements in the propagation and characteristics of the MJO are leading to improved skill in the intra-seasonal range. Although predictability in the extra-tropics is modest there also appears to be an extra-tropical tele-connection from ENSO to the NAO which offers the possibility of improved (European) predictability during some winters (Ineson and Scaife, 2008). Parameterisation of convection seems to be important in representing the MJO, the ENSO/IOD and mean tropical state. Resolution by itself is not sufficient, however, even up to 10 km.

Tele-connections are a primary problem as is bias in the ENSO/IOD mode. The tele-connections stem from variations in tropical convection, which depends on getting both the mean state and variability right. For example, if the mean convection is too low in the eastern Indian Ocean, proper anomalies of reduced convection will not be made, and hence sufficiently strong tele-connections that are driven by reduced convection will not result. Representing monsoons well and predicting active and break periods is still a problem. The West African Monsoon (WAM) is a system where land convection and precipitation are also very important. Apart from the WAM precipitation being wrong in most of the models from the first week of forecast, there are some studies that point at the tele-connection with the summer climate over the Mediterranean region, in particular the eastern Mediterranean.

Parameterisation of physical processes is a very important component of model development. Errors in the representation of fast physics processes remain a key limiting factor in the skill of our models across all timescales from NWP to sub-seasonal to seasonal. Efforts to develop these representations are therefore of crucial importance for many applications. Increasingly, parameterization has a stochastic component to it. Several ways of including stochastic processes are being tested, generally with positive results. More complex forms such as cellular automata have the potential to allow nonlocal processes which might be important in representing for example the MJO.

Insufficient blocking is a major problem of many models and can affect forecasts on all timescales. The frequency of occurrence of blocking is somehow improved to a level commensurate with that estimated from ERA-40, through increased orographic drag, the revised convection scheme of Bechtold et al 2008 for Euro-Atlantic blocking, and by the convection scheme in the case of North Pacific. (Jung et al 2010). Improvements in blocking have also been associated with improvements in the mean state of a model used for seasonal forecasting at the UKMO. Improvements in both resolution and parametrization schemes can produce such improvements (Scaife et al, 2010, 2011, Tennant et al 2010, Matsuedo et al 2009).

Land conditions such as snow cover and soil moisture can give useful extended range predictability and therefore land should be initialised as well as possible. In the absence of an appropriate land reanalysis, nudging schemes such as ALI might work well and be cost effective.

Relaxation experiments can be used to identify remote origins of extended range prediction and give an idea of how much skill could be gained by reducing forecast error in these regions such as the tropics or the stratosphere. There is some evidence for significant benefits from better stratospheric representation. This will be further quantified by a CLIVAR-lead model study (SHFP) which is in progress to evaluate the importance of the stratosphere for extended range prediction. Further, operational models are now starting to resolve the stratosphere as part of their model development.

No operational model has a fully comprehensive upper ocean model interacting with the atmospheric model through the wave field. Some models do have a wave model built in and the physics of coupling is being made more comprehensive, but much more research work is needed to quantify the importance of improved upper ocean physics for intra-seasonal forecasting.

Presently the atmosphere and ocean components are analysed and initialised separately. Quite sophisticated schemes are generally used to analyse the atmospheric state such as 4d-var or EnKF. Ocean analysis techniques tend to be less sophisticated than their atmospheric counterparts but EnKF and EnOI techniques are being developed as are 3d-var techniques. Whereas a few years ago some schemes were univariate- analysing temperature but letting salinity wander, this is much less the case today and any useful scheme would be multi-variate. Coupled data assimilation is frequently raised as an objective. Weak coupling is already being developed, whereby the first guess is provided by the coupled model. Strongly coupled 4d-var assimilation in an operational setting still seems some way off but schemes which estimate cross covariances between atmosphere and ocean are being developed. Despite the fact that some data assimilation systems include a bias correction term in the atmospheric and the oceanic data assimilation systems (Balmaseda et al 2005), this bias is not usually used in the forecast. At the start of the forecast the bias term is switched off.

In the meeting agenda, three regional projects are mentioned, for South Asia, Africa and South America. Assuming that these projects have to deal with applications to for example hydrology, agriculture or disease prevention, most will use rainfall forecasts from EPS/monthly and/or seasonal prediction systems. It would be very useful if these regional projects could give a feedback to the modelling community about the "quality" of rainfall simulations over the respective regions. This may be done by establishing a small coordination group, who should ensure some consistency among the modelling input to these projects, as well as common verification tools and metrics. The goal of this exercise is to evaluate how different models/ensemble systems behave in terms of simulating convective rainfall over various tropical regions, both in terms of climatological properties (mean and variability) and in terms of forecast reliability (for example realistic spread-error relationship). The coordination group may also take care of interactions with for example the three CLIVAR monsoon panels. This project links model development with applications. The project could be broadened to tele-connections driven by tropical convection: this is the main source of seasonal predictability and clearly challenges most forecast systems. The importance of ENSO/IOD and its flavours is strong, yet every system still has strong ENSO/IOD biases.

Applications of extended-range and seasonal forecasts are still being developed. An example of malaria prediction is given, but there are other examples such as the Ganges river discharge (Webster and Hoyas 2004) and wheat production forecasts developed in Australia (Meinke and Stone 2005). Adapting the application model to run smoothly off model output at model scales is non trivial. Recalibration might be useful.

Acknowledgements

I would like to thank my colleagues Alberto Arribas, Andy Brown, Gilbert Brunet, David Burridge, Paco Doblas-Reyes, Harry Hendon, Arun Kumar, Michele Rienecker, Adam Scaife, Richard Swinbank and Frederic Vitart, for their helpful comments and input to improving the manuscript and Franco Molteni and Tim Stockdale for interesting discussions.

References

Alexander et al 2004: The atmospheric response to realistic Arctic sea ice anomalies in an AGCM during winter. J Clim, 17, 890-905.

Allen, M. R. & Stainforth, D. A., 2002: Towards objective probabilistic climate forecasting. Nature 419, 228.

Anderson, D., 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 ECMWF Tech Memo 503 www.ecmwf.int

Arribas Alberto, M. Glover, A. Maidens, K. Peterson, M. Gordon, C. MacLachlan, R. Graham, D. Fereday, J. Camp, A.A. Scaife, P. Xavier, P. McLean, A. Colman, S. Cusack 2011: The GloSea4 ensemble prediction system for seasonal forecasting. Mon Weath Rev doi: 10.1175/2010MWR3615.1

Baldwin, M.P. and T.J. Dunkerton, Stratospheric harbingers of anomalous weather regimes, Science, 244, 581-584, 2001.

Balmaseda M., Laura Ferranti, Franco Molteni and Tim N.Palmer 2010 Impact of 2007 and 2008 Arctic ice anomalies on the atmospheric circulation: Implications for long-range predictions, 136, 1655–1664.

Balmaseda M., D. Dee A Vidard and D. Anderson 2007: A multivariate treatment of bias for sequential data assimilation: application to the tropical oceans. QJ Roy Met Soc, 133, 167-179, part A, DOI:10.1002/qj.12.

Bechtold P. M.Koehler, T Jung, P Doblas-Reyes M. Leutbecher, M Rodwell and F Vitart, 2008. Advances in simulating atmospheric variability with the ECMWF model; from synoptic to decadal scales. Q.J.R.Meteorol.Soc., 134, 1337-1351.

Betts, A. K and P. Viterbo, 2005: Land-surface, boundary layer and cloud-field coupling over the south-western Amazon in ERA-40. J. Geophys. Res., 110, D14108, doi:10.1029/2004JD005702.

Boer, G.J. (2009). Climate trends in a seasonal forecasting system. Atmos-Ocean, 47, 123–138 doi:10.3137/AO1002.2009.

Boer, G.J. and K. Hamilton, 2008: QBO influence on extratropical predictive skill. Climate Dynamics, 31, 987-1000.

Bougeault P. and 21 others 2010: The THORPEX interactive Grand Global Ensemble. BAMS DOI:10.1175/2010BAMS2853.1

Boville BA. 1984. The influence of the polar night jet on the tropospheric circulation in a GCM. J. Atmos. Sci. 41: 1132–1142.

Brunet G. and 13 others 2010: Collaboration of the weather and climate communities to advance sub-seasonal to seasonal prediction BAMS 10.1175/2010BAMS3013.1

Buizza R., M. Leutbecher, L. Isaksen, J. Haseler 2010: Combined use of EDA- and SV-based perturbations in the EPS. ECMWF Newsletter No. 123, p22-28.

Cassou C. 2008: Intraseasonal interaction between the Madden–Julian Oscillation and the North Atlantic Oscillation. Nature 455, 523-527 | doi:10.1038/nature07286.

Chikamoto, Y., H. Mukougawa, T. Kubota, H. Sato, A. Ito, and S. Maeda, 2007: Evidence of growing bred vector associated with the tropical intraseasonal oscillation, Geophys. Res. Lett., 34, L04806, doi:10.1029/2006GL028450.

Chen Mingyue, Wanqiu Wang, and Arun Kumar 2010:Prediction of Monthly-Mean Temperature: The Roles of Atmospheric and Land Initial Conditions and Sea Surface Temperature. J Clim, 23, 717-726.

Coelho C.A.S., D. B. Stephenson, F. J. Doblas-Reyes, and M. Balmaseda 2006: The skill of empirical and combined/calibrated coupled multi-model South American seasonal predictions during ENSO. Advances in Geosciences, 6, 51–55, SRef-ID: 1680-7359/adgeo/2006-6-51

Davey M.K. et al 2006: Multi-model multi-method multi-decadal ocean analyses from the ENACT project. Clivar Exchanges No 38, Vol 11, no 3, July 2006.

Delsole T and J Shukla 2011 Model Fidelity versus skill in seasonal forecasting. Submitted.

Dirmeyer, Paul A., Zhichang Guo, Xiang Gao, 2004: Comparison, Validation, and Transferability of Eight Multiyear Global Soil Wetness Products. *J. Hydrometeor*, **5**, 1011–1033.

Doblas-Reyes, F.J., A. Weisheimer, M. Deque, N. Keenlyside, M. McVean, J.M. Murphy, P. Rogel, D. Smith and T.N. Palmer (2009). Addressing model uncertainty in seasonal and annual dynamical seasonal forecasts. Quarterly Journal of the Royal Meteorological Society, 135, 1538-1559, doi:10.1002/qj.464.

Doblas-Reyes, F.J., R. Hagedorn, T.N. Palmer and J.-J. Morcrette (2006). Impact of increasing greenhouse gas concentrations in seasonal ensemble forecasts. GeophyResLet, 33, L07708, doi:10.1029/2005GL025061.

Douville, H., 2004: Relevance of soil moisture for seasonal atmospheric predictions: is it an initial value problem? Climate Dyn., 22, 429-446.

Douville, H. and F. Chauvin, 2004: Relevance of soil moisture for seasonal climate predictions: a preliminary study. Climate Dyn., 16, 19-736.

Drusch M., and P. Viterbo, 2007: Assimilation of screen-level variables in ECMWF integrated forecast system: A study on the impact on the forecast quality and analyzed soil moisture. Mon. Wea. Rev., 135, 300-314.

Fennessy M and J Shukla (1999) Im pact of initial soil wetness on seasonal atmospheric prediction. J Clim, 12, 3167-80.

Ferranti, L., T. N. Palmer, F. Molteni, E. Klinker, 1990: Tropical-Extratropical Interaction Associated with the 30-60 Day Oscillation and Its Impact on Medium and Extended Range Prediction. Journal of the Atmospheric Sciences:Vol. 47, No. 18, pp. 2177-2199.

Ferranti, L., and P. Viterbo, 2006: The European summer of 2003: Sensitivity to soil water initial conditions. J. Climate, 19, 3659–3680.

Folland C.K. et al 2006: The 2005/6 winter in Europe and the United Kingdom: part 2, prediction techniques and their assessment against observations. Weather, 61, 337-345.

Göber, M., Zsótér, E. and Richardson, D. S. (2008), Could a perfect model ever satisfy a naïve forecaster? On grid box mean versus point verification. Meteorological Applications, 15: 359–365. doi: 10.1002/met.78

Gottschalck J. plus 13 others 2010: A Framework for assessing Operational MJO forecasts: A project of the Clivar MJO working group. BAMS 10.1175/2010BAMS2816.1

Graham R et al: The 2005/6 winter in Europe and the United Kingdom: part 1, how the Met Office forecast was produced and communicated. Weather, 61, 337-345.

Hagedorn R., P. Doblas-Reyes and T Palmer, 2006: A Real Application of seasonal forecasts – malaria early warnings. ECMWF Newsletter No 117 Spring 2006. See also 2 February 2006 *Nature*, vol. **439**, 576–579, doi: 10.1038/nature04503).

Hagedorn R. 2010 On the relative benefits of TIGGE multi-model forecasts and reforecasts and reforecast-calibrated EPS forecasts. ECMWF Newsletter, 124, 17-23.

Hagedorn R. 2008: Using the ECMWF reforecast dataset to calibrate EPS forecasts. ECMWF Newsletter, 117, 8-13.

Hagedorn, Renate, Thomas M. Hamill, Jeffrey S. Whitaker, 2008: Probabilistic Forecast Calibration Using ECMWF and GFS Ensemble Reforecasts. Part I: Two-Meter Temperatures. *Mon. Wea. Rev.*, **136**, 2608–2619. doi: 10.1175/2007MWR2410.1

Hamill, Thomas M., Renate Hagedorn, Jeffrey S. Whitaker, 2008: Probabilistic Forecast Calibration Using ECMWF and GFS Ensemble Reforecasts. Part II: Precipitation. *Mon. Wea. Rev.*, **136**, 2620–2632.
doi: 10.1175/2007MWR2411.1

Hamill T and J. S. Whitaker, 2007: Ensemble calibration of 500-hPa geopotential height and 850-hPa and 2-m temperatures using reforecasts. Mon. Wea. Rev., 135, 3273–3280.

Hamill T., J Whittaker and S Mullin 2006: Reforecasts, an important dataset for improving weather predictions. BAMS, Jan, 33-

Hamilton Rosie Eade, Richard Graham, Adam Scaife, Doug Smith and Anna Maidens 2011. Forecasting the frequency of extreme daily events on seasonal timescales, in preparation.

Higgins, R. W., J.-K. E. Schemm, W. Shi, and A. Leetmaa, 2000: Extreme precipitation events in the western United States related to tropical forcing. J. Climate, 13, 793--820.

Hou, D., Z. Toth, Y. Zhu, and W. Yang (2008), Impact of a stochastic perturbation scheme on NCEP global ensemble forecast system, in Proceedings of the 19th AMS

Conference on Probability and Statistics, 21-24 January 2008, New Orleans, Louisiana.

Houtekamer, P. L., M. Charron, H. L. Mitchell, and G. Pellerin (2007), Status of the global EPS at Environment Canada, in Proceedings of the ECMWF Workshop on Ensemble Prediction, 7-9 November 2007, pp. 57–68.

Hudson D and O Alves (2007) BMRC Res. Lett. No. 8, cawcr.gov.au/bmrc/pubs/researchletter/reslett_08.pdf

Hudson Debra, Oscar Alves, Harry H. Hendon and Andrew G. Marshall 2010a: Bridging the Gap between Weather and Seasonal Forecasting: Intraseasonal Forecasting for Australia. Accepted for publication in Quarterly Journal of the Royal Meteorological Society.

Hudson Debra, Oscar Alves, Harry H. Hendon, Guomin Wang 2010b: The impact of atmospheric initialisation on seasonal prediction of tropical Pacific SST. Clim Dyn DOI 10.1007/s00382-010-0763-9.

Ineson S. and Scaife A.A (2008). The role of the stratosphere in the European climate response to El Nino. Nature Geoscience, 2, 32-36.

Johnson C. and R. Swinbank 2009: Medium range multi-model ensembles combination and calibration. Q J Roy Met Soc, 135, 777-794. doi:10.1002/qj.383.

Jung T. 2010: Athena Project. ECMWF Newsletter No 124, 8-9.

Jung T., G. Balsamo, P. Bechtold, A. Beljaars, M. Kohler, M. Miller, J.-J. Morcrette, A. Orr, M. Rodwell and A. Tompkins 2010a: The ECMWF model climate: Recent progress through improved physical parametrizations. ECMWF Tech Memo 623.

Jung, T., M. J. Miller, T. N. Palmer, 2010b: Diagnosing the Origin of Extended-Range Forecast Errors. Mon. Wea. Rev., 138, 2434–2446. doi: 10.1175/2010MWR3255.1

Jung, T. and F. Vitart, 2005: Medium-range weather forecasting in the extratropics during wintertime with and without an interactive ocean. ECMWF TECH MEMO No 470.

Koster, R.D., et al 2004: Regions of strong coupling between soil moisture and precipitation. Science, 305, 1138-1140.

Koster and Coauthors, 2006: GLACE: The Global Land–Atmosphere Coupling Experiment. Part I: Overview. J. Hydrometeor.,7, 590–610.

Koster, R.D., S. Mahanama, T.J. Yamada, G. Balsamo, M. Boisserie, P. Dirmeyer, F. Doblas-Reyes, C.T. Gordon, Z. Guo, J.-H. Jeong, D. Lawrence, Z. Li, L. Luo, et al., 2010. The Contribution of Land Surface Initialization to Subseasonal Forecast Skill: First Results from the GLACE-2 Project. Geophy. Res. Letters, 37, L02402. DOI: 10.1029/2009GL04167.

Kumar Arun 2007: On the Interpretation and Utility of Skill Information for Seasonal Climate Predictions. Mon Weath Rev, 135, 1974-1984.

Kumar Arun 2009: Finite Samples and Uncertainty Estimates for Skill Measures for Seasonal Prediction. Mon Weath Rev, 137, 2622-2631.

Kumar Arun, Mingyue Chen, Wanqiu Wang 2010 An analysis of prediction skill of monthly mean climate variability. Clim Dyn DOI 10.1007/s00382-010-0901-4, Accepted.

Kumar A and F Yang 2003: Comparative influence of snow and SST variability on extratropical climate in northern winter. J Clim,16, 2248-2261.

Lawrence, D, and P. J. Webster, 2002: The boreal summer intraseasonal oscillation and the South Asian monsoon. J. Atmos. Sci., 59, 1593-1606.

Lin, H., and G. Brunet, 2009: The influence of the Madden-Julian Oscillation on Canadian wintertime surface air temperature. Mon. Wea. Rev., 137, 2250-2262.

Lin, H., G. Brunet, J. Fontecilla, 2010a: Impact of the Madden-Julian Oscillation on the intraseasonal forecast skill of the North Atlantic Oscillation. Geophys. Res. Lett., 37, L19803, doi:10.1029/2010GL044315.

Lin, H., G. Brunet, and R. Mo, 2010b: Impact of the Madden-Julian Oscillation on wintertime precipitation in Canada. Mon. Wea. Rev., 138, 3822-3839.

Lin, H., G. Brunet and J. Derome. 2009. An observed connection between the North Atlantic Oscillation and the Madden-Julian Oscillation. J. Climate, 22:364-380.

Liniger, M.A., H. Mathis, C. Appenzeller and F. J. Doblas-Reyes (2007). Realisticgreenhousegasforcingandseasonalforecasts. Geophys Res Let, 34, L04705, doi:10.1029/2006GL028335.

Luo, J.-J., S.K. Behera, Y. Masumoto, T. Yamagata (2011). Impact of global ocean surface warming on seasonal-to-interannual climate prediction. J.Climate, inpress.

Marshall A.G., D.Hudson, M.C. Wheeler, H.H. Hendon, O. Alves 2010: Assessing the Simulation and Prediction of Rainfall Associated with the MJO in the POAMA Seasonal Forecast System.

Marshall and Scaife (2010). Improved predictability of stratospheric sudden warming events in an atmospheric general circulation model with enhanced stratospheric resolution. *J. Geophys. Res.*, 115, D16114, doi:10.1029/2009JD012643.

Matsueda, M., R. Mizuta and S. Kusunoki (2009), Future change in wintertime atmospheric blocking simulated using a 20-km-mesh atmospheric global circulation model,J.Geophys.Res,114,D12114.

Matsueda, M., M. Kyouda, Z. Toth, T. Miyoshi, H. L. Tanaka, and T. Tsuyuki: On the predictability of a blocking occurred on 15th December 2005. Third THORPEX International Science Symposium. 14-18 September 2009, Monterey, USA. http://air.geo.tsukuba.ac.jp/~tanaka/papers/paper225.pdf

Meinke H. and R. Stone 2005:Seasonal and inter-annual climate forecasting: the new tool for increasing preparedness to climate variability and change in agricultural planning and operations. Clim Change, 70, 221-253.

Mo, K. C., and R. W. Higgins, 1998: Tropical convection and precipitation regimes in the western United States. J. Climate, 11, 2404--2423.

Moncrieff, M.W., 2010: The multiscale organization of moist convection and the intersection of weather and climate. *AGU Geophys. Monog,* **189**, *Why does Climate Vary?* Ed. D. Sun and F. Bryan, 3-26.

Moncrieff, M.W., and Co-authors: The Year of Tropical Convection (YOTC): The Scientific Basis. *Bull. Amer. Meteorol. Soc*., submitted

Moncrieff, M.W., and E. Klinker, 1997: Mesoscale cloud systems in the Tropical Western Pacific as a process in general circulation models. *Quart. J. Roy. Met. Soc*., **123,** 805-827.

Moncrieff, M.W., M. Shapiro, J. Slingo, and F. Molteni, 2007: Collaborative research at the intersection of weather and climate. *WMO Bulletin*, **56**, 204-211.

Morcrette J-J, G. Mozdzynski and M Leutbecher, 2008: A reduced radiation grid for the ECMWF IFS. Mon. Weath. Rev., 136, 4760-4772.

Nobre C and 10 others 2010: Addressing the complexity of the earth system. BAMS doi: 10.1175/2010BAMS3012.1

Norton, W.A., 2003: Sensitivity of northern hemisphere surface climate to simulation of the stratospheric polar vortex. Geophys. Res. Let., 30, 1627.

Palmer T.N. 2006: Predictability of Weather and Climate: from theory to practice. P1-29, in Predictability of weather and Climate, CUP, pp702.

Palmer T.N., F Doblas-Reyes, A Weisheimer and M Rodwell 2008: Toward seamless prediction. Calibration of climate change projections using seasonal forecasts. BAMS, 459-470. See also reply to Scaife et al 2009 in BAMS Oct 2009, p1551-4 DOI:10.1175/2009BAMS2916.1

Palmer T. and A. Weisheimer 2009: Diagnosing the causes of bias in climate models: why is it so hard? 1-13. ECMWF Seminar Proceedings 2009.

Palmer, T.N., R. Buizza, F. Doblas-Reyes, T. Jung, M. Leutbecher, G.J. Shutts, M. Steinheimer, A. Weisheimer 2009: Stochastic Parametrization and Model Uncertainty. ECMWF Tech Memo 598.

Park, Y.-Y., R. Buizza, and M. Leutbecher, 2008: TIGGE: preliminary results on comparing and combining ensembles. Q. J. R. Meteorol. Soc., 134, 2029-2050
Also published as ECMWF Technical Memorandum No. 548

Rashid, H.A., H.H. Hendon, M.C. Wheeler, and O. Alves, 2010: Prediction of the Madden-Julian Oscillation with the POAMA dynamical prediction system. *Climate Dyn.* DOI 10.1007/s00382-010-0754-x.

Rienecker, M.M., R. Kovach, C.L. Keppenne, and J. Marshak, 2010: NASA's ocean observations for climate analyses and prediction. *BAMS* (submitted).

Saha, S., and co-authors, 2010: The NCEP Climate Forecast System Reanalysis. *Bull. Am. Meteorol. Soc*., **91**, 1015–1057.

Saith, N. and Slingo, J. (2006) The role of the Madden-Julian Oscillation in the El Nino and Indian drought of 2002. International Journal Of Climatology, 26 (10). pp. 1361-1378. ISSN 0899-8418.

Saji, N. H., B. N. Goswami, P. N. Vinayachandran, and T. Yamagata. 1999. A dipole mode in the tropical Indian Ocean. Nature 401:360–363.

Scaife A.A., J.R. Knight, G.K. Vallis, C.K. Folland 2005. A stratospheric influence on the winter NAO and North Atlantic surface climate. Geophys. Res. Let., 32, L18715.

Scaife A. A. and J.R. Knight (2008), Ensemble simulations of the cold European winter of 2005/6. Quart. J. Roy. Met. Soc., 134, 1647-1659.

Scaife A.A., T. Woollings, J.R. Knight, G. Martin and T. Hinton (2010). Atmospheric Blocking and Mean Biases in Climate Models. J. Clim., 23, 6143-6152.

Scaife A.A., Copsey D., Gordon C., Harris C., Hinton T., Keeley S., O'Neill A., Roberts M. and Williams K. 2011: Improved Atlantic Blocking in a Climate Model. Geophys. Res. Lett., submitted.

Scaife A.A, C. Buontempo, M. Ringer, M Sanderson, C Gordon and J. Mitchell 2009: Towards Seamless Prediction: Calibration of climate change projections using seasonal forecasts. BAMS, 1549-155. DOI:10.1175/2009BAMS2753.1

Shapiro and others 2010 An Earth-system Prediction Initiative for the 21st Century. BAMS doi: 10.1175/2010BAMS2944.1

Shi Li, Harry H. Hendon, Oscar Alves, Matthew C. Wheeler, David Anderson and Guomin Wang 2010: On the Importance of Initializing the Stochastic Part of the Atmosphere for Forecasting the 1997/98 El Niño. Climate Dynamics, accepted.

Shongwe, M.E., C.A.T. Ferro, C.A.S. Coelho and G.J. van Oldenborgh 2007: Predictability of cold spring seasons in Europe. Mon. Wea. Rev., 135, 4185-4201, doi:10.1175/2007MWR2094.1.

Smith, D. M., S. Cusack, A. W. Colman, C. K. Folland, G. R. Harris and J. M. Murphy, 2007, Improved surface temperature prediction for the coming decade from a global climate model, Science, 317, 796-799.

Stephenson D., C Coelho, F. Doblas-Reyes, and M Balmaseda 2005. Forecast assimilation: a unified framework for the combination of multi-model weather and climate predictions. Tellus A, 57, 253-264.

Stockdale T.N., D. L. T. Anderson, J. O. S. Alves & M. A. Balmaseda 1998: Global seasonal rainfall forecasts using a coupled ocean–atmosphere model. Nature 392, 370-373 doi:10.1038/32861

Stockdale T., D Anderson, M Balmaseda, F Doblas-Reyes, L Ferranti, K. Mogensen, T. N. Palmer, F. Molteni and F. Vitart 2010: ECMWF Seasonal Forecast System 3 and its prediction of Sea Surface Temperature. Climate Dynamics, in press.

Sugiura N. Toshiyuki Awaji, Shuhei Masuda, Takashi Mochizuki, Takahiro Toyoda, Toru Miyama, Hiromichi Igarashi, and Yoichi Ishikawa 2008: Development of a 4-dimensional variational coupled data assimilation system for enhanced analysis and prediction of seasonal to interannual climate variations. J Geophys Res. 113, C10017, doi:10.1029/2008JC004741

Takaya Y, F. Vitart, G. Balsamo, M. Balmaseda, M. Leutbecher and F. Molteni 2010: Implementation of an ocean mixed layer model in IFS. ECMWF Tech Memo 622.

Tennant W.J., G Schutts, A Arribas and S.Thompson 2010 Using a stochastic Kinetic Energy Backscatter Scheme to Improve MOGREPS probabilistic forecast skill. Monthly Weather Review 2010 ; e-View doi: 10.1175/2010MWR3430.1

Thomson M.C., F. J. Doblas-Reyes, S. J. Mason, R. Hagedorn, S. J. Connor, T. Phindela, A. P. Morse and T. N. Palmer 2006: Malaria early warnings based on seasonal climate forecasts from multi-model ensembles. *Nature*, vol. **439**, 576–579, doi: 10.1038/nature04503).

Toth Z., M. Pena and A. Vintzileos 2007: Bridging the gap between weather and climate forecasting; research priorities for intraseasonal prediction. B. Am Met Soc 88, 1427-9.

Vecchi, G. A., and N. A. Bond, 2004: The Madden-Julian Oscillation (MJO) and northern high latitude wintertime surface air temperatures. Geophys. Res. Lett., 31, L04104, doi: 10.1029/2003GL018645.

Vitart, F., 2009: Impact of the Madden Julian Oscillation on tropical storms and risk of landfall in the ECMWF forecast system. Geophys. Res. Lett., 36, L158 02, doi:10.1029/2009GL039089.

Vitart, F, 2005: Monthly Forecast and the summer 2003 heat wave over Europe: a case study. *Atmos. Sci. Lett*., **6**, 112-117.

Vitart F. 2004. Monthly forecasting at ECMWF. Mon Weather Rev.132: 2761-2779.

Vitart F. and F Molteni 2010: Simulation of the Madden-Julian Oscillation and its teleconnections in the ECMWF forecast system. Q. J. R. Meteorol. Soc 136:842-855. DOI:10.1002/qj.623.

Vitart F. and F. Molteni 2009a: Dynamical extended-range prediction of early monsoon rainfall over India. MWR, 137, 1480-1492.

Vitart F. and F. Molteni 2009b: An experiment with a 46-day ensemble prediction system. ECMWF Newsletter No 121, p25-29.

Vitart F. and 9 others 2008: The new VarEPS-monthly forecasting system: a first setep towards seamless prediction. Q J Roy Meteor Soc, 134, 1789-1799 DOI:10.1002/qj.322.

Waliser D. and others 2009: MJO Simulation Diagnostics. J Clim., 22, 3006-3030.

Waliser, D.E., M.W. Moncrieff, 2008: Year of Tropical Convection (YOTC) Science Plan,WMO/TD-No. 1452, WCRP -130, WWRP/THORPEX- No 9, 26 pp.

Wang Bin and 27 others 2009: Advance and prospectus of seasonal prediction: assessment of the APCC/CliPAS 14-model ensemble retrospective seasonal prediction (1980–2004). Clim Dyn (2009) 33:93–117 DOI 10.1007/s00382-008-0460-0

Wang Wanqiu, Mingyue Chen, and Arun Kumar, 2010: An Assessment of the CFS Real-Time Seasonal Forecasts, 2010: Weather and Forecasting, 25, 950-969. DOI: 10.1175/2010WAF2222345.1

Weaver A, C Deltel, E Machu, S. Ricci, N. Daget 2005: A multivariate balance operator for variational data assimilation. QJRMS, 131, 3605-3625.

Weaver, A.T., J. Vialard, and D.L.T. Anderson, 2003: Three- and four-dimensional variational assimilation with a general circulation model of the tropical Pacific Ocean, Part 1: formulation, internal diagnostics and consistency checks. Monthly Weather Review, 131, 1360-1378.

Webster, P. J., A. M. Moore, J. P. Loschnigg, and R. R. Leben. 1999. Coupled ocean-atmosphere dynamics in the Indian Ocean during 1997-98. Nature 401:356–360.

Webster, P. J. and C. Hoyas. 2004. Forecasting monsoon rainfall and river discharge variability on 20-25 day time scales. Bulletin of the American Meteorological Society 85(11):1745–1765.

Weigel A.P., Baggenstos D., Liniger M.A., Vitart F. and C. Appenzeller. 2008. Probabilistic verification of monthly temperature forecasts. Mon. Wea. Rev. 136, 5162-5182.

Weisheimer A and 9 others 2009 ENSEMBLES: A new multi-model ensemble for seasonal-to-annual predictions- skill and progress beyond DEMETER in forecasting tropical Pacific SSTs. Geo Res Lett. 36, L21711, doi:10.1029/2009GL040896.

Weller et al 2010: Assessment of Intraseasonal to interannual climate prediction and predictability. Nat Res Council, National Academies Press http://www.nap.edu/catalog/12878.html.

Wheeler, M., and H. H. Hendon, S. Cleland, H. Meinke, and A. Donald, 2009: Impacts of the Madden-Julian oscillation on Australian Rainfall and circulation. J. Climate, 22, 1482--1498.

Wheeler M. and H Hendon 2004: An all season real-time multivariate MJO index: development of an index for monitoring and prediction. Mon Weather Rev., 132, 1917-1932, doi:10.1175/1520-0493(2004)132<1917:AARMMI>2.0.CO;2.

Whitaker J., Xue Wei and F Vitart 2006: Improving week-2 forecasts with multimodel reforecast ensembles. Mon Weather Rev 134, 2279-84.

Wilks D.S. and T Hamill 2007: Comparison of ensemble MOS methods using GFS reforecasts. Mon Weath Rev, 135, 2379-2390.

Woolnough, S. J., F. Vitart and M. A. Balmaseda, 2007: The role of the ocean in the Madden-Julian Oscillation: Implications for MJO prediction. Quart. J. Roy. Meteor. Soc., 133, 117-128

Yang, S.-C., M. Cai, E. Kalnay, M. Rienecker, G. Yuan, and Z. Toth, 2006: ENSO Bred Vectors in coupled ocean-atmosphere general circulation models. *J. Climate,* **19**, 1422–1436.

Yang, S.-C., E. Kalnay, M. Cai and M. Rienecker, 2007: Bred vectors and forecast errors in the NASA coupled general circulation model. *Mon. Wea. Rev.*, **136**, 1305-1326, DOI: 10.1175/2007MWR2118.1.

Yang S.-C., C Keppenne, M Rienecker and E Kalnay 2009: Application of coupled bred vectors to seasonal-to-interannual forecasting and data assimilation. J Clim 22, 2850-2870.

Yang, S.-C., M. Rienecker, and C. Keppenne, 2010: The impact of ocean data assimilation on seasonal-to-interannual forecasts: A case study of the 2006 El Niño Event. *J. Clim.* **23**, 4080–4095.

Yin Yonghong, Oscar Alves and Peter R. Oke 2010: An ensemble ocean data assimilation system for seasonal prediction. Monthly Weather Review Pre-publication e-View doi: 10.1175/2010MWR3419.1

Zhang S., A Rosati, M. Harrison, R Gudgel and W Stern 2008: GFDL's Coupled Ensemble Data Assimilation System, 1980-2006 coupled reanalysis and its impact on ENSO forecasts. WCRP extended abstract. Available from the GFDL website.

Zhang S., M. Harrison, A Rosati, and A. Wittenberg 2007: System design and evaluation of coupled ensemble data assimilation for global oceanic climate studies. Monthly Weather Rev., 135, 3541-3564.

Zhang S., M. Harrison, A. Wittenberg, A. Rosati 2005: J Anderson and V Balaji: Initialisation of an ENSO Forecast system using a parallelized Ensemble Filter. Mon Weather Rev., 133, 3176-3201.

Zsoter E 2006: Recent developments in extreme weather forecasting. ECMWF Newsletter, 107, 8-17.