Seasonal prediction at the Met Office

contributed by the monthly/seasonal prediction group

Bernd Becker, Anca Brookshaw, Andrew Colman, Mike Davey, Margaret Gordon, Richard Graham, Matt Huddleston, Sarah Ineson, Bruce Ingleby, Peter McLean

Met Office, Bracknell
RG12 2SZ United Kingdom
mike.davey@metoffice.com

Abstract

The Met Office has long had an interest in seasonal forecasting, in terms of both scientific research and practical applications. Currently we operate both statistical and dynamical systems, which will be described in this article. The dynamical coupled ocean-atmosphere forecast system has a fully interactive ocean component for which the initial upper ocean state is an essential ingredient. The ocean also has a fundamental but non-interactive role in the statistical and 2-tier dynamical systems that make use of the high degree of sea surface temperature anomaly persistence on monthly to seasonal timescales.

Current public seasonal forecasts can be found at http://www.metoffice.com/research/seasonal/ , along with information about the systems and their performance.

1. A brief history

1.1. Statistics

The present statistical schemes have their origins in research in the 1980s on the connection between sea surface temperature (SST) anomalies and tropical rainfall variations, in particular in the African Sahel and Brazilian Nordeste regions (Folland et al. 2001, Ward et al. 1993). That research established connections between seasonal rainfall and pre-season large-scale SST anomaly patterns in historical observational data, and led to the introduction of real-time forecasts for those regions (Sahel from 1986, Nordeste from 1987), using multiple linear regression and discriminant analysis to issue forecasts in terms of historically equiprobable quint categories (Very Dry, Dry, Average, Wet, Very Wet). The forecasts were primarily sent by fax to National Meteorological and Hydrological Services (NMHSs) in the relevant regions, and also to select research institutes and organisations such as the UN Food and Agriculture Organisation. The statistical schemes have gradually expanded, to include East Africa (Mutai et al. 1998), UK/Europe (Colman 1997, Colman and Davey 1999), the North Atlantic Oscillation index (Rodwell and Folland 2002), and global-mean annual-mean temperature (Folland and Colman 2001). Means of communication have also changed: email has replaced fax for direct dissemination, and the internet has enabled public access to the forecasts. Many of the statistical forecasts have appeared in the Experimental Long-Lead Forecast Bulletin (now online at http://www.iges.org/ellfb) since its inception in 1993.

1.2. GCMs

By the early 1990s computer power had advanced to the point that it became feasible to extend Numerical Weather Prediction (NWP) models to the monthly/seasonal range, using relatively coarse ‘climate resolution’ and building on the experience gained through the development of the same General Circulation

Models (GCMs) for climate prediction and research in the Hadley Centre component of the Met Office. The Met Office established an ensemble 2-tier seasonal prediction system based on the HadAM series of atmospheric GCMs. The SST required to drive the AGCM was prescribed using persistence of monthly-scale SST and sea-ice anomalies observed at the time of the forecast start, and the ensemble was created using analysed atmospheric conditions at staggered start times.

Development was further stimulated by the collaborative European PROVOST research project (PROVOST = PRediction Of Variability On Seasonal to interannual Timescales, 1996-1999), and the increasing demand for seasonal forecast information as the potential benefits of such forecasts became more apparent. The PROVOST project focussed on potential predictability, principally using AGCM ensembles driven by observed SSTs. The results demonstrated significant skill, estimates of value, and the benefits of a multi-model approach (see the special issue of Q.J.Roy.Met.Soc., 2000). The Met Office contribution (Graham et al. 2000) included a comparison of skills obtained using both observed and statistically-predicted SST fields, and demonstrated that the 2-tier forecast skill was often close to that obtained in the corresponding simulations that used observed SST.

The first Southern African Regional Climate Outlook Forum (SARCOF1) in 1997 was a major step forward in bringing together forecast providers and users, and set the template for several other regional consensus forecast fora that meet regularly in Africa, Asia and South America (Basher et al. 2001). The Met Office made a substantial contribution to the organisation of SARCOF1, and provided forecasts from the 2-tier AGCM system. The provision of African forecasts expanded to become a service available to all NMHSs from January 1998, with global precipitation and temperature seasonal forecast maps being made available via password-protected internet pages. That service (which still continues) includes the use of ‘skill masks’ these are display options that mask out regions where skill (using a ROC skill measure, based on 18 years of retrospective forecast data) falls below a prescribed threshold.

The turn of the calendar from 1999 to 2000 and the threat of the ‘millennium bug’ brought a request from UK government for long-range forecasts leading up to the end of the year. Calibrated tercile probability forecasts were produced using the 2-tier AGCM system: predictions slightly favouring warmer than normal conditions for the UK turned out to be good advice. That experience helped lead to the establishment of the first public seasonal forecast service, described in section 2.1.

The PROVOST project also included steps toward using coupled ocean-atmosphere GCMs (CGCMs) for seasonal prediction. By that time the Met Office had developed a CGCM that used the same HadAM2 AGCM coupled to a regional ocean GCM of the tropical Pacific, with a focus on predicting the El Niño cycle and its impacts (Ineson and Davey 1997), and for PROVOST we produced sets of trial retrospective forecasts with that system. Over the past 2 years a new globally-coupled CGCM has been developed, as described in section 2.2.

2. Current systems and products

2.1. AGCM

The 2-tier system currently uses the HadAM3 version of the AGCM, which is the atmospheric component of HadCM3, the coupled model used for climate research and greenhouse gas climate predictions (Pope et al. 2000). HadAM3 has a 2.5 latitude by 3.75 longitude horizontal grid, and 19 vertical levels. Each week a 9-ensemble forecast is run to 6 months range, with the ensemble created by using initial conditions from atmospheric NWP analyses staggered at 6-hour intervals. The statistical SST prediction algorithm makes use of the weekly NCEP OI_V1 SST and sea-ice analysis. (Note: a change to OI_V2 (Reynolds et al. 2002) will be made when OI_V1 is discontinued at the end of 2002.)

The output is used to update weekly the password-protected 'NMS' site on the Met Office website, which contains global and regional graphics showing probabilities of 3-month-average above/below normal precipitation and 850hPa temperature, and ensemble-mean anomaly patterns. The weekly 2-tier ensemble is also the basis for The Monthly Outlook, which is a commercial product with forecast information for week 2 and weeks 3 and 4 ahead, principally for the UK (see http://www.metoffice.com/monoutlook/index.html); and for the long-range content of commercial site-specific temperature and precipitation forecasts (see http://www.weatherXchange.com).

Each month, 3 successive weeks of the 2-tier forecasts are used to create a 27-ensemble forecast, with products that are freely available on our website. Again, global and regional maps of probabilities of above/below normal precipitation and temperature are provided: menus provide access to various regions, variables and forecast periods. (See http://www.metoffice.com/research/seasonal/monthly/index.html) To give users information about the skill of the forecast system in a readily-understandable form, maps of Heidke skill scores are also provided. These scores are based on 20 years of retrospective forecasts using the same 2-tier system, for the same start-month, forecast range and period as the current forecast. Among many possibilities, we selected the Heidke score because it is closely related to a simple fraction-correct measure. The Heidke score is an adjustment of 'fraction correct' to take into account the observed frequency of occurrence of the event being predicted, and is scaled to have a maximum value of 1 (perfect forecasts), and a value of 0 if skill is that of 'chance' (e.g. forecasts drawn from the historical record at random).

Our Heidke score S is defined as:

$$S = (H/T - F) / (1 - F) = (H - C) / (T - C)$$

where

T is the total number of times the event is predicted to occur,

H is the number of correct predictions (or "hits"), i.e. the number of times the predicted event is the same as the observed event

C is the average number of correct forecasts out of T cases that would be obtained by chance. This is T times F, where F is the observed frequency of the event.

For example, if in 17 years above normal temperature is observed 9 times in a particular season, predicted above normal 8 times, and correctly predicted above normal 6 times, then the score is

$$(6/8 - 9/17) / (1 - 9/17) = 0.47$$

Here 'predicted above normal' means that the forecast probability of above normal was larger than the forecast probability of below normal. 'Correctly predicted above normal' means that the forecast was 'predicted above normal' and the observation was above normal in that season.

In addition to the above services, the 2-tier system is also used to provide seasonal prediction information to UK government departments, such as the Dept. for International Development.

2.2. CGCM: the GloSea system

GloSea is the name of the Met Office CGCM-based system for global seasonal forecasting. The GloSea CGCM is derived from the climate version of the Unified Model, HadCM3. Details of the model physics and discussion of the performance of HadCM3 can be found in Gordon et al. (2000). The atmospheric component of the GloSea CGCM is the HadAM3 AGCM: effectively the same as that used in the 2-tier system. A number of oceanic and coupling enhancements have been made for seasonal forecasting purposes. These include increased vertical ocean resolution (from 20 in HadCM3 to 40 levels in GloSea), a variable spatial horizontal grid which gives increased meridional ocean resolution in the tropics, and the recently developed Hadley Centre coastal tiling scheme which enables specification of the land-sea mask at the ocean resolution. The ocean component of the GloSea CGCM has a zonal grid spacing of 1.25°: the meridional grid spacing is 0.3° near the equator increasing to 1.25° poleward of the mid-latitudes. Like HadCM3, the GloSea CGCM contains no flux corrections or relaxations to climatology.

Ocean initial conditions for forecasts are taken from analyses generated using the ocean GCM component of GloSea, with assimilation of sub-surface temperature observations. Surface fluxes of momentum, heat and water are taken from ECMWF atmospheric analyses, and ocean model temperatures in the top layer are constrained to be close to surface observations. In the ocean analysis phase, there is slow timescale relaxation of temperature and salinity to climatology (Levitus and Boyer 1994, Levitus et al. 1995) at all model levels. The assimilation method is based on the Met Office FOAM scheme (an optimal interpolation type scheme): the performance of FOAM for real-time short range operational ocean forecasting is described by Bell et al. (2000). Assimilation of thermal data into the ocean model near the equator can lead to a dynamically unbalanced state with unrealistic deep overturning circulations, and the equatorial bias correction scheme of Bell et al. (2002) has been implemented to overcome this problem. (See Huddleston et al. 2002.)

Atmospheric and land-surface initial conditions for the CGCM forecasts are taken from atmospheric analyses.

Ensemble techniques are required to explore the probability space of the predictions. To generate the forecast ensemble we follow the method developed at ECMWF for use with their SYSTEM 2 seasonal forecasts. Perturbations are applied to the forecast system which aim to represent the uncertainty in observations of wind stress and sea surface temperature. Wind stress perturbations (sampled from the anomaly differences between independent monthly wind climatologies) are applied to the momentum fluxes that force the ocean analyses in such a way as to generate several ocean analyses for each forecast start time. At the start of the forecasts, sea surface temperature perturbations, similarly sampled from independent SST analyses, are added to the top 40m (with a ramp to zero at 40m) of the temperature field of the ocean analysis.

The differences between the ocean analyses thus produced can be surprisingly large: in equatorial Pacific upper ocean temperature sections the differences can be a few degC, despite the assimilation of plentiful in situ observational data in each analysis. This appears to be partly because the assimilation is not tightly constrained to the observations, and partly because the wind stress perturbation samples can occasionally be quite large. Measures to reduce this spread are being investigated.

The GloSea system is part of a real-time multi-model framework, at present involving the Met Office and ECMWF. The CGCM has been set up to run on the computing facility at ECMWF, using an infrastructure similar to that of ECMWF SYSTEM2. Real-time forecasts comprise a 40-ensemble that operates in 'burst' mode from a single start time: each of five ocean analyses has eight SST perturbations added at the start time, to generate 40 different initial ocean states. The ocean analyses themselves are separately generated at the Met Office, then transferred to ECMWF for each forecast.

Along with the 40-ensemble forecast each month, a 5-ensemble set of retrospective forecasts is run with the same start time-of-year, extending back over 15 years. This set is used to estimate the systematic error in the SST predictions as a function of start time and lead time, and hence to calibrate the SST forecasts by subtracting that systematic drift.

Real-time GloSea forecasts started production in January 2002, and are undergoing trials before being made publically available. Fig. 1 shows forecast plumes for Niño3 SST anomalies, starting from the beginning of October 2002. The colour differences indicate groups of forecasts from the five different ocean analyses: some groups are distinctly different at month 1, with increasing overlap with increasing lead time. Forecast spread can vary substantially from case to case (see Fig. 2 in section 3), but is generally larger than desirable. This may be partly due to the substantial initial ocean analysis spread, and partly due to the natural sensitivity of the CGCM which tends to be overactive in the tropical Pacific.

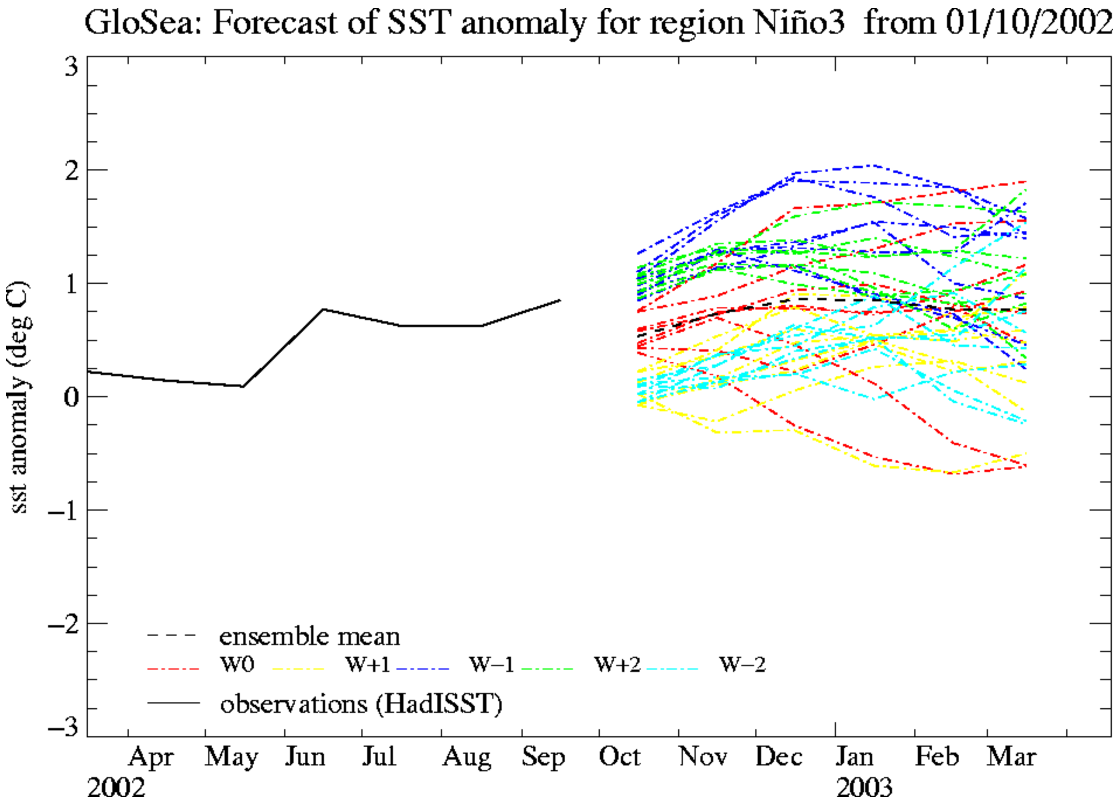

Figure 1: : forecast trajectories for Niño3 SST anomalies, from the GloSea system, for forecasts starting on 1 October 2002. The different colours indicate forecast groups with ocean initial conditions from different ocean analyses.

2.3. Statistics

The established set of regional statistical forecasts is being maintained. One development has been the increasing use of combined statistical and 2-tier AGCM forecast information: this combination is facilitated by adapting the statistical schemes to provide forecasts on the same grid as the AGCM. Where advantageous, the statistical and dynamical forecasts are weighted by a skill measure to produce an overall forecast. An extra simple statistical scheme, persistence of the previous year's observed conditions, is included in this procedure for regions such as the Sahel where decadal signals are substantial. Another development, in

response to end-user requests, is the inclusion of information such as the probability of 'extreme' anomalies (e.g. an anomaly larger than observed in the last 10 years). The most recent east Africa forecast is an example of these extensions.
(See e.g. http://www.metoffice.com/research/seasonal/regional/east_africa/index.html)

3. Some results from the DEMETER project

The Met Office is participating in the European DEMETER project (DEMETER = Development of a European Multi-model Ensemble system for seasonal to interannual climate prediction). Sets of 9-ensemble retrospective forecasts are being produced using both the 2-tier and GloSea systems, alongside those of other partners, and extensive assessments (ongoing) are available on the project website (http://www.ecmwf.int/research/demeter/).

For a coupled model, a key measure is the performance in the tropical Pacific with regard to prediction of SST variability. Fig 2 shows GloSea forecast mean, terce and range information for 3-month averages at 3-month lead (i.e. for 4-6 months ahead), for SST anomalies in the Niño3.4 region. (Note that the anomalies are relative to the model forecast climatology.) The forecasts capture the observed behaviour quite well (correlation 0.88 between observed and ensemble mean anomalies for the period shown), but with larger variance than observed, as reflected in the too-warm peak values in the major 1997/98 El Niño event.

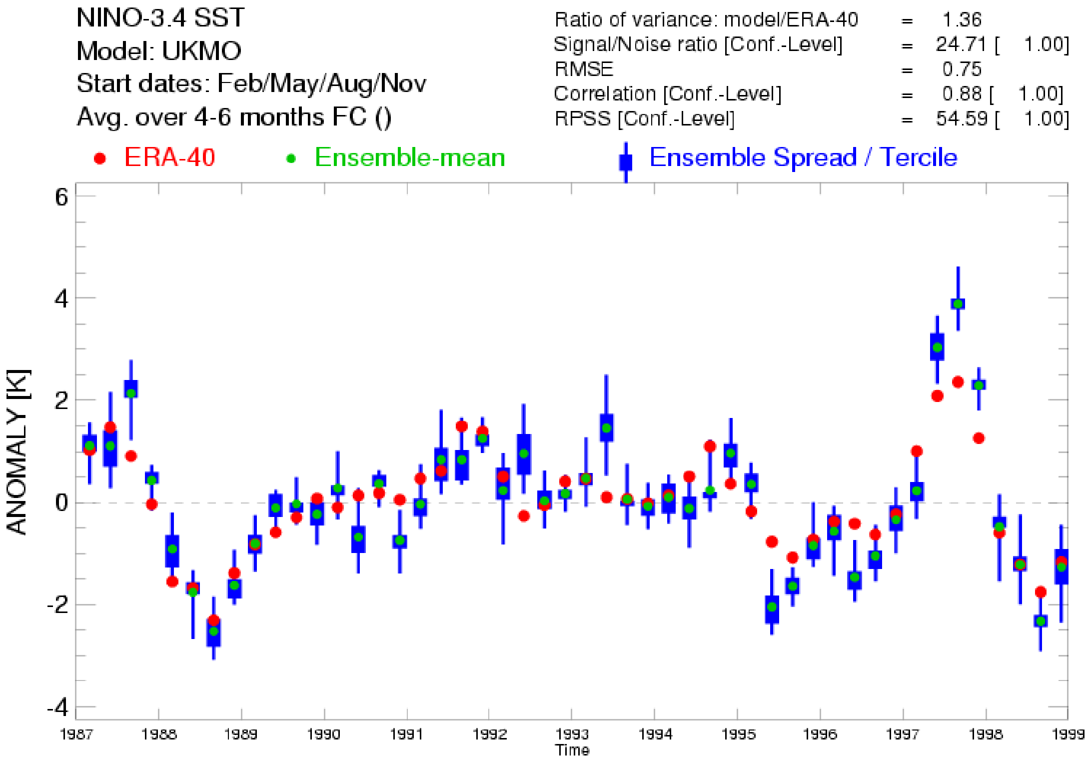

Figure 2 Retrospective forecast data forNiño3.4 SST anomalies, for 9-ensemble forecasts made using the GloSea CGCM system. (From the DEMETER project website.)

It is useful to compare results from the GloSea and 2-tier systems, because they have essentially the same atmospheric component. (Note that there are some minor AGCM configuration differences to accommodate coastal tiling in the CGCM, and the ensemble strategy is different in that the CGCM uses a 'burst' mode with one atmospheric analysis while the 2-tier uses 9 staggered atmospheric analyses.) Because persistence is generally a very good prediction strategy for SST over 1-2 months, one might expect the AGCM and CGCM performance to be very similar over that period. The fully interactive nature and more complete physics of the CGCM should make it the better forecast system at longer ranges, but this advantage might be offset by model errors in the more complex system and the tendency of the CGCM to drift towards its intrinsic climate as a forecast proceeds.

A comparison of ROC scores (cf Graham et al., 2000) from the two systems has been made, using probabilistic predictions of above-normal seasonal-mean 2m temperature and precipitation. Scores for six geographical regions were calculated over the 12-year period 1987-1998. Comparisons for 2m temperature (Fig. 3) show a degree of scatter about the diagonal in all seasons, indicating that for individual regions there are often substantial differences in scores.

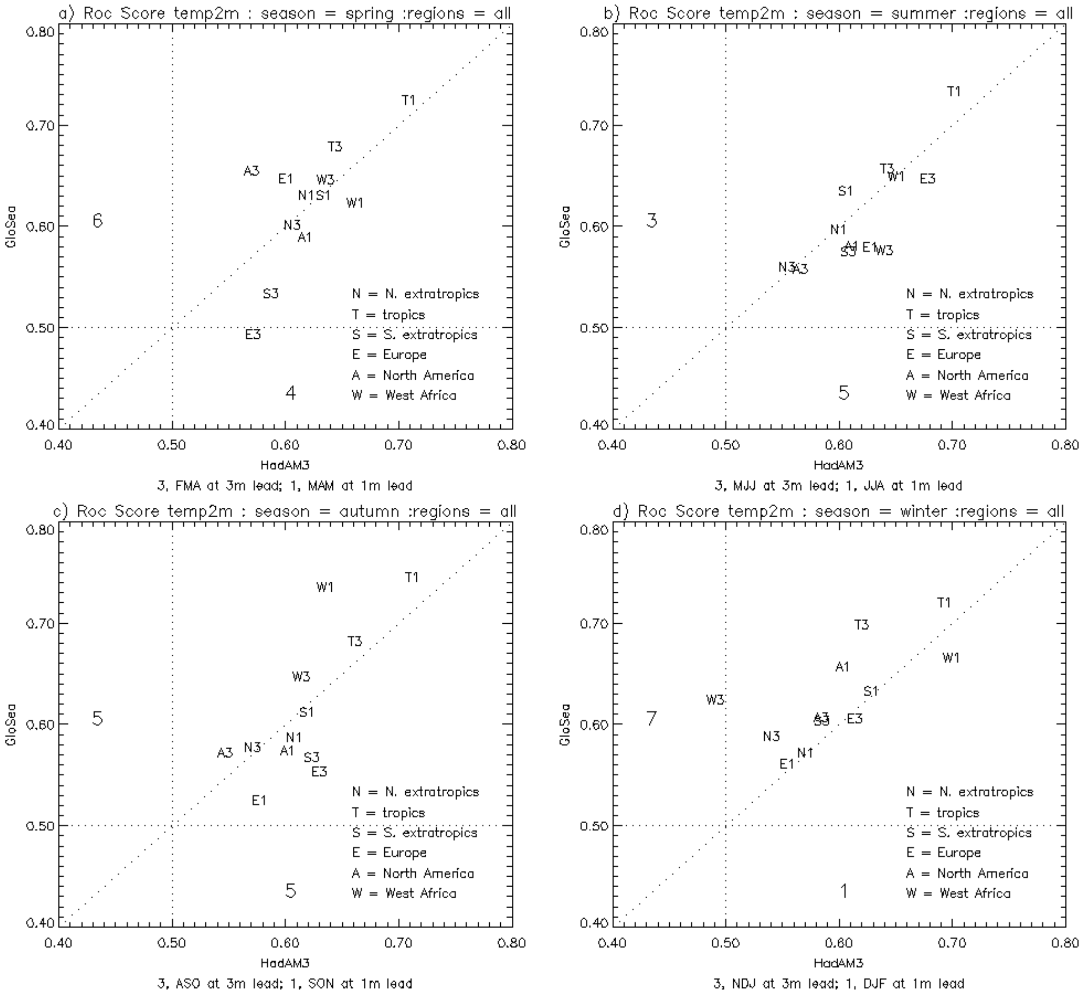

Figure 3: A comparison of ROC scores for the event '2m temperature above the climate normal' for various seasons, regions and lead times, using retrospective forecast data from 9-ensemble forecasts made using the GloSea CGCM system and the HadAM3 2-tier AGCM system as part of the DEMETER project. The scores are calculated for the12-year period 1987-1998. The numbers in the upper left (lower right) quadrant of each panel indicate how many times the GloSea (2-tier) system had the better score by a margin of at least 2%, with that score exceeding 0.5 .

The number of times each model gains a better score than the other is indicated next to the appropriate axis in each panel; in counting these wins for each model only cases in which the difference in the scores exceeds 2%, and the winning model has a score exceeding 0.5 are considered. As indicated by the number of wins, the overall performance of GloSea and HadAM3 is similar in northern spring, autumn and summer, though for the summer season HadAM3 appears generally more successful in this limited sample. Particular benefits from GloSea are evident in the long-lead predictions for the winter season, though the 1-month lead predictions are generally somewhat better with the HadAM3 model.

A similar analysis of precipitation scores (not shown) has results again often similar for both models, but with GloSea achieving the greater number of wins in each season. In particular the CGCM achieves better northern summer season forecasts for the West Africa region, and better long-lead predictions of winter precipitation over Europe.

4. Next steps

With the GloSea system recently established, the next focus will be on developing forecast products from that system. To aid calibration developments, the hindcast dataset will be expanded with the aim of matching the size of the real-time forecast 40-ensemble. There will be increasing emphasis on multi-model products,and work is needed to determine the most beneficial way of combining the different sources of forecast data.

It is anticipated that the Met Office will participate fully in the WMO infrastructure for long range forecasts (ILRF) that is in the planning stages. That infrastructure initially envisages a comprehensive set of seasonal products with a common structure being provided (via internet sites) by centres that have a global seasonal forecasting capacity. (See the report of the Nov 2001 ILRF meeting, at http://www.wmo.ch/web/www/DPS/reports/)

Acknowledgements:

The development and production of seasonal forecasts at the Met Office is supported by the Government Meteorological Research programme. The DEMETER project is supported by the European Commission under contract no. EVK2-CT-1999-00024.

References:

Basher, R., C. Clark, M. Dilley, M. Harrison (2001). Coping with the climate: a way forward. A multi-stakeholder review of Regional Climate Outlook Forums.
See http://www.cip.ogp.noaa.gov/docs/pretsum.pdf

Bell, M.J., R.M. Forbes and A. Hines (2000). Assessment of the FOAM global data assimilation system for real-time operational ocean forecasting. *J. Mar. Sys*., 24, 249-275

Bell, M.J., M.J. Martin and N.K. Nichols (2002). Assimilation of data into an ocean model with systematic errors near the equator. *Q. J. R. Meteorol. Soc.*, to appear.

Colman, A.W. and M.K. Davey (1999). Prediction of summer temperature, rainfall and pressure

in Europe from preceding winter north Atlantic ocean temperature. *Int. J. Clim*., 19, 513-536.

Colman, A.W. (1997). Prediction of summer central England temperature from preceding north Atlantic winter sea surface temperature. *Int.J.Clim* 17 1285-1300.

Folland, C.K and A.W. Colman (2001). Empirical prediction of the global temperature anomaly for 2002. *Experimental Long lead Forecast Bulletin*, 10, December 2001.

Folland, C.K., A.W. Colman, D.P. Rowell, and M.K. Davey (2001). Predictability of northeast Brazil rainfall and real-time forecast skill, 1987-98. *J.Climate* 14 1937-1958.

Gordon, C., C. Cooper, C.A. Senior, H. Banks, J.M. Gregory, T.C. Johns, J.F.B. Mitchell and R.A. Wood (2000). The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. *Clim Dyn*, 16, 147-168

Graham. R.J., A.D.L. Evans, K.R. Mylne,M.S.J. Harrison and K.B. Robertson (2000). An assessment of seasonal predictability using atmospheric general circulation models. *Q.J.R. Meteorol. Soc*, 126, 2211-2240.

Huddleston, M.R., M.J. Bell, M.J.Martin and N.K. Nichols (2002). Assessment of wind stress errors using bias corrected ocean data assimilation. *Submitted to Q. J. R. Meteorol. Soc.*

Ineson, S. and M.K. Davey (1997). Interannual climate simulation and predictability in a coupled TOGA GCM. *Mon. Weath. Rev.*, 125, 721-741.

Levitus, S. and T.P. Boyer (1994). World ocean atlas 1994, Vol 4: temperature. NOAA/NESDIS E/OC21, US Department of Commerce, Washington, DC, pp117

Levitus, S., R. Burgett, and T.P. Boyer (1995). World ocean atlas 1994, Vol 3: salinity. NOAA/NESDIS E/OC21, US Department of Commerce, Washington, DC, pp99

Mutai, C.C., M.N. Ward and A.W. Colman (1998). Prediction of East Africa seasonal "short rainfall" rooted in evidence for widespread SST-forced variability during October-December. *Int.J.Clim*. 18 975-997.

Pope, V.D., M.L. Gallani, P.R. Rowntree, and R.A. Stratton (2000). The impact of new physical parameterizations in the Hadley Centre climate model: HadAM3. *Climate Dynamics*, 16, 2-3, Pp123-146.

Reynolds, R.W., N.A. Rayner, T.M. Smith, D.C. Stokes and W. Wang (2002): An improved in situ and satellite SST analysis for climate. *J. Climate*, 15, 1609-1625.

Rodwell, M.J and C.K. Folland (2002). Atlantic air-sea interaction and seasonal predictability. *Quart. Journal Royal Meteor. Soc.*, 128, 1413-1443.

Ward, M.N., C.K.Folland, K.Maskell, A.W.Colman, D.P.Rowell and K.Lane (1993). Experimental seasonal forecasting of tropical rainfall at the Meteorological Office. In: *Prediction of Interannual Climate Variations,* ed. J.Shukla, NATO ASI series, Vol 16 pp 197-216