Assessing the skill of seasonal forecasts

Geert Jan van Oldenborgh

Introduction

Seasonal forecasts are predictions of the average weather up to one year ahead. Ordinary weather forecasts are limited to one or two weeks because of the chaos of the highly non-linear weather system. However, in some regions, in some seasons, the average weather is influenced in a predictable way by slowly varying boundary conditions, such as sea surface temperature (SST), ice or land conditions. These factors can have a local or a remote effect. Remote influences are called teleconnections.

The most important factor influencing the global weather on seasonal time scales is the El Niño – Southern Oscillation (ENSO) phenomenon in the eastern equatorial Pacific Ocean. There, SST can be up to five degrees above normal; this situation is called El Niño, the Christmas boy, as it often peaks at the end of the year. SST below normal is denoted by La Niña, the girl. El Niño and La Niña last for many months and can be predicted reasonably well up to half a year ahead. ENSO predictions are used to make seasonal forecasts of the weather in regions and seasons affected by ENSO teleconnections. The forecast is often computed as a shift in the mean weather depending on the strength and phase of ENSO. This means that simple, often linear, forecast models can be very useful in seasonal forecasting. In fact statistical models based on linear ENSO teleconnections are used in many locations throughout the world.

In contrast, dynamical seasonal forecast models, such as the system run operationally at the European Centre for Medium-Range Weather Forecasts (ECMWF), consist of ensembles of integrations of coupled General Circulation Models (GCMs) of the ocean and atmosphere, similar to the models used to make the weather forecast. If the models were perfect, this approach would yield better forecasts, both for ENSO itself and for the teleconnections that give rise to seasonal forecasts, as all physical processes thought to be relevant are included. In practice model errors limit the skill, and there has been some debate whether the theoretical advantages of GCMs translate in better forecasts than those produced by statistical models [1].

Methods

KNMI has made a systematic investigation [2] of the skill of the ECMWF seasonal forecast model [3], in collaboration with the seasonal forecasting group at the ECMWF. The skill was compared to that of statistical models over the period 1987-2001. The comparison considered both ENSO predictions and seasonal forecasts for temperature and precipitation. Shown here are results for the current version of the operational forecast model, named System-2 (S2).

Firstly, for ENSO forecasts, we compared the ECMWF model forecasts with those of a set of statistical models. Two of these are used operationally at the National Centers for Environmental Prediction (NCEP): the Markov model [4] and Constructed Analogue model [5] (CA). We also consider the ENSO-CLIPER model of Landsea and Knaff [1]. A baseline is given by the Cliper model: the optimal combination of climatology and persistence.

Secondly, ECMWF seasonal forecasts for temperature and precipitation were compared with those of a straightforward statistical model (STAT). This model uses persistence and lagged regressions of observations over 1901-1986 against two patterns of SST. The first pattern is ENSO, the second one the wider decadal ENSO-like pattern. Where the training period indicated a significant relationship between these patterns and observed temperature or precipitation some time later, the same relationship was used to make ‘forecasts’ for the verification period 1987-2001.

A problem in seasonal forecasting is the lack of data for a thorough verification. The ECMWF has made historical forecasts starting in 1987 for calibration, so only the 15 years 1987-2001 were available. Moreover, the skill of seasonal forecasts depends very strongly on the season and the region. This means one cannot combine forecasts from two different regions, or two different seasons, to increase the number of independent forecasts to compute the skill. Due to this lack of data we use the simplest possible measure of skill: the correlation of the ensemble mean. This is appropriate when biases in the mean state and variance are known and corrected for.

Even a measure as simple as the correlation coefficient has large uncertainties in samples of only 15 data points: if the true skill is zero, there still is a 5% chance that the correlation is 0.44: in those areas the system was just lucky. Conversely, in other areas it

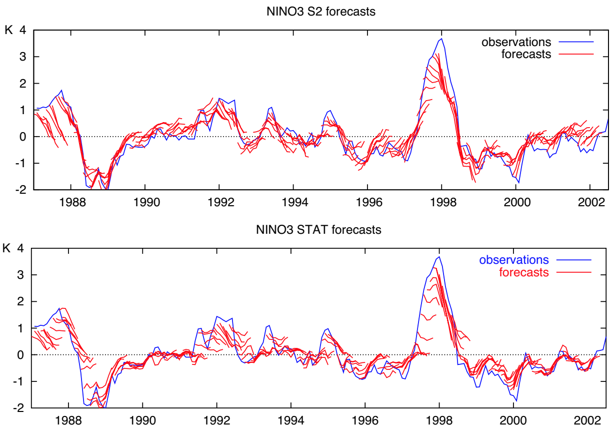

Figure 1. 6-month forecasts of the Niño3 index by the ECMWF *seasonal forecast model S2 (top) and a statistical model* STAT *(bottom).*

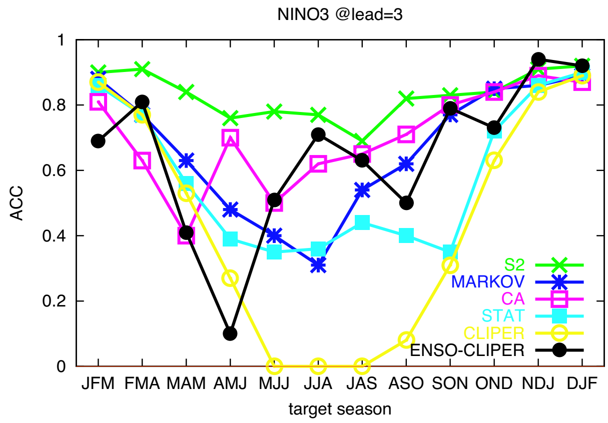

Figure 2. Skill of the Niño3 forecasts at lead time 3 months as function of the target season.

The skill of seasonal forecasts depends very strongly on the season and the region

may have been unlucky a few times. When considering large numbers of possible regions and seasons it is therefore inevitable that some areas of relatively high correlation occur by chance.

El Niño - Southern Oscillation forecasts

The forecasts of the current ECMWF seasonal forecast system (S2) and the statistical forecast system (STAT) of the Niño3 index (SST anomalies averaged over 5°S-5°N, 150°-90°W) over 1987-2001 are shown in Figure 1. The differences between the two are visible in the onset of El Niño (peaks) and La Niña (troughs). S2 correctly predicted the onset and decay of most El Niño and La Niña events. The exceptions are the failure to predict the 1987 El Niño and the tendency in 2001 to start the 2002 El Niño too early. The systematical underestimation of the strength of El Niño is due to a known model deficiency. In contrast, the statistical model failed to predict the onset of every El Niño and La Niña except 1999.

The same difference in behaviour can also be seen in the skill as a function of the target season, with an example for lead-time +3 months shown in Figure 2. (Lead time is defined here as the number of months between analysis time and the beginning of the forecast period.) The strong drop in prediction skill of the SST-based statistical models for forecasts through the boreal spring is known as the spring barrier. This barrier is reduced in the more sophisticated statistical models, and smallest in the dynamical ECMWF model. In making forecasts of SST for boreal winter, however, the statistical models do just as well as the coupled GCM.

Global temperature forecasts

We consider the 2-metre temperature T2m, and compare the seasonal forecasts of deviations from climatology (normal for the time of year) against the observed temperature, represented by the NCEP/NCAR reanalysis[6]. Over the oceans T2m is highly correlated with SST. The results from forecasts from November for December-February are shown in Figure 3 (left panels) for the S2 and STAT models. The statistical model has no forecasts (white) in areas where there is not enough data in the training period 1901-1986, or where persistence and teleconnections were not strong enough to base a forecast on. The colours indicate the skill of the forecasts over 1987-2001. Roughly, red denotes a good forecast, dark orange a useful one, light orange is doubtful and yellow useless.

One sees that temperature over the Indian, Pacific and equatorial Atlantic Oceans is forecast reasonably well by both models. However, over land the skill is lower than over the ocean. In northern South America and along the North American northwest coast one sees well-known ENSO teleconnections in winter as areas where both models have positive skill.

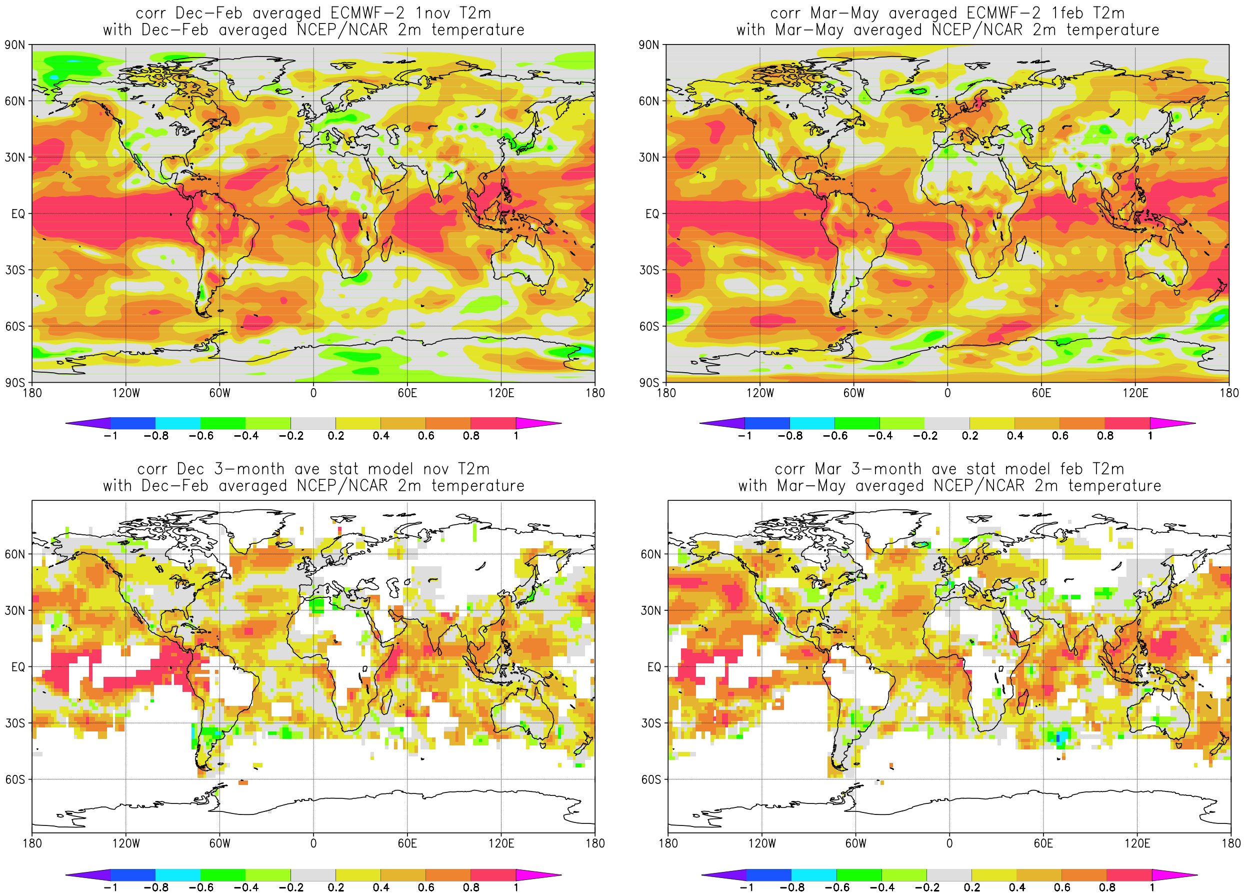

Figure 3. Skill of the T2m forecasts of the ECMWF model S2 (top) and the statistical model STAT (bottom) for December–February (left) and March–May (right) at lead +1 month. Roughly, red denotes a good forecast, dark orange a useful one, light orange is doubtful and yellow useless. White areas did not have enough data or not enough skill to make a forecast.

Comparing the skill plots, one sees that over the oceans the ECMWF model clearly outperforms the statistical model. Over land, the statistical model is hindered by lack of data in many tropical regions. In North America it is slightly better than the ECMWF model, which misrepresents the effects of El Niño along the US-Canada border. In Europe neither model shows any skill in predicting the temperature in winter.

In March-May (Figure 3, right panels) one sees comparable patterns over the oceans. Over land, melting snow and ice gives rise to predictability in North America and Europe in spring. The dynamical model uses this slightly better than the statistical persistence forecast.

Global precipitation forecasts

Precipitation forecasts of the S2 and STAT models are verified against the Global Precipitation Climatology Project[7)] analysis, which merges rain gauge observations with satellite estimates.

The strongest effects of ENSO on precipitation occur in eastern Indonesia and the western Pacific during the dry season (Augustus-November). The ECMWF model has skill r=0.8 in these areas in the last three months (September-November from August), see Figure 4 (left panels). These skills are higher than the skill of STAT. As the STAT model is based on ENSO teleconnections, this indicates that there are other factors than ENSO that give rise to predictability. A strong point of a GCM is that these factors are 'automatically' included and increase the skill of the forecasts.

In Australia one sees some skill along the east coast. However, the ECMWF model overextends the telecon-

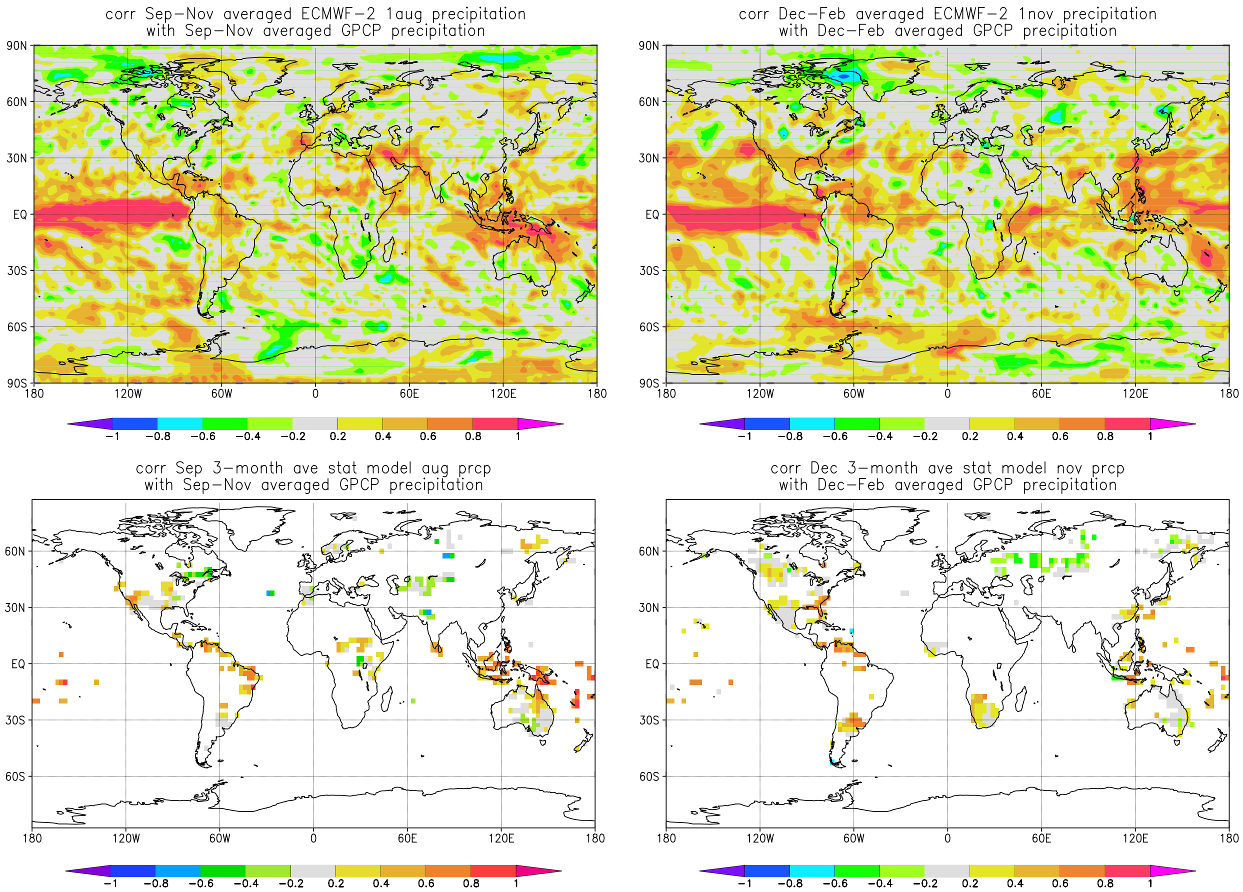

Figure 4. Skill of the precipitation forecasts of the ECMWF model S2 (top) and the statistical model STAT (bottom) for September–November (left) and December–February (right) at lead +1 month.

nection toward the west, leading to forecasts without skill in other parts of Australia. Model deficiencies can give rise to spurious forecasts, when the model itself indicates it can predict the weather when in fact it cannot. Precipitation in Central America and parts of the Caribbean was forecast fairly well by the ECMWF model, as was rainfall in parts of the Middle East and the Sahel. In Europe the model showed skill in the Iberian Peninsula. The statistical model performed much worse in these areas.

The extratropical ENSO teleconnections are strongest in the Americas in December-February. The skill of forecasts for this season is shown in the right panels of Figure 4. The ECMWF model reproduces the historical teleconnection patterns well in Florida, northern South America and southern Brazil. El Niño related rainfall along the coast of Ecuador and northern Peru was also forecast well by the ECMWF model, whereas the linear statistical model did not see a strong enough historical teleconnection to base forecasts on. On the west coast of the US the ECMWF model reproduced the long-term historical pattern of no ENSO effects. In contrast, the teleconnections observed over 1987-2002 were quite strong. It is not clear whether this change is predictable.

Outside the Americas there are also ENSO teleconnections in the season December-February. The ECMWF model used these to make reasonably successful forecasts in eastern China and India in this season. It also showed some skill in southern Africa and Morocco.

Overall view

Due to the limited number of years in the verification period, few of the differences between the ECMWF and statistical forecasts noted above are statistically

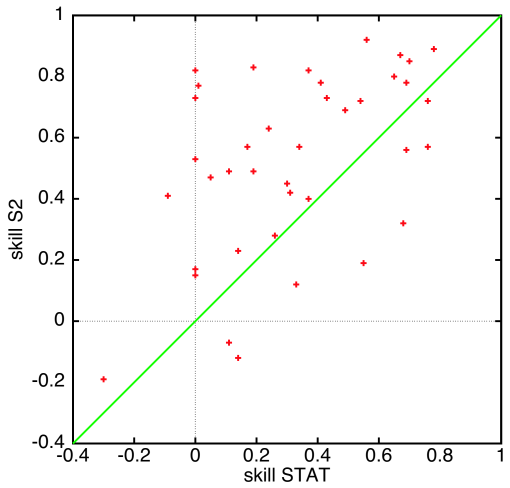

Figure 5. The correlation skill score of the +2 month precipitation forecasts of the ECMWF *seasonal forecast model S2 versus that of the statistical model* STAT *for 40 regions and seasons where skill is expected due to* ENSO *teleconnections.*

significant. The skill of precipitation forecasts for 40 regions and seasons where we expected skill due to ENSO teleconnections is shown together in Figure 5 for a lead-time of +2 months. In 32 out of the 40 cases the ECMWF model performed better than the statistical model. This result is very unlikely to be due to chance, implying that the ECMWF model performs better overall than this statistical model.

Conclusions

The performance of the ECMWF seasonal forecasts has been compared with that of some statistical ENSO forecast models and a simple but global statistical model. The ECMWF forecast model is shown to be better on average than these statistical models in

- predicting El Niño through the spring barrier
- predicting SST in the Indian and Pacific Oceans
- forecasting rain in areas of known ENSO teleconnections.

Many other aspects of performance have not been investigated here, particularly in mid-latitudes. Nonetheless, our results show that there are already a significant number of seasonal forecast 'targets' where the ECMWF numerical model is outperforming straightforward statistical techniques.

All maps have been produced at the KNMI Climate Explorer web site (*http://climexp.knmi.nl*). Readers are invited to investigate the skill in other areas and seasons.

1) Landsea, C.W. and J.A. Knaff, 2000. *How much skill was there in forecasting the very strong 1997-98 El Niño?* Bull. Amer. Meteor. Soc., **81**, 2107-2119.
2) Oldenborgh, G.J. van, M.A. Balmaseda, L. Ferranti, T.N. Stockdale and D.L.T Anderson, 2003. *Did the ECMWF seasonal forecast model outperform a statistical model over the last 15 years?* Technical Memoranda 418, ECMWF, Reading, UK, 32 pp, to appear in J. Climate.
3) Anderson, D.L.T., T. Stockdale, M.A. Balmaseda, L. Ferranti, F. Vitart, P. Doblas-Reyes, R. Hagedorn,

T. Jung, A. Vidard, A. Troccoli and T. Palmer, 2003. *Comparison of the ECMWF seasonal forecast systems 1 and 2, including the relative performance for the 1997/8 El Niño*. Technical Memoranda 404, ECMWF, Reading, UK, 93 pp.

4) Xue, Y., A. Leetmaa and M. Ji, 2000. *ENSO Prediction with Markov Models: The Impact of Sea Level*. J. Climate, **13**, 849-871.

5) Dool, H.M. van den, 1994. *Searching for analogues, how long must one wait?* Tellus, **46A**, 314-324.

6) Kalnay, E. and 20 co-authors, 1996. *The NCEP/NCAR 40-year reanalysis project*. Bull. Amer. Meteor. Soc., **77**, 437-471.

7) Huffman, G. J., R.F. Adler, B. Rudolf, U. Schneider and P.R. Keehn, 1995. *A Technique for Combining Satellite Data, Raingauge Analysis and Model Precipitation Information into Global Precipitation Estimate*. J. Climate, **8**, 1284-1295.