# Probabilistic seasonal forecast verification with the Climate Explorer

*G. J. van Oldenborgh*[*] *and C.A.S. Coelho*[‡]

KNMI

**Seasonal climate forecasts are made using multi-model ensembles. Contrary to climate change projections, the skill of the forecasts can be verified against observations using old forecasts and hindcasts. In practice the small number of forecasts (18-45 yrs) is a severe limitation, as the skill depends strongly on the region and season.**
**We present a web-based system to produce charts and maps of the skill of operational and research seasonal forecast systems using a variety of measures. It is part of the KNMI Climate Explorer, and presently contains data from the ECMWF S2, NCEP CFS and IRI ECHAM4.5 operational forecast systems, as well as the Demeter research experiment. The verification measures have been developed in the RCLIM project, and include deterministic measures such as the ensemble mean correlation, RMSE and MAE, as well as probabilistic measures such as the Brier Score, its decomposition into resolution, reliability and uncertainty, and the ROC curve. These are available both for time series (area-averaged or all grid points in a region) and as spatial maps.**

## 1. Verification of European spring temperature

As an example of the way the verification system can be used we consider the forecasts for spring temperatures in Europe. All figures in this analysis were copied from the web pages.

### 1.1 Deterministic score maps

From statistical evidence it is known that there is some skill in forecasting spring temperatures near sea and around the climatological snow line based on persisted winter temperature. The physical mechanism for this is persistence of anomalous SST and snow (Fig. A).

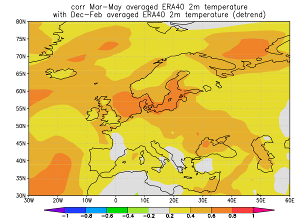


***Figure A.*** *Correlation of spring (MAM) 2m temperatures with the preceding winter (DJF) in ERA-40 (1958-2001); the effects of SST and snow persistence are clearly visible.*

The dynamical models starting Feb 1 (available Feb 15) also show skill in the correlation of the ensemble mean in this area and season (Fig. B). Note that the differing periods mean one cannot directly compare the skill scores of the models, although the Demeter version of S2 shows very similar skill over 1958-2002.

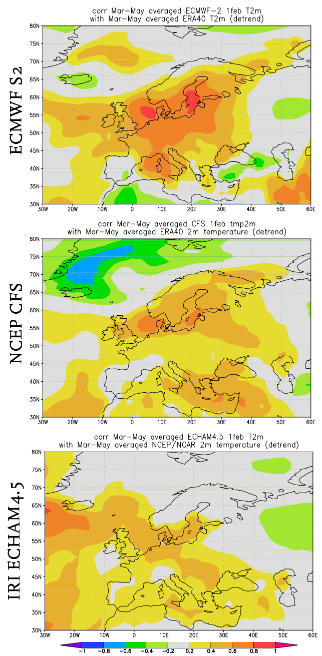


***Figure B.*** *Correlation of modelled and re-analysed spring temperature in ECMWF S2 (1987-2005), NCEP CFS (1981-2005) and IRI ECHAM4.5 (1968-2005).*

### 1.2 Probabilistic score maps

The skill of seasonal forecasts usually only allows probabilistic forecasts. A useful measure to verify these is the ROC score, which measures whether it is possible to increase the Hit Rate over the False Alarm Rate for a given threshold. It is 0.5 for no skill and 1 for perfect skill.

Before computing the skill score, the data is detrended and biases of the forecast model are removed in the mean and variance (using a jack-knife procedure). The climatology is by default computed over all available observations, not just the forecast period. Skill scores are undefined when the threshold is outside the observed range during the forecast period.

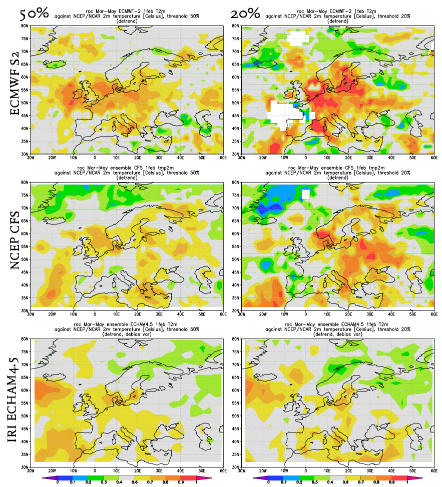


***Figure C.*** *ROC scores of median (left) and 20% coldest (right) spring temperatures in ECMWF S2 (5–40 members, 1987-2005), CFS (15–30 members, 1981-2005) and IRI ECHAM4.5 (12–24 members, 1968-2005).*

In Fig. C one sees that the ROC score for the median (left) is less robust (more noisy) than the correlation coefficient. The ROC scores for forecasts for the 20% coldest winters are much better than for the median in the S2 and CFS models. We speculate that the absence of snow does not lead to a useful forecast, but the presence of snow gives a significantly higher probability of a cold spring.

### 1.3 Probabilistic diagrams

It is often useful to verify that decadal variability and the changing number of ensemble members do not inflate the skill scores. We look at the forecasts for Warszawa, interpolated to 52°N, 21°E. These also emphasize the need for bias corrections.

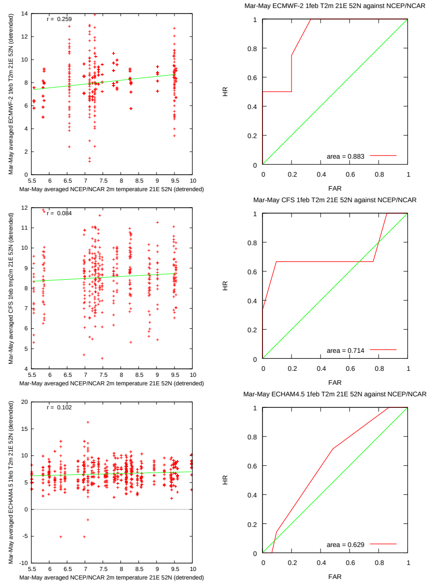


***Figure D.*** *Scatterplots (left) and 20% threshold ROC curves (right) of spring temperatures in S2, CFS and ECHAM4.5.*

The skill indeed comes from a few ensemble members with very cold weather in years with cold springs.

## 2. Conclusion

**The Climate Explorer web verification system is the first web-based verification system that allow anyone to quickly and conveniently**

- **compute skill scores for various forecast systems, as maps or time series,**
- **compare the skill of different systems,**
- **investigate apparent skill in certain areas.**

**We plan to add data of more forecast systems (ENSEMBLES data when it becomes available), implement more sophisticated bias correction schemes, add confidence intervals to the skill scores and make the system more user-friendly.**

***References***
*G.J. van Oldenborgh et al.* Evaluation of atmospheric fields from the ECMWF seasonal forecasts over a 15 year period. *J. Climate, 2005,* **18**, *2970-2989, corr. 5188-5198.*

*KNMI, De Bilt ‡*University of Reading*

Koninklijk Nederlands Meteorologisch Instituut