Project no. GOCE-CT-2003-505539

Project acronym: ENSEMBLES

Project title: ENSEMBLE-based Predictions of Climate Changes and their Impacts

Instrument: Integrated Project

Thematic Priority: Global Change and Ecosystems

Milestone M5.2: Prototype of an automatic system for forecast quality assessment of seasonal-to-decadal hindcasts

Due date of deliverable: 1 March 2006
Actual submission date: 8 March 2006

Start date of project: 1 September 2004 Duration: 60 Months

Geert Jan van Oldenborgh, KNMI

Revision 1

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)		
Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the Consortium (including the Commission Services)	

Milestone M5.2: Prototype of an automatic system for forecast quality assessment of seasonal-to-decadal hindcasts

Introduction

Even more than in weather forecasts, the skill of seasonal forecasts depends very strongly on the geographical location and season of the year. Also, different forecast systems have strong and weak points in various regions and seasons. For a user of seasonal forecasts it is therefore very important to know what skill the different systems have achieved in past forecasts and hindcasts in order to be able to judge the quality of current forecasts for his/her region.

Verification of seasonal forecasts is available from the various centres (e.g., ECMWF, NCEP, IRI, UKMO) and research experiments (e.g., Demeter). However, these are static pictures, not complete (for all months and lead times) and not directly comparable due to the use of different verification measures and colour scales.

We introduce here the first web site that allows for the dynamic generation of skill score maps and diagrams from a variety of seasonal forecast models using different skill scores.

Description

Withing the ENSEMBLES project we have constructed a web site that lets the user generate verification plots in real time. The site is part of the KNMI Climate Explorer (climexp.knmi.nl). It consists of three parts

1. A full set of monthly data of the hindcasts of the ECMWF S2, NCEP CFS and IRI ECHAM4.5 forecasts systems, plus the Demeter research experiment.
2. A large set of deterministic and probabilistic verification routines: correlation, RMS, MAE, ROC, RPS, RPSS (with respect to climatology), Brier Score and decomposition, Brier Skill Score (with respect to climatology).
3. A set of web pages that allows anybody to generate verification curves and maps from 1. and 2.

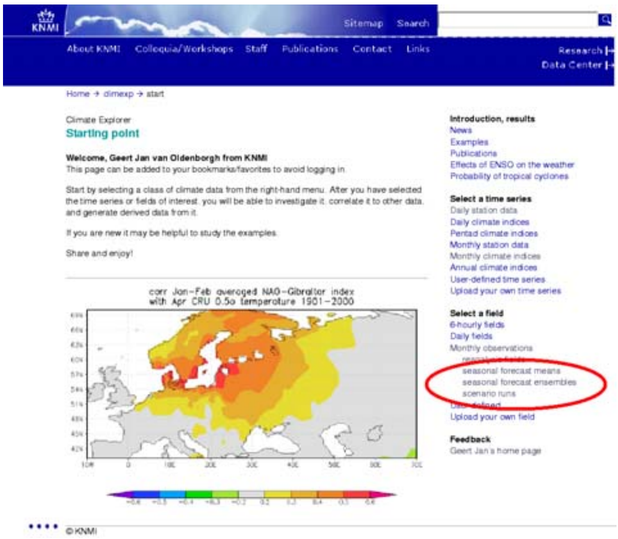
In the following examples are shown of the current set-up. Planned extensions are

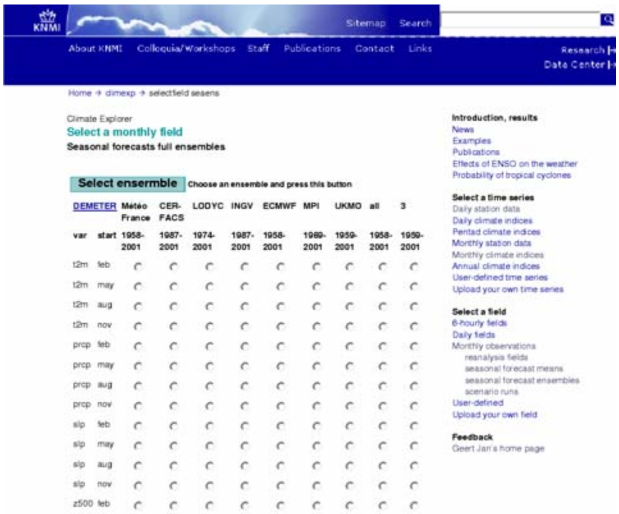
1. More data from other operational centres and the ENSEMBLES data servers
2. A connection to the public ECMWF seasonal to decadal ENSEMBLES server to allow the user to select data there and seamlessly verify at KNMI. This will allow for the first time verification studies of statistics of daily data, for instance the numbers of windstorms, high precipitation events, extreme snowfall.
3. More verification measurees, in particular skill scores such as the RPSS and BSS with respect to more sophisticated models than climatology and persistence: damped persistance, optimal normal correlactions and regression to Nino3.4.
4. A more user-friendly user-interface consisting of a single web page rather than the current multi-step procedure.

We envisage these extensions to be complete by summer 2006.

Example

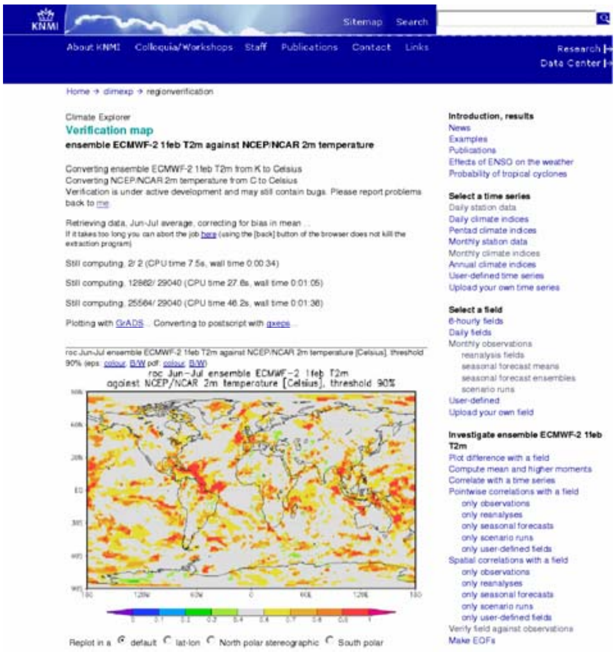
On the web site climexp.knmi.nl, the data sets available for verification are under `seasonal forecast ensembles':





One selects a field from the collection of Demeter, ECMWF, NCEP and IRI data, in this case the 1Feb T2m forecasts from the ECMWF System-2 operational seasonal forecast system was selected (only visible when one scrolls down the page). The next page gives a list of options to investigate this field, among them is `verify against observation'. This brings up the main verification form, with choices for the dataset to verify against (only relevant ones are shown), the verification measure, the threshold, season and plot options:

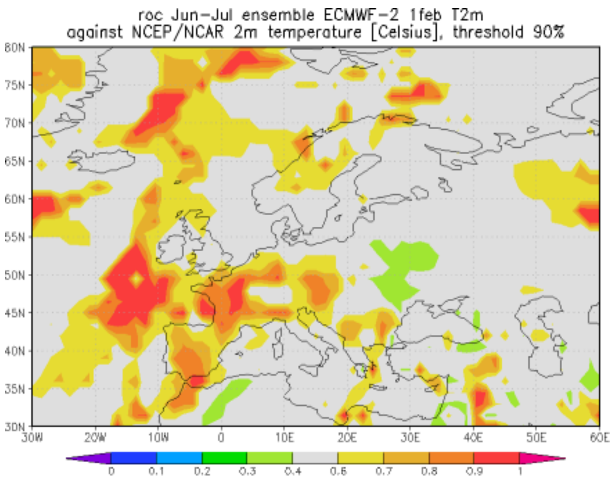
I chose the area under the ROC curve for very warm (90%) June-July, starting from Feb.1 analyses (this is not visible on the screendump, one has to scroll down to see all options). The verifying dataset is the NCEP/NCAR reanalysis, the units are automatically converted to agree with each other. The production of the verification map take less than two minutes, a typical time; most of this time is needed to read the forecast data from the data store:

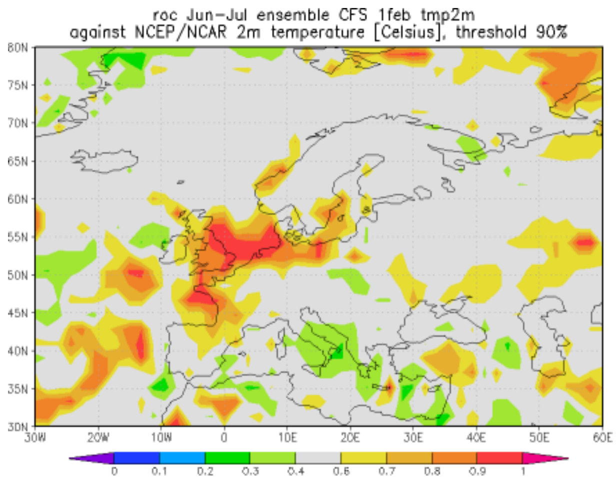


The resulting map is available in a variety of formats (PNG, PDF, EPS), and also the underlying data can be downloaded if the user prefers her own graphics software.

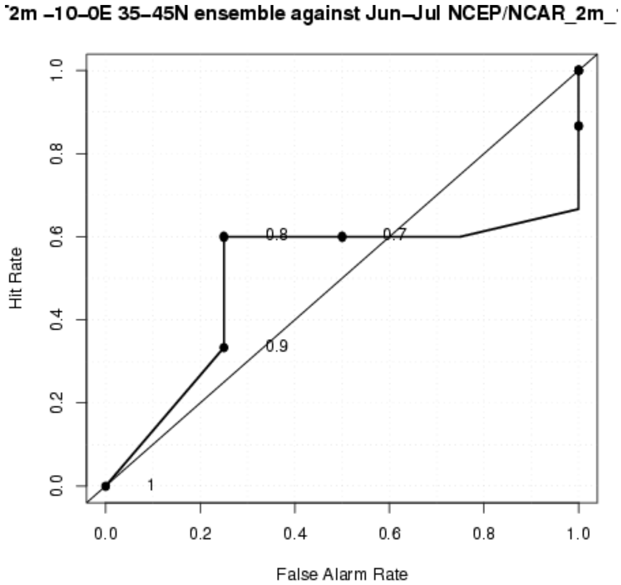
This score show quite good skill in predicting heat waves in southwestern Europe (the Iberian peninsula and France), indicated by the red colours for ROC areas larger than 0.5 (which is obtained by a system without any skill). There is very little skill in these areas in predicting the mean climate (e.g., van Oldenborgh et al, 2005). We suspected that this skill is due to soil moisture depletion: when the winter rains have been much weaker than normal, the soil dries up in summer, leading to higher temperatures on average (see also Ferranti and Viterbo, 2006), but first some cross-checks have to be made.

The same map can easily be made for a different forecasting system, the NCEP CFS. That system does not show skill in this area, casting doubt on the ECMWF skill score.

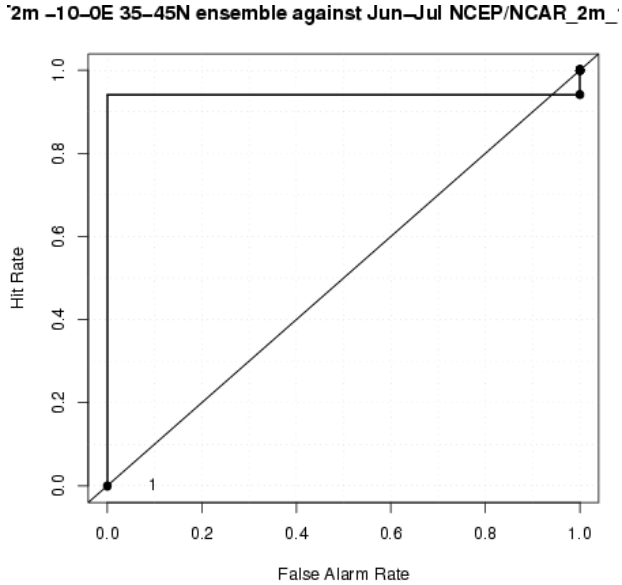




Further analysis of the area-averaged temperature over the Iberian Peninsula shows that the skill only is high when one chooses a threshold that is much higher than 80%, which excludes the years in the hindcast period with only 5 ensemble members. The curves shown are ROC vcurves that plot the Hit Rate against the False Alarm Rate for different sensitivities (number of ensemble members needed to issue a forecast). A system with no skill give a diagonal line with equal Hit Rates and False Alarm Rates. The area under the curve is a measure of the quality of the system.

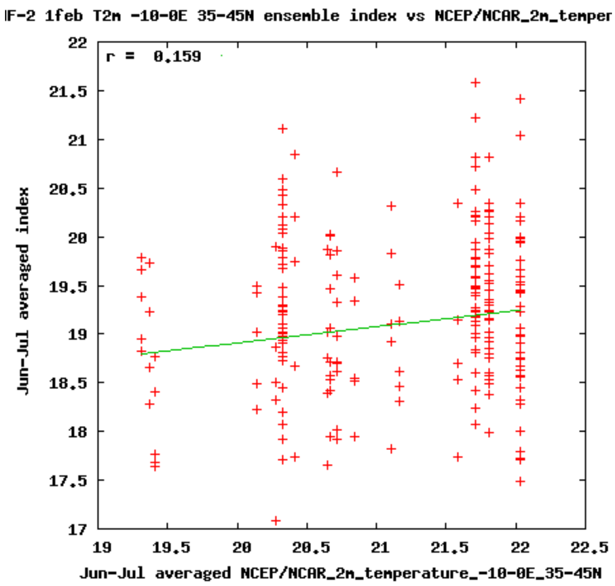


79% threshold



90% threshold

There is essentially no skill in the 79% threshold ROC curve. The same can be seen in scatterplots of the data: the reason that the warm summers in Spain and Portugal were forecast well is that they all fell in the last few years, for which there are 40 ensemble mebers. The apparent skill in the maps is therefore due to a coincidence of recent warm years and a changing ensemble size.



Conclusions

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 (ENSMEBLES data when it becomes available), make the system more user-friendly, and add confidence intervals to the skill scores.

Geert Jan van Oldenborgh & Caio Augusto Dos Santos Coelho, 7 March 2006

References

L. Ferranti and P. Viterbo, 2006: The European summer of 2003: sensitivity to soil water initial conmditions, submitted to J. Climate.

G.J. van Oldenborgh, M.A. Balmaseda, L. Ferranti, T.N. Stockdale en D.L.T. Anderson, *Evaluation of atmospheric fields from the ECMWF seasonal forecasts over a 15 year period*
J. Climate, 2005, 18, 16, 2970-2989, corr. J. Climate, 2005, 18, 5188-5198.