



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 persistence, optimal normal correlations 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':

Home → dimep → start

Climate Explorer
Starting point

Welcome, Geert Jan van Oldenborgh from KNMI
This page can be added to your bookmarks/favorites to avoid logging in.

Start by selecting a class of climate data from the right-hand menu. After you have selected the time series or fields of interest, you will be able to investigate it, correlate it to other data and generate derived data from it.

If you are new it may be helpful to study the examples.

Share and enjoy!

Introduction, results
News
Examples
Publications
Effects of ENSO on the weather
Probability of tropical cyclones

Select a time series
Daily station data
Daily climate indices
Pentad climate indices
Monthly station data
Monthly climate indices
Annual climate indices
User-defined time series
Upload your own time series

Select a field
6-hourly fields
Daily fields
Monthly observations
reanalysis fields
seasonal forecast means
seasonal forecast ensembles
scenario runs
User-defined
Upload your own field

Feedback
Geert Jan's home page

Home → dimep → selectfield/ensemb

Climate Explorer
Select a monthly field
Seasonal forecasts full ensembles

Select ensemble Choose an ensemble and press this button

| DEMETER | Meteo France | CERN | LODYC | INGV | ECMWF | MPI | UKMO | all | 3 |
|-----------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| var start | 1958-2001 | 1967-2001 | 1974-2001 | 1987-2001 | 1958-2001 | 1959-2001 | 1958-2001 | 1958-2001 | 2001 |
| 12m feb | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 12m may | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 12m aug | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 12m nov | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| prop feb | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| prop may | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| prop aug | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| prop nov | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| slp feb | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| slp may | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| slp aug | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| slp nov | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| z500 feb | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |

Select a time series
Daily station data
Daily climate indices
Pentad climate indices
Monthly station data
Monthly climate indices
Annual climate indices
User-defined time series
Upload your own time series

Select a field
6-hourly fields
Daily fields
Monthly observations
reanalysis fields
seasonal forecast means
seasonal forecast ensembles
scenario runs
User-defined
Upload your own field

Feedback
Geert Jan's home page

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:

Home → dimep → select

Climate Explorer
Field
ensemble ECMWF-2 1Feb T2m

Found ensemble members 0 to 39
ecmwf_2m_feb_00.dat
X axis: whole world in 240 1.50° steps. first point at 0.00° E
Y axis: regular grid with 120 1.50° steps. first point at 90.00° S
Z at 1.00
Monthly data available from Feb1967 to Jul2006 (222 months)
Variable t2m (° surface 2 metre temperature) in K, defined at 2 level 0

Extract timeseries
latitude: 30 °N 90 °N (leave second field blank for one point)
longitude: 0 °E 360 °E
boundaries: halfway grid points
make average set of grid points
units convert to Celsius leave in K
Make time series

Create a field with derived data
New time scale: yearly
New variable: number of 12m 1Feb T2m
Threshold: greater than 25 K
Make new field

Apply yearly high-low-pass filter
Apply a yearly high-pass | running-mean | filter
with cut-off value of 2 years
Filter

Download ensemble ECMWF-2 1Feb T2m
The ECMWF member states do not permit us to give you access to the raw data. Please consult the ECMWF [FAQ](#) or [seasonal forecasting](#) website for further information.

Introduction, results
News
Examples
Publications
Effects of ENSO on the weather
Probability of tropical cyclones

Select a time series
Daily station data
Daily climate indices
Pentad climate indices
Monthly station data
Monthly climate indices
Annual climate indices
User-defined time series
Upload your own time series

Select a field
6-hourly fields
Daily fields
Monthly observations
reanalysis fields
seasonal forecast means
seasonal forecast ensembles
scenario runs
User-defined
Upload your own field

Investigate this field
Plot difference with a field
Compute mean and higher moments
Correlate with a time series
Pointwise correlations with a field
only observations
only reanalyses
only seasonal forecasts
only scenario runs
only user-defined fields
Spatial correlations with a field
only observations
only reanalyses
only seasonal forecasts
only scenario runs
only user-defined fields
Verify field against observations
Make EOFs

Feedback
Geert Jan's home page

Home → dimep → regionverificationform

Climate Explorer
Field verification
ensemble ECMWF-2 1Feb T2m

Converting ensemble ECMWF-2 1Feb T2m from K to Celsius

Verifying Temperature field
Temperature 1858-now anomalies: HadCRUT1 (Jones & Parker T2m/SST analysis)
1857-now variance adjusted: HadCRUT1, HadCRUT1v
Land 1851-now anomalies: GISS-TEMP (Jones T2m analysis) number of stations:
variance adjusted: CRUTEM1, CRUTEM1v
1901-2000: CSIRO New 1.5° analysis (land only) OM World, New World (SARGE)
Air 1800-1997: COADS Tat
Temperature 1.5° 2.5° ERA-40
T2m NCEP/NCAR

Map verification measures
 Correlation of the ensemble mean
 Root mean square error (RMSE) of the ensemble mean
 Mean absolute error (MAE) of the ensemble mean
 Brier score (alternative)
 BSS wrt climatology
 Resolution
 Reliability
 Uncertainty
 tercile RPS alternative quintile RPS
 tercile RPSS wrt climatology
 Area under under the ROC curve, alternative1, alternative2
 Only compute the netcdf files with observations and forecasts.

Timeseries verification measures
 Plot likelihood
 Deterministic scores for the ensemble mean (correlation, root mean squared error, and mean absolute error)
 Brier score
 Plot reliability diagram
 Compute Ranked Probability Score for terciles
 Plot ROC curve for number of ensemble members below some threshold (alternative1 alternative2)
 Plot ROC curve varying the model threshold
 Only compute the observations/forecasts table

Threshold
For measures that require a threshold, use 90 %
Use bits show R loglike

Introduction, results
News
Examples
Publications
Effects of ENSO on the weather
Probability of tropical cyclones

Select a time series
Daily station data
Daily climate indices
Pentad climate indices
Monthly station data
Monthly climate indices
Annual climate indices
User-defined time series
Upload your own time series

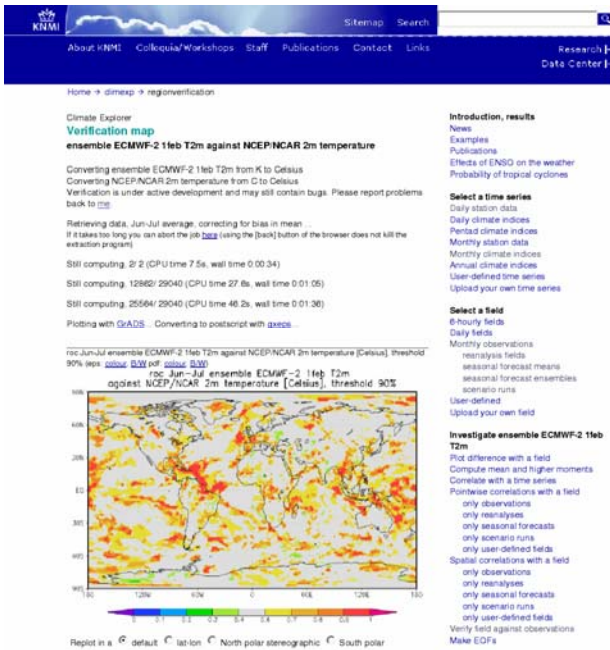
Select a field
6-hourly fields
Daily fields
Monthly observations
reanalysis fields
seasonal forecast means
seasonal forecast ensembles
scenario runs
User-defined
Upload your own field

Investigate this field
Plot difference with a field
Compute mean and higher moments
Correlate with a time series
Pointwise correlations with a field
only observations
only reanalyses
only seasonal forecasts
only scenario runs
only user-defined fields
Spatial correlations with a field
only observations
only reanalyses
only seasonal forecasts
only scenario runs
only user-defined fields
Verify field against observations
Make EOFs

Feedback
Geert Jan's home page

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

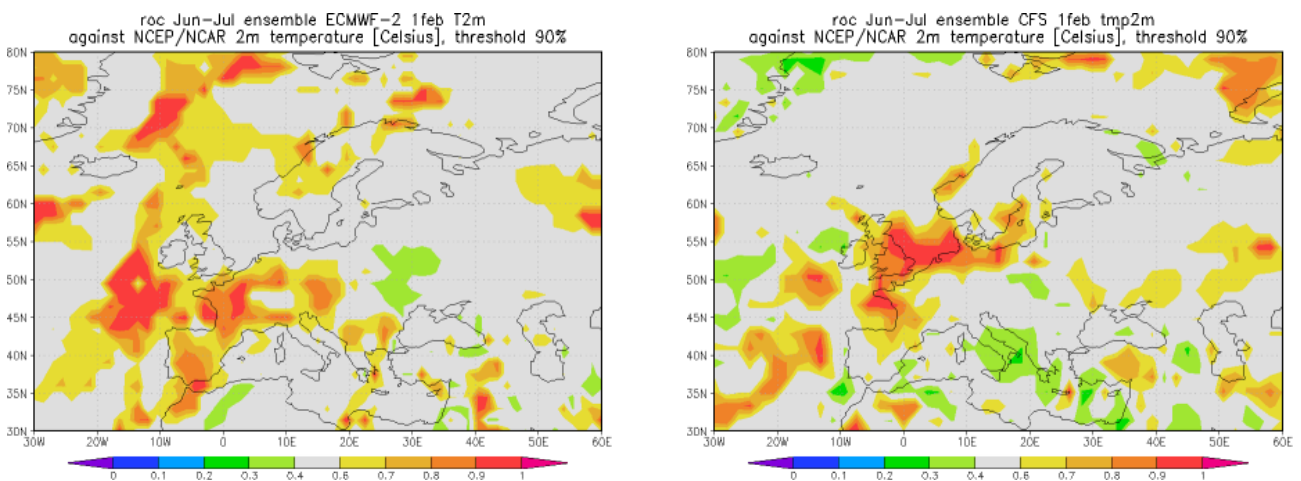
this time is needed to read the forecast data from the data store:



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.

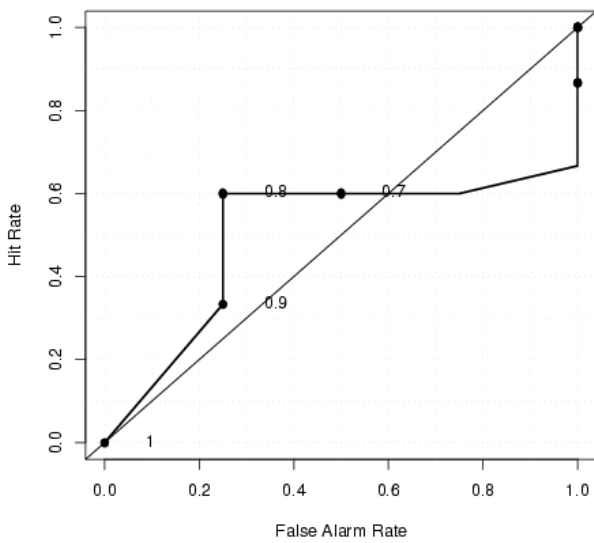
The same map can easily be made for a different forecasting system, the NCEP CFS. That system does not show skill in this area, casting doubt on the ECMWF skill score.



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

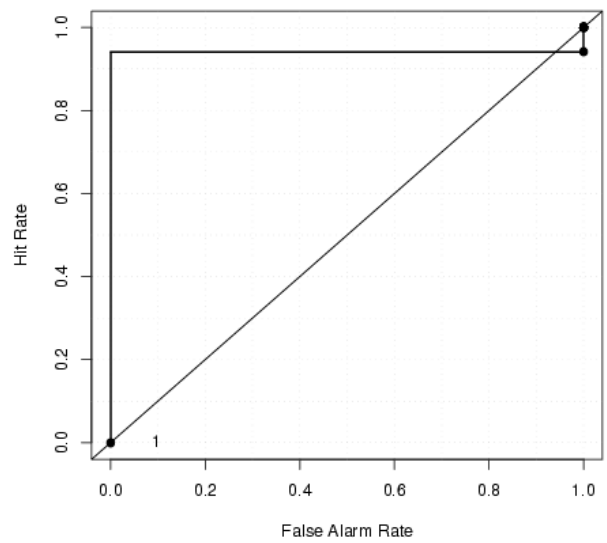
needed to issue a forecast). A system with no skill give a diagonal line with equal Hit Rates and False Alarm Rates. The area under the curve is a measure of the quality of the system.

'2m -10-0E 35-45N ensemble against Jun-Jul NCEP/NCAR_2m_'



79% threshold

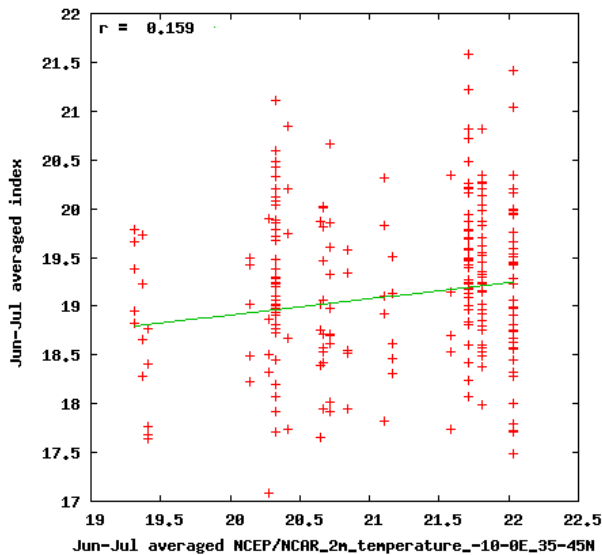
'2m -10-0E 35-45N ensemble against Jun-Jul NCEP/NCAR_2m_'



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 members. The apparent skill in the maps is therefore due to a coincidence of recent warm years and a changing ensemble size.

If-2 1feb T2m -10-0E 35-45N ensemble index vs NCEP/NCAR_2m_temper



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 conditions, submitted to J. Climate.

G.J. van Oldenborgh, M.A. Balmaseda, L. Ferranti, T.N. Stockdale and 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.