Seasonal-to-decadal climate Prediction for the improvement of European Climate Services

THEME ENV.2012.6.1-1

G.A. no 308378

WP 2.1 ***Forecast quality of s2d systems***

Deliverable 21.1 *R-based verification package available and built into the Climate Explorer web service*

Date: 03/11/2014

Deliverable Title	R-based verification package available and built into the Climate Explorer web service		
WP number and title	WP 2.1: Forecast quality of s2d systems		
Lead Beneficiary	*KNMI*		
Contributors	*G. J. van Oldenborgh*	*KNMI*	
	David Stephenson	*UNEXE*	
	Stefan Siegert	*UNEXE*	
	Chris Ferro	*UNEXE*	
	Chloe Prodhomme	*IC3*	
	Omar Bellprat	*IC3*	
	Nicolau Manubens	*IC3*	
	Louis-Philippe Caron	*IC3*	
Creation Date	14/10/2014	**Version number**	0.3
Deliverable Due Date	31/10/2014	**Actual Delivery Date**	03/11/2014
Nature of deliverable		R - Report	
	X	P - Prototype	
		D - Demonstrator	
		O - Other	
Dissemination Level/ Audience		PP - Public	
	X	PU - 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	

Version	Date	Modified by	Comments
0.1	14/10/2014	G. J. van Oldenborgh	Initial version
0.2	30/10/2014	G. J. van Oldenborgh	Included milestone 2.1 and screenshots
0.3	3/11/2014	David Stephenson, Chloe Prodhomme	Better explanation of aims, added IC3 work

INDEX

1. Executive summary

The skill of seasonal and decadal climate forecasting systems is generally evaluated using verification performance measures of sets of past forecasts (hindcasts) and corresponding observations. To assess whether there has been improvement in skill, it is necessary to have suitable methods for comparative verification that can fairly allow for differences in the number of forecasts issued (ensemble size), quantify sampling uncertainty in differences of scores, and account for dependencies in the data (e.g. serial time-dependence in the observations). UNEXE and IC3 partners in the SPECS project have developed such methods and have implemented them in two new open-access R software packages: SpecsVerification and s2dverification. Furthermore, KNMI has incorporated the SpecsVerification tools into the widely-used KNMI Climate Explorer web application, which provides a quick and user-friendly interface for non-specialists.

2. Project objectives

With this deliverable, the project has contributed to the achievement of the following objectives (see DOW Section B.1.1.2):

No.	Objective	Yes	No
1.	To achieve an objective exhaustive *evaluation* of current forecast quality from dynamical, statistical, and consolidated systems to identify the factors limiting s2d predictive capability	X	
2.	To test specific hypotheses for the improvement of s2d predictions, including novel mechanisms responsible for high-impact events using a *process-based verification* approach		X
3.	To develop innovative methods for a comprehensive *forecast quality assessment*, including the maximum skill currently attainable	X	
4.	To facilitate the *integration of multidimensional observational data* of the atmosphere-ocean-cryosphere-land system as sources of initial conditions, and to validate and calibrate climate predictions	X	
5.	To achieve an *improved forecast quality at regional scales* by better initialising the different components, an increase in the spatial resolution of the global forecast systems and the introduction of important new process descriptions		X
6.	To assess the best alternatives to characterise and deal with the *uncertainties in climate prediction* from both dynamical and statistical perspectives for the increase of forecast reliability		X
7.	*To achieve reliable and accurate local-to-regional predictions* via the combination and calibration of the information from different sources and a range of state-of-the-art regionalisation tools	X	

No.	Objective	Yes	No
8.	*To illustrate the usefulness* of the improvements for specific applications and develop methodologies to better communicate actionable climate information to policy-makers, stakeholders and the public through peer-reviewed publications, e-based dissemination tools, multi-media, examples for specific stakeholders (energy and agriculture), stakeholder surveys, conferences and targeted workshops		X
9.	*To support* the European contributions to *WMO research initiatives* on s2d prediction such as the GFCS and enhance the European role on the *provision of climate services* according to WMO protocols by creating examples of improved tailored forecast-based products for the GPCs and participating in their transfer to worldwide RCCs and NHMSs.	X	

3. Detailed report on the deliverable

The software packages can be downloaded from the R CRAN site (http://CRAN.R-project.org/) and are documented here:

- http://www.rdocumentation.org/packages/SpecsVerification
- http://www.rdocumentation.org/packages/s2dverification

These packages are still under development with new ideas that have emerged from SPECS partners at the 3rd general assembly. More specifically, on one side we are planning to optimize the s2dverification package to make it more suitable to handle the huge dataset, we have to deal with in the climate forecasting field. On the other side, we will work on an efficient and optimized interface between SpecsVerification and s2dverification. This will be beneficiary for the users of both packages, improving the big dataset handling in SpecsVerification and adding lot of new scores and statistical methods in s2dverification. The package s2dverification is not (yet) implemented on climate explorer, but we are working on improving the portability of the package for new users (CNRM).

This deliverable fulfills the aims of the Task 2.1.1 by creating new tools in these areas:

1. Fair scores for ensemble were recently shown to avoid biases in the perceived ensemble quality due to finite number of ensemble members (Ferro, 2013, DOI 10.1002/qj.2270). The fair Brier Score and the fair CRPS were implemented as part of the package. Furthermore, methods to analyze the difference of fair CRPS and fair Brier Scores between two ensemble forecasting systems were developed. Confidence intervals for these score differences are provided automatically to facilitate significance tests.

   Advanced rank histogram analysis was implemented as part of the package. It is possible to plot rank histograms in raw mode, and on probability paper to assess deviations from flatness, following results from Broecker (2008, DOI 10.5194/npg-15-661-2008). Significance tests for flatness of rank histograms published in Jolliffe and Primo (2008, DOI 10.1175/2007MWR2219.1) are part of the package.

The reliability diagram for probability forecasts of binary events is improved by adding consistency bars to assess sampling fluctuations, following Broecker and Smith (2007, DOI 10.1175/WAF993.1). Brier score decomposition of these forecasts is implemented to assess the attributes Reliability and Resolution of probability forecast, following Murphy (1973, DOI 10.1175/1520-0450(1973)012<0595:ANVPOT>2.0.CO;2).

2. Ensemble dressing is a versatile method to transform discrete ensemble forecasts into smooth continuous forecast distributions. An ensemble dressing method called Affine Kernel Dressing published in Broecker and Smith (2007, DOI 10.1111/j.1600-0870.2008.00333.x) has been implemented as part of the package. It involves a correction of ensemble mean and standard deviation by an affine transformation of the ensemble, and adjustment of the width of the dressing kernels. The dressing parameters are calculated by optimizing the continuously ranked probability score (CRPS) of the dressed ensemble over a training period. An analytical expression for the CRPS was derived and implemented, significantly improving the speed and accuracy of the calculations. A function to visualize dressed ensemble forecasts is provided as part of the package, and differences between CRPS of two dressed ensemble forecasts can be analyzed.

3. Commonly used statistical tests of hypothesis, also termed inferential tests, which are available to meteorologists and climatologists all require independent data in the time series to which they are applied. Unfortunately, most of the time series, which are usually handled, are actually serially dependent. A common approach to deal with such a serial dependence is to replace in those statistical tests the actual number of data by an estimated effective number of independent data which is computed from a classical and widely used formula relying on the autocorrelation function. In spite of being perfectly demonstrable under some hypotheses, this formula provides unreliable results on practical cases, for two different reasons. First, the formula has to be applied using the estimated autocorrelation function, which bears a large uncertainty due to the usual shortness of the available time series. Second, the derivation of this formula is done under the hypothesis of identically distributed data, which is often not valid in real climate or meteorological problems.

   Following the results of Guemas et al, (2013), seven new functions have been developed and included in this R package to address this. They allow for a new method to estimate the equivalent sample size of a time series (which is necessary for parametric tests but also for bootstrapping methods to choose the window length for the block bootstrap). Its performance is better than previous existing methods for small samples such as the ones we have in climate prediction.

   Climate predictions initialized from an observationally-constrained state systematically drift toward the attractor of an unconstrained model, i.e., develop model bias, which makes the use of a-posteriori correction methods essential to disentangle the climate signal of interest from the bias. Fuckar et al. (2014) proposes correcting the model bias by applying a linear regression of the bias on the observationally-based state. This new method shows a reduction of error with respect to established bias correction methods for tropical and high-latitude sea surface temperature, and the Northern Hemisphere sea ice extent in near-term climate predictions from the first forecast year up to the fifth forecast year. This method is now available in the s2dverification package.

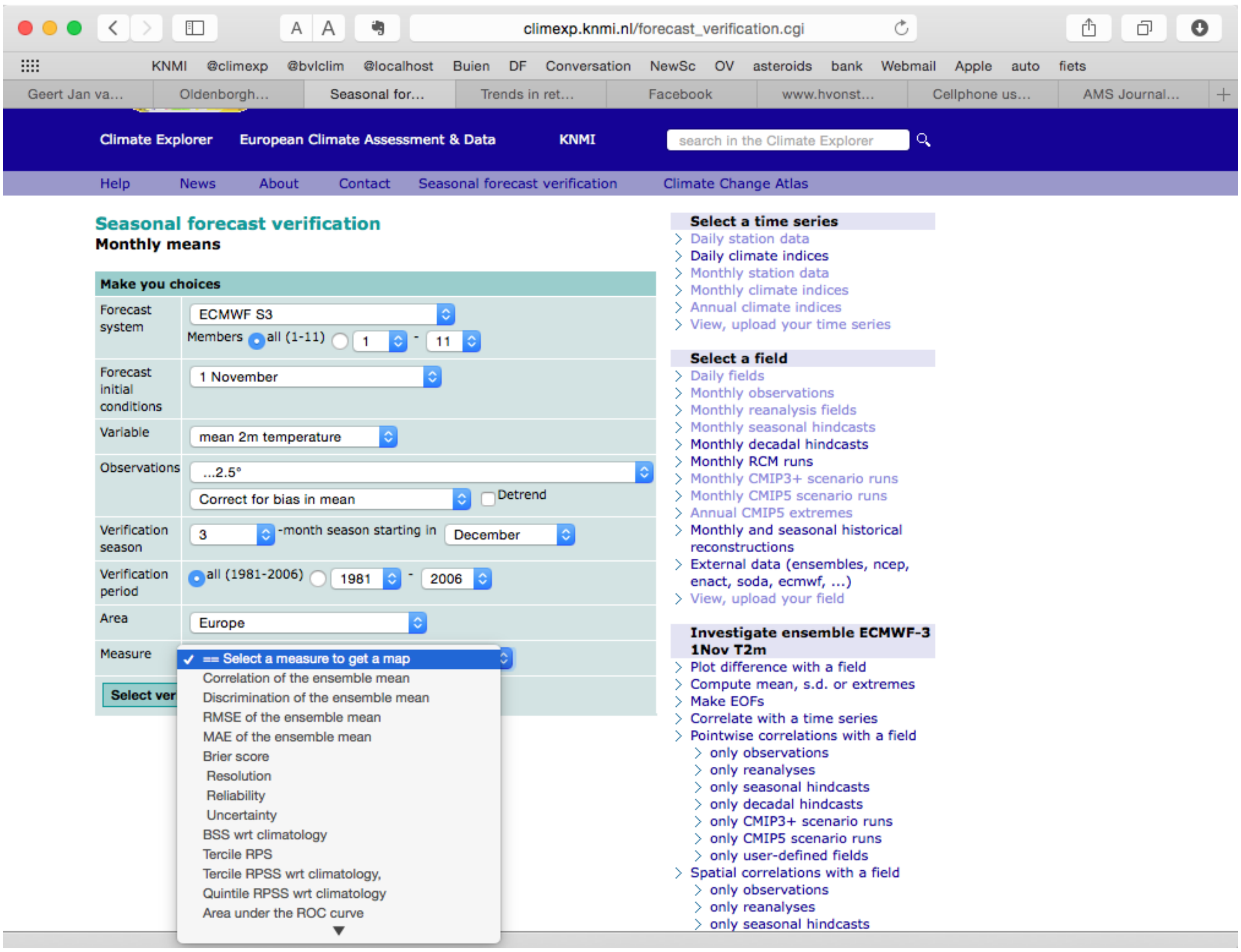
4. Finally, these R routines have been incorporated in the KNMI Climate Explorer web site, which has a user-friendly interface to plot skill scores for a variety of seasonal forecast systems. At

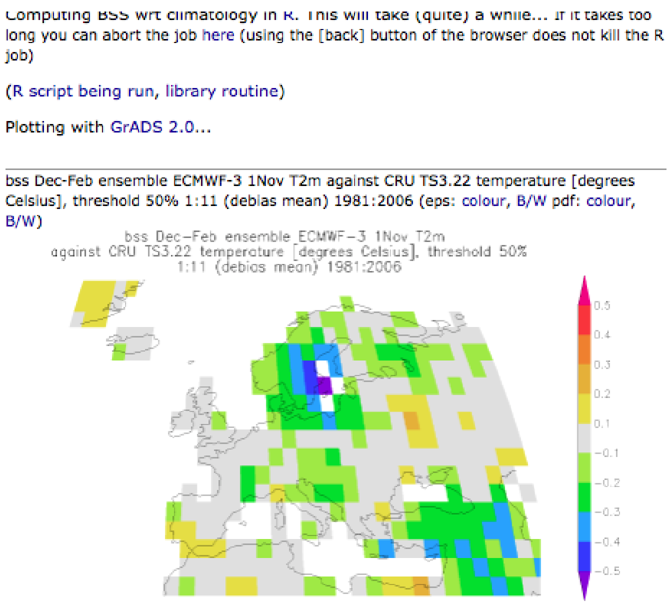
the moment the database of hindcasts only runs up to the DEMETER experiments, but this task will be continued by adding the ENSEMBLES and current forecast system hindcasts to the extent that these will be provided by the forecasting centres. We will also update the routines to the newer version being developed at U Exeter and IC3, which will become available in the next month.

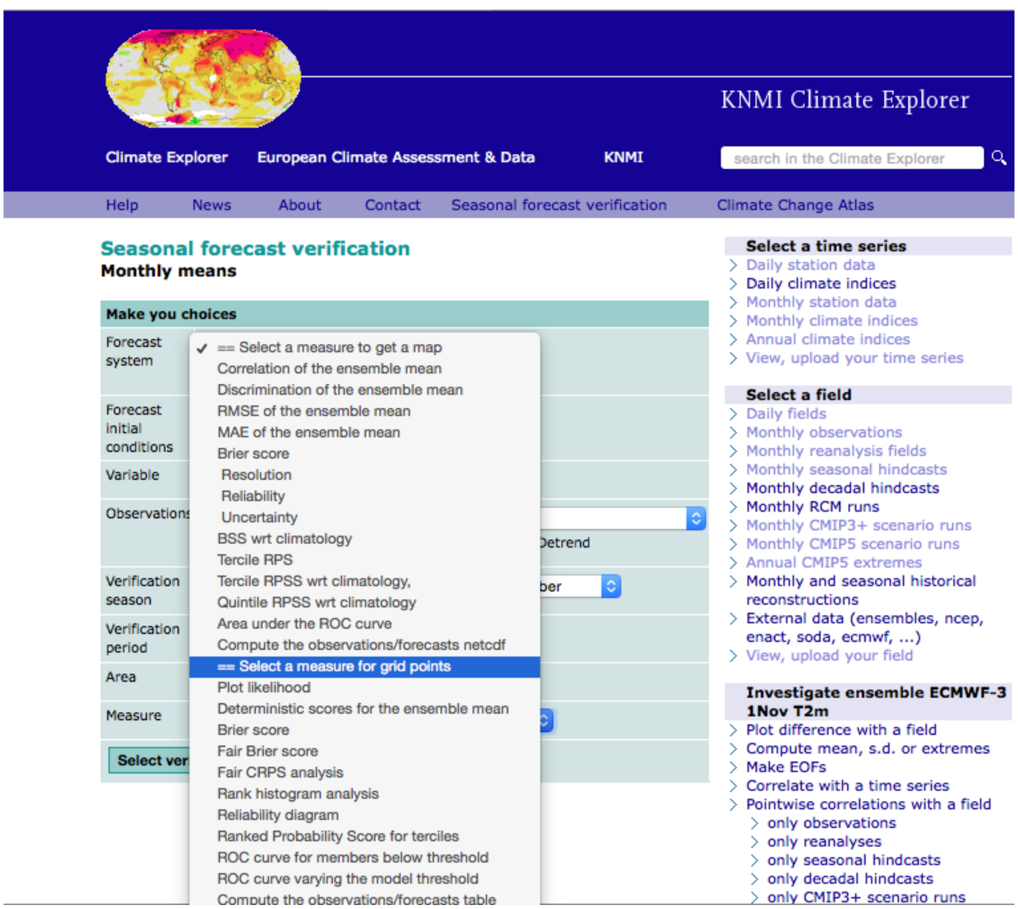
Some screenshots of the forecast system at the KNMI Climate Explorer, currently only working with data from older seasonal forecast systems.

- The list of verification measures that produces a map for the value at each grid point,
- An example map (Brier Skill Score for winter temperatures in Europe from a Nov 1 start date).
- The list of verification measures that consider all grid points in the region indicated on the map,
- An example of this analysis (rank histogram for the winter forecasts)

(It should be noted that this is for an old forecast system, the hindcast data from the current systems is not yet available.)



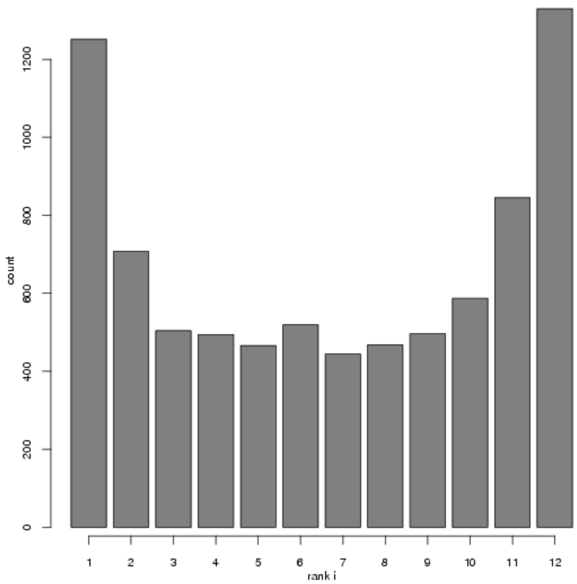

Computing BSS wrt climatology in R. This will take (quite) a while... If it takes too long you can abort the job here (using the [back] button of the browser does not kill the R job)
(R script being run, library routine)
Plotting with GrADS 2.0...
bss Dec-Feb ensemble ECMWF-3 1Nov T2m against CRU TS3.22 temperature [degrees Celsius], threshold 50% 1:11 (debias mean) 1981:2006 (eps: colour, B/W pdf: colour, B/W)
bss Dec−Feb ensemble ECMWF−3 1Nov T2m
against CRU TS3.22 temperature [degrees Celsius], threshold 50%
1:11 (debias mean) 1981:2006
0.5
0.4
0.3
0.2
0.1
−0.1
−0.2
−0.3
−0.4
−0.5


KNMI Climate Explorer
Climate Explorer
European Climate Assessment & Data
KNMI
search in the Climate Explorer
Help
News
About
Contact
Seasonal forecast verification
Climate Change Atlas
Seasonal forecast verification
Monthly means
Make you choices
Forecast system
Forecast initial conditions
Variable
Observations
Verification season
Verification period
Area
Measure
Select ver
== Select a measure to get a map
Correlation of the ensemble mean
Discrimination of the ensemble mean
RMSE of the ensemble mean
MAE of the ensemble mean
Brier score
Resolution
Reliability
Uncertainty
BSS wrt climatology
Tercile RPS
Tercile RPSS wrt climatology,
Quintile RPSS wrt climatology
Area under the ROC curve
Compute the observations/forecasts netcdf
== Select a measure for grid points
Plot likelihood
Deterministic scores for the ensemble mean
Brier score
Fair Brier score
Fair CRPS analysis
Rank histogram analysis
Reliability diagram
Ranked Probability Score for terciles
ROC curve for members below threshold
ROC curve varying the model threshold
Compute the observations/forecasts table
Detrend
ber
Select a time series
Daily station data
Daily climate indices
Monthly station data
Monthly climate indices
Annual climate indices
View, upload your time series
Select a field
Daily fields
Monthly observations
Monthly reanalysis fields
Monthly seasonal hindcasts
Monthly decadal hindcasts
Monthly RCM runs
Monthly CMIP3+ scenario runs
Monthly CMIP5 scenario runs
Annual CMIP5 extremes
Monthly and seasonal historical reconstructions
External data (ensembles, ncep, enact, soda, ecmwf, ...)
View, upload your field
Investigate ensemble ECMWF-3 1Nov T2m
Plot difference with a field
Compute mean, s.d. or extremes
Make EOFs
Correlate with a time series
Pointwise correlations with a field
only observations
only reanalyses
only seasonal hindcasts
only decadal hindcasts
only CMIP3+ scenario runs

Computing Fair CRPS Analysis in R. This will take (quite) a while... If it takes too long you can abort the job here (using the [back] button of the browser does not kill the R job)

(R script being run, library routine)

ensemble ECMWF-3 1Nov T2m verified against CRU TS3.22 temperature [degrees Celsius] 1981:2006 in 30:75N, -30:50E correcting for bias in mean (raw data, PDF)



Verification measures in R have been kindly provided by the RCLIM initiative and ongoing research on verification by the EU project SPECS

4. References

Not applicable.

5. List of publications

Peer reviewed articles:

- *Guemas V., Auger L, Doblas-Reyes F., 2013, Hypothesis testing for auto-correlated short climate time series. Journal of Applied Meteorology and Climatology, doi:10.1175/JAMC-D-13-064.1.*
- *Fuckar N, Volpi D, Guemas V, Doblas-Reyes F. A posteriori adjustment of near-term climate predictions: Accounting for the bias dependence on the observed initial state, in preparation for Geophys. Res. Lett.*

Plan for future publication:

More papers envisaged using these new tools to compare forecasting systems (in collaboration with all SPECS partners).

6. Efforts for this deliverable

How many person-months have been used up for this deliverable? (includes Milestone 2.1)

Not that this is the effort to date, part of the work will be performed over the next few months.

Partner	Person-months (actual)	Person-months (in-kind)	Period covered
1. KNMI	1	-	M12-18
2. U Exeter	12	3	M1-18
6. IC3	7.14	-	M1-18
Total	**20.14**	**3**	

7. Sustainability

The verification routines are being used in the evaluation of the KNMI empirical seasonal forecast system (Task 5.1.1) and the U Reading empirical decadal forecast system (Task 5.1.2).

The s2dverification package has been used during the ITCP-WCRP summer school on extreme events and will be use soon by the CNRM.

This deliverable has benefited from a wide engagement with all the other partners and WPs. An online survey was first conducted involving all partners to find out how they did verification and what they thought were issues that needed to be addressed by new methods. Two verification workshops have taken place – the most recent one at the 3rd General Assembly in Oct 2014 involved around 50 people from across the whole of SPECs and from the EUPORIAS project. Some of the new tools have also recently been implemented on verification platforms of other projects e.g. the German MiKlip project.