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

THEME ENV.2012.6.1-1

G.A. no 308378

Final Report

Date:11/04/2017

Title	Final Report		
Lead Beneficiary	*BSC*		
Contributors	Francisco J. Doblas-Reyes	BSC/IC3	
	Mar Rodríguez	BSC/IC3	
	Michel Déqué	METEOFRANCE	
	Johanna Baer	UHAM	
	Klaus Wyser	SMHI	
	Antje Weisheimer	UOXF	
	Caio Augusto dos Santos Coelho	INPE	
	Matteo De Felice	ENEA	
	Adam A. Scaife	MetOffice	
	David Stephenson	UNEXE	
	José Manuel Gutiérrez	CSIC	
	Tim Stockdale	ECWMF	
	Eric Guilyardi	CNRS	
Start Month	01/11/2012	**End month**	31/01/2107
Delivery Date	11/04/2017	**Actual Delivery Date**	11/04/2017

Version	Date	Modified by	Comments
1	10/04/2017	Francisco Doblas-Reyes (BSC)	First draft

INDEX

1 Main scientific and technical results

SPECS has delivered a new generation of European climate forecast systems, including initialised Earth System Models (ESMs) and efficient regionalisation tools to produce quasi-operational and actionable local climate information over land at seasonal-to-decadal time scales with improved forecast quality and a focus on extreme climate events, and provide an enhanced communication protocol and services to satisfy the climate information needs of a wide range of public and private stakeholders. This challenge is at the centre of SPECS' overarching objectives, aiming to push the boundaries of the capability in climate forecasting and dissemination of information for the development of climate services. The objectives have been achieved through a set of scientific and technical results that are summarised in the following, including a few examples of illustrative outcomes, taken out from the myriad of interesting results obtained in the project.

1.1 Process evaluation and forecast quality assessment

Emerging climate services in Europe provide seasonal-to-decadal (s2d) forecasts to private and public sectors to guide strategic planning. In these services, longer time scales are considered than what National Meteorological and Hydrological Services (NMHSs) traditionally do. For instance, climate adaptation strategies are informed by s2d forecasts and information on the quality of these forecasts because forecast quality is as important as the product itself. Both natural variability and trends provide skill in s2d forecasts. Skilful s2d forecasts rely on the correct representation of physical processes in the climate system having an impact on meteorological variables that exceeds the noise from fast atmospheric variability. On seasonal time scales, varying sea surface temperatures (SST) during for instance the El Niño Southern Oscillation (ENSO) events and anomalous snow cover can provide predictability locally and to remote regions via teleconnections. On the longer term, the radiative effect of changing greenhouse gas concentrations and aerosols and slow oceanic processes are relevant. It is hence fundamental to identify these processes, assess their representation in current climate forecast systems, provide suggestions to improve them in the new generation of forecast systems and to inform interested stakeholders on the physical reasons behind specific forecast quality levels. The correct representation of high-impact climatic events and extremes are most relevant because they expose the vulnerability of many sectors of the society.

The assessment of forecast quality is necessary for deciding which forecast systems can be considered both skilful and reliable and whether the quality of the new forecast systems has improved. Forecast quality assessment is also necessary for creating stakeholder credibility in the climate services developed. SPECS has comprehensively assessed the forecast performance of s2d hindcast datasets available. Thorough forecast quality assessment has been performed to explore the improvements in performance across the new generation of forecast systems developed in SPECS. New forecast quality evaluation tools have been developed and made publicly available using the R language responding to the feedback provided by the consortium and other research and operational initiatives; they have become world-leading verification system for s2d prediction and climate services. The common tools have been used to comprehensively assess existing s2d hindcast datasets, in operational systems (EUROSIP, US National Multi-Model Ensemble, APCC/CliPAS) and also the predictions performed in SPECS. Systematic mean biases have been explored using process-based metrics and metrics related to key physical processes and design of forecasting systems.

As an example, the temperature and precipitation skill in the EUROSIP forecast systems have been assessed for winter and summer seasons and was presented as part of deliverable D21.3. Figure 1 shows that in summer the skill is significant around the Mediterranean Sea (for Météo-France, ECMWF, Glosea5), Central (ECMWF, NCEP) and Eastern Europe (ECMWF) for temperature. The skill is mainly due to the warming trend. The system presenting skill over the largest region is ECMWF. The skill of precipitation is very sparse and low, the only region for which several systems have skill is the Iberian Peninsula. For both precipitation and temperature, the skill is relatively poor in winter for all the considered systems (not shown).

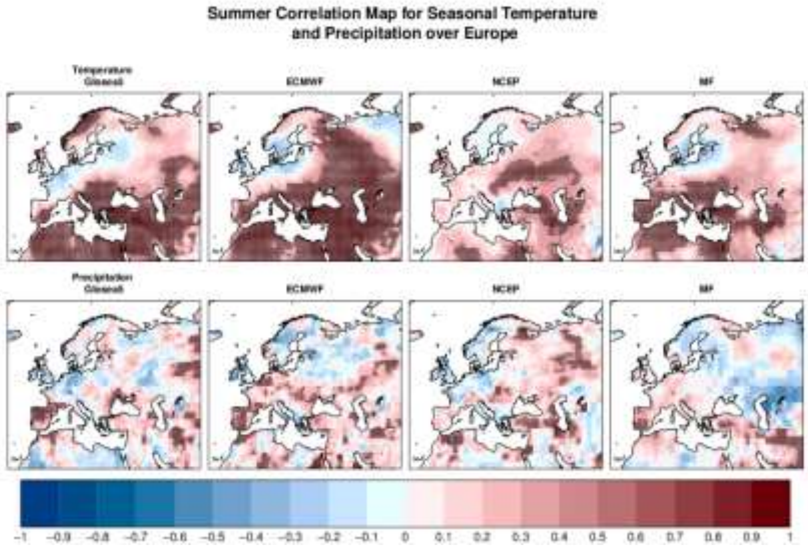


Figure 1: Correlation for summer season. Seasonal (JJA) temporal correlation for European region over 1992-2012. The four maps on the top and the bottom correspond to seasonal temperature and precipitation, respectively. Dots mark the areas where the skill is significant at 95% confidence level.

A particularly relevant study discusses the role of observational uncertainty for the verification of seasonal predictions (Massonnet et al., 2016). The study evaluates the correlation skill to predict Niño3.4 SST anomalies and finds that models score systematically better when evaluated against more sophisticated observational data products. Figure 2 illustrates that the correlation is systematically higher when using the satellite derived ESA-CCI SST product as when compared to the in-situ SST product ERSST4 consistent across a range of different models and members thereof. ESA-CCI is in addition independent of ERSST4 and independent of the models.

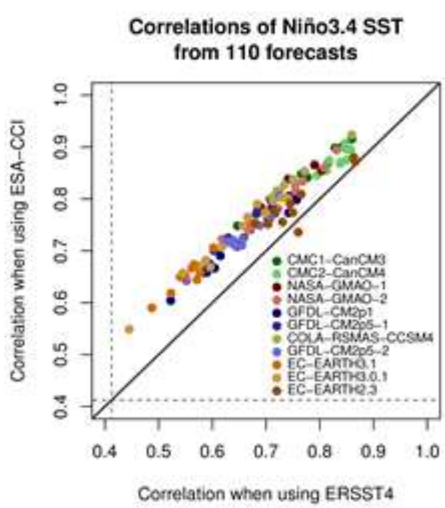


Figure 2: Systematic dependence of correlation on verification product. Each dot highlights the relationship between the correlation of one August SST forecast verified against the ESA-CCI and ERSST4. The reference period is 1993 to 2009. There are 110 forecasts in total (11 models, 10 members each). The solid line is the 1:1 line and delimits regions where ESA-CCI or ERSST4 scores better. The dashed lines are the threshold above which correlations are significant at the 0.05 level.

1.2 Improved initialisation and ensemble generation

The essential, and practical, difference between a climate scenario and a forecast is the start of the numerical model from an initial state of the system as close as possible to available observed conditions at the time of the forecast start. Therefore, a good forecast must use the most accurate initial state available. The availability of long time series of ocean/atmosphere/land/sea-ice initial conditions is critical for initialising larger samples of hindcasts that can provide results robust enough to extract valid conclusions for the stakeholders to make informed decisions and produce actionable climate information. The frequency of the sampling has been also suboptimal until now. Decadal hindcasts in previous initiatives have been started every five years during the last 50 years, which results in a too small sample to assess the quality of decadal predictions reliably. As the initial state is not perfectly known from the observations, the best way to take into account the uncertainty is to produce ensemble forecasts.

SPECS has worked on improving the description of the initial state of the slowly evolving part of the system (ocean, soil, cryosphere). The work has focused on the assessment of the impact of improved soil moisture, snow cover and sea-ice initial conditions. The impact of the hindcast sample size has been also evaluated. The analysis has been based upon a set of coordinated experiments and the evaluation of the forecast skill improvements involved the tools developed in the project

Climate forecasting being essentially probabilistic, one of the most common ways of estimating a probability forecasts with dynamical systems consists in running ensembles of simulations started from slightly different initial conditions. However, the estimation of the optimal set of initial conditions given a certain amount of computing resources is not trivial. The work has focused on sensitivity experiments individually testing the different ensemble generation techniques or initialization methods in a particular forecast system. The results of the ensemble generation technique/initialisation methods pointed at the large difficulty of improving over the current benchmark solutions usually employed in an operational context, in particular as far as the impact on the forecast reliability is concerned.

An example of the sensitivity experiments to the initialisation of the individual model components is offered by the evaluation of the seasonal prediction skill for European summer temperatures in two sets of seasonal hindcast experiments initialised on the first of May each year over 1981-2010 and performed with the Max Planck Institute Earth System Model (MPI-ESM) using the same initialisation setup and model configuration but different soil schemes (similar experiments were performed by other forecast systems). For one hindcast set the MPI-ESM bucket soil scheme is used, for the other set a new 5-layer soil-hydrology scheme is used. Both hindcast sets are initialised from fully-coupled assimilation experiments where different atmospheric and oceanic variables are nudged, but no direct assimilation of soil-moisture data is performed. The experiments on initial soil moisture have shown an improvement in seasonal skill for European summer temperatures only if the five-layer soil-hydrology scheme is switched on (Fig. 3). Possible causes for this skill improvement were investigated. On the one hand, the indirect soil-moisture assimilation is more realistic causing a different behaviour of land-atmosphere coupling in the five-layer scheme compared to the bucket scheme. On the other hand, the prediction of the atmospheric blocking frequency is improved, reflecting more realistic persistence of large-scale weather patterns over Europe.

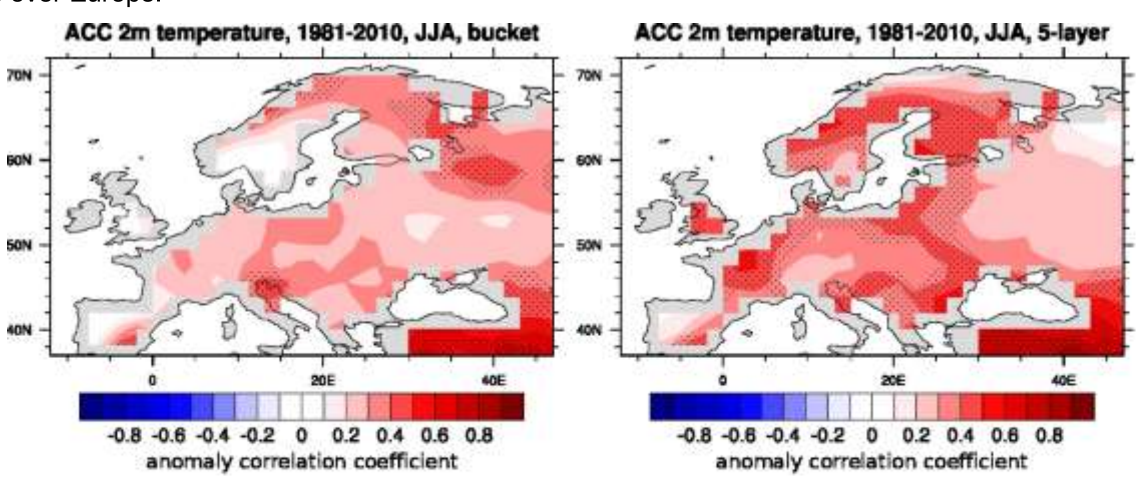
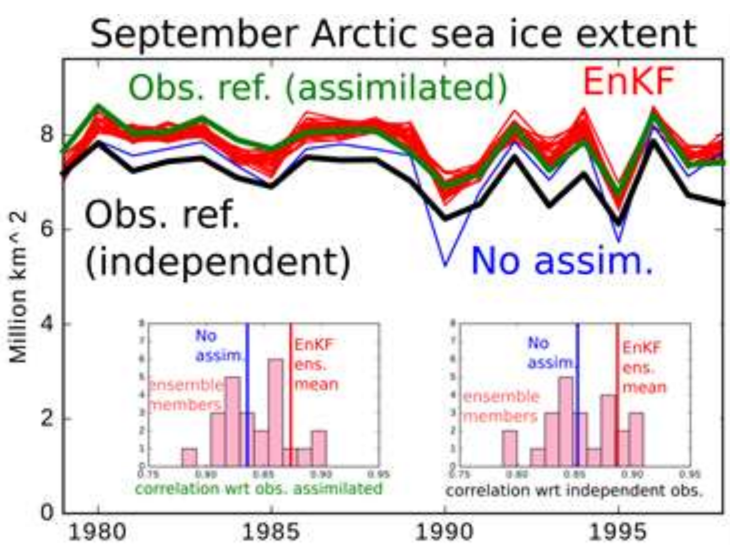


Figure 3: Anomaly correlation coefficient (ACCs) for JJA-mean 2-meter air temperature over Europe, computed from the ensemble mean of seasonal hindcasts performed with the MPI-ESM forecast system started each year on 1 May within

1981-2010 with respect to ERA-Interim reanalysis data. Results are shown for both the bucket soil scheme (left), and the 5-layer soil-hydrology scheme (right). Dotted regions indicate significant ACCs at the 95% level obtained from a distribution of 1000 re-sampled 10-member ensemble means.

The hindcasts with soil moisture driven by reanalyses show, as expected, a significant increase in temperature predictability, with the regional models (RACMO and ALADIN) as with the global model (ARPEGE). More interesting is the fact that precipitation skill is improved as well, and to some extent, mean sea level pressure over Europe. The forecast quality of boreal summer heat waves is further increased when starting on the first of June instead of the first of May, and additional skill is obtained by using observed SSTs. These results have no practical implication in actual forecasting, but show that the occurrence of heat waves is not pure atmospheric chaos, and could be predictable at seasonal scale with improved models. A SPECS Technical note (#6) has been issued with these results, and will be followed by a publication in the peer-reviewed literature.

Concerning innovative ensemble generation methods, an ensemble Kalman Filter (EnKF) for assimilation of sea ice concentration has been implemented in both the NEMO3.6 ocean and EC-Earth3.2 climate models. The EnKF presents notable advantages over simpler assimilation techniques, such as nudging, in that it takes into account the covariances between observed and non-observed state variables. This is a critical point, given that in polar regions sea-ice thickness is usually poorly observed while it is in fact the major source of sub-seasonal to seasonal prediction for sea-ice extent. The EnKF has been run successfully in the two model configurations. The results obtained so far should be interpreted with caution, but preliminary results (Figure 4) demonstrate the added value of the filter on the model results. The simulated variability of summer sea-ice extent is generally better simulated when constrained by the observations available. In coupled mode, the monthly assimilation of sea ice concentration is sufficient to prevent the model from drifting away to its own climatology, as it would be in a free run (blue lines).



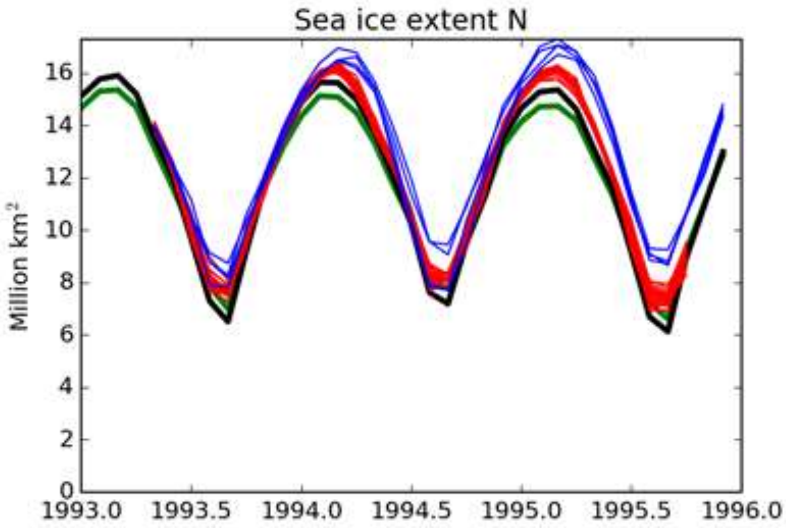


Figure 4: (Top) September Arctic sea ice extent (oceanic area covered by at least 15% of ce) in the ensemble stand-alone data assimilation run ("EnKF", red), in the free run ("No assim.", blue) and in two observational references: the OSI-SAF product (green) that is also the assimilated product, and the National Snow and Ice Data Center (black) estimate, to be considered as more independent. The two insets show the distribution of correlation between time series when either the OSI-SAF (left panel) or NSIDC (right panel) observational reference is used for assessment. (Bottom) Sea ice extent in a 1993-1996 series of integrations of the coupled EC-Earth3.2 model in which the EnKF assimilation was turned on (red, 25 ensembles), off (blue, free run, 5 ensembles) and in the same two observational sources as for the top figure

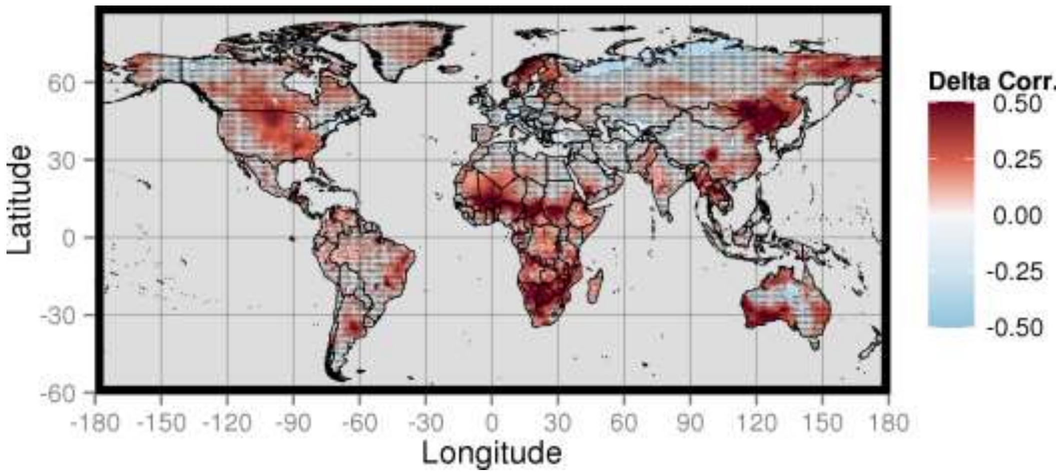
1.3 Forecast system improvement

The improved representation in dynamical models of some key physical processes was expected to increase the quality of climate information for s2d time scales, particularly over land. The factors considered by SPECS include the climate-vegetation processes that have been shown to influence variability and extreme events, the representation of air-sea interaction and upper ocean temperature treatment, relevant because of the persistence and spatial structure of upper ocean temperature, and the explicit inclusion of varying aerosol concentrations (both natural and anthropogenic) and solar irradiance variations in the forecasts. Some of the impacts of these processes are evidenced through more realistic teleconnections, which are required to take advantage of the remote sources of skill over land. Hence, teleconnections have received special attention by the partners in the project. Some of these factors have, in spite of their obvious relevance, been neglected in climate forecasting.

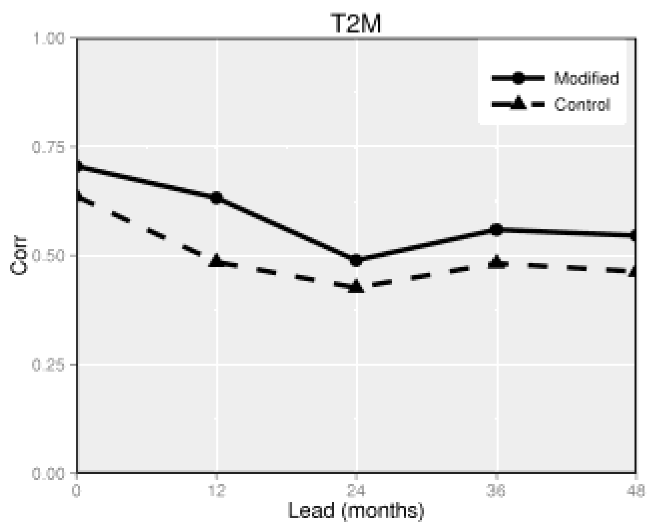
Apart from the use of more realistic physical processes borrowed from model development carried out in contexts other than climate prediction, increased model resolution has been considered as a source of improved models and, potentially, of better climate information. The standard horizontal resolution of most operational forecast systems prior to SPECS has been ~1-2° for both atmosphere and ocean components. At this resolution forecast systems have pervasive systematic errors that affect several aspects of the forecast quality. SPECS has investigated the contribution to s2d forecast quality from an increase in model resolution towards global resolutions close to 0.25°. As in other cases, sets of coordinated multi-model hindcasts were performed to assess the relative merits of resolution using new high-resolution ocean reanalysis datasets. Experiments where the resolution was increased only in the atmosphere or the ocean were carried out to try to isolate the role of the individual model components. This was particularly relevant for the optimisation of computational resources across the various facets available to improve forecast quality. Moving towards resolutions of 0.25° has placed Europe at the forefront of international efforts in this area.

The basis for this work is to improve both reliability and skill of s2d climate predictions in forecast systems. The traditional approach to weather and climate forecasting has been based on deterministic models of the Earth system. However, the inevitable approximations needed to solve the equations of weather and climate in a model are a major source of forecast error and uncertainty on all time scales. Pioneering research has been carried out about how these uncertainties can be accounted for in all the components of climate forecast systems. Stochastic parameterisation has been thoroughly tested as a new promising methodology to address model error and uncertainty.

The more complete representation of the effective vegetation cover variability in EC-Earth2.3 (Alessandri et al., 2016) leads to improved potential predictability at decadal (up to five forecast years) scale. Figure 5a shows a global map of correlation differences for near-surface air temperature for forecast years four and five. The correlation differences that did not pass the significance test at 10% level are dotted. For each grid point, the null hypothesis of getting as high or higher correlation differences simply by chance were tested with a Monte Carlo bootstrap method (1,000 repetitions). The experiment with the modified parameterization displays a significant improvement over Central US Great Plains, Nordeste Brazil, Sahel, Middle East, Southeast Asia including the Indian Peninsula, subtropical southern Africa and Australia, and boreal forests over East Asia, North America and Northeast Europe. Overall, the inclusion of the new parameterization leads to an enhancement of the land averaged correlation at all lead times (Fig. 5b) with a significant (5% significance level) improvement in potential skill at all lead times.



a)



b)

Figure 5: (a) Improved vs. control differences of correlation with the reference historical simulation for two year averages of T2M with 36-months lead time. Areas that did not pass a significance test at the 10% level are dotted. (b) Global-average correlations over land for modified and control experiment. Marks indicate significant differences of the correlations between the two experiments that passed a significance test at the 5% significance level.

Building on the experience with new high-resolution systems, the high-resolution Met Office decadal climate prediction system (DePreSys3) has shown the first skilful predictions of the winter NAO one year ahead (Fig. 6; Dunstone et al., 2016). It represents a key step to providing European winter climate services beyond the seasonal timescale. The mechanisms of this long-range NAO skill rely on two main drivers. The first one is the variability in the tropical Pacific (ENSO), which is skilfully predicted by DePreSys3 out to one year ahead. The second one is the variability in the stratospheric polar vortex, which is also skilfully predicted out to one year ahead and is found to be driven by the low-frequency 11-year solar cycle. The North Atlantic sea surface temperatures and the Arctic sea ice were found not to be driving the second winter NAO skill (despite being implicated as drivers of the observed NAO and first winter predictions) and so are areas identified for possible future improvement as climate models improve and their drift is reduced.

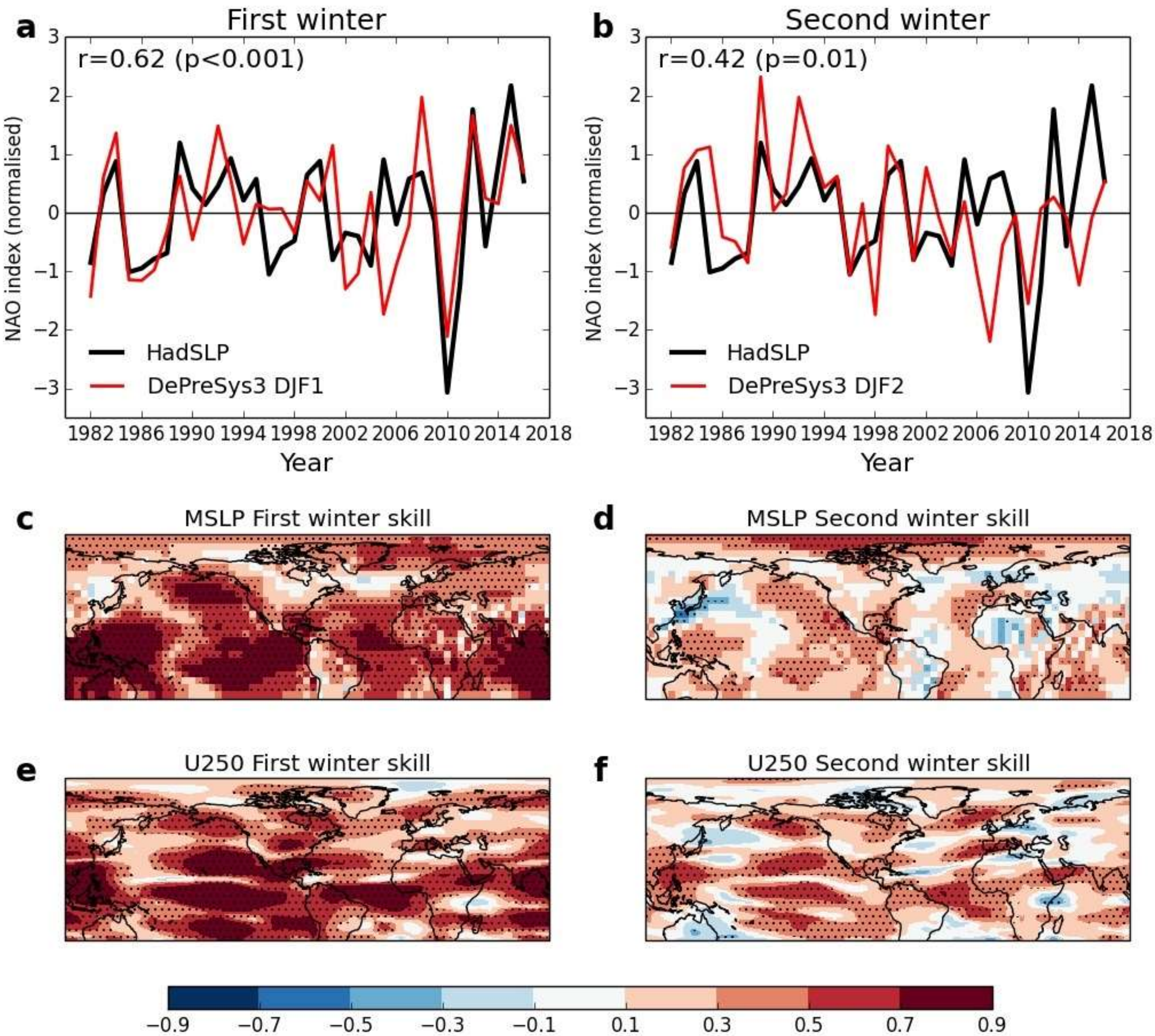


Figure 6: First winter (DJF1) results (a,c,e) and second winter (DJF2) results (b,d,f). a,b, Time series of the NAO from the DePreSys3 ensemble mean (red line) and the HadSLP observations (black line). Time series have been normalized by their standard deviation. The year refers to the January of the DJF period. c,d, Maps of skill (correlation) for predicting the first- and second-winter MSLP. e,f, As c,d, but for predicting the upper-tropospheric (250 hPa) zonal winds. Stippled regions are significant at the 5% level according to a Student's t-test.

The stochastic dynamics technique offers an example of the efforts made to increase the adequacy of the ensemble spread through stochastic techniques. This method consists in introducing random perturbations/corrections of model errors to the atmospheric prognostic fields, such as temperature, vorticity and specific humidity. Model errors are estimated using atmospheric nudging in constraint reforecast runs. Nudging increments are stored each day of the nudged reforecast and make up a correction population from which perturbations are randomly drawn in forecast mode, using cross-validation to remove the year of the forecast. Corrections are consistent in space and between variables as they are drawn by randomly selecting a date from the nudged run, and consistent with forecast time (at a monthly scale) by drawing corrections for the corresponding lead-time. There are two distinct parameters to be defined for the technique: the strength of the relaxation, noted τ, in the nudged re-forecast (which influences the amplitude of corrections) and the random selection method for the perturbations in forecast mode (frequency at which corrections are changed, use of sequences or periodic averages of corrections). In a study presented in Batté and Déqué (2016), perturbations were defined by a 30-day weakly nudged CNRM-CM run (ARPEGE-Climate v6.1 T127L91, NEMO ORCA1L42, GELATO). An analysis in spectral space of the amplitude of the perturbations, averaged over the nudged re-forecast months and 200 hPa layers of the atmosphere shows that differences in perturbation amplitude between upper and lower layers of the atmosphere are more pronounced for specific humidity than for temperature and streamfunction. A quadratic decomposition of the perturbation terms showed that they were mostly made of inter-month variance, justifying their use as a means of increasing ensemble spread. However, limited impact is found in terms of skill. Instead, the technique considerably improves systematic errors in the system, such as the Northern Hemisphere mid- and high-latitudes 500 hPa geopotential height bias, which is substantially reduced (Figure 7). Both S5D and SMM exhibit North Atlantic weather regime statistics closer to ERA-Interim than the REF ensemble, suggesting that not only the model mean state but also its variability is improved. The lack of significant gains in skill over the area could be due to the weak nudging coefficient used for this study, which intended to allow the model to drift closer to its own equilibrium than that of the reference, but could imply that the perturbation population does not properly sample model errors at the seasonal range. To test this hypothesis, several experiments with a more recent version of the CNRM-CM coupled seasonal forecasting system were run, using different values for τ. More details are provided in the deliverable D44.1.

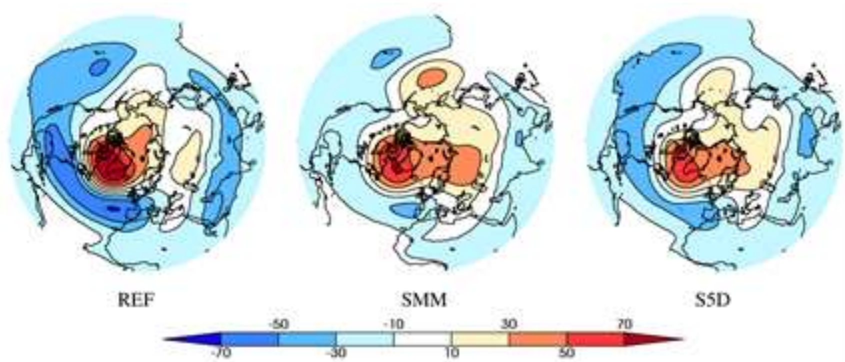


Figure 7: Bias with respect to ERA-Interim for 500 hPa geopotential height in DJF over 1979-2012 in the REF, SMM and S5D ensemble experiments. Figure 6 in Batté and Déqué (2016).

1.4 Regionalisation, calibration and combination

Policymakers require quantitative regional s2d climate predictions to aid efforts in preparing for unusual climate variations or for adapting to a changing climate. Process-based (dynamical) s2d predictions have many advantages, but their production is complex, challenging, computationally expensive and prone to biases. The skill of current dynamical forecast systems is limited, particularly over Europe. It therefore seemed prudent to compare the forecast quality of dynamical predictions to that of independent forecast systems based on simpler, statistically-based, physically-motivated alternatives. This type of assessment has been useful in predictions for many years, particularly for ENSO forecasts. SPECS built a prototype empirical prediction system for land climate variables for s2d time scales. The empirical predictions, which are now produced regularly, have been used both as a benchmark with which to assess and compare the dynamical forecast systems and as an independent forecast system to be combined with the other prediction sources. The former use also allowed identifying processes that are at the source of the skill and diagnosing dynamical forecast system biases. It is important to note that the empirical predictions are probabilistic in nature.

Global forecast systems are in many cases unable to provide actionable information at the spatial scale required by the stakeholders. This is a valid hypothesis even with the increases in horizontal resolution undertaken in the project. Hence, regionalisation/downscaling methods were developed and tested in SPECS. The goal was to provide suitable products for a range of applications both in SPECS and in other European initiatives. Although there are both empirical/statistical and dynamical (based on regional climate models) approaches to downscaling climate information, the effort mainly focused on statistical methods due to the enormous amount of hindcasts to be downscaled for each specific case. This large number of hindcasts is due to the need to estimate the forecast quality of a system to be provided along with a prediction and is the result of the large ensembles with many start dates typical of an s2d hindcast sample. Statistical downscaling is based on empirical relationships derived between a set of predictands based on observed quantities and either reanalysis (in the perfect prognosis approach) or predictions (model output statistics, MOS, approach). The perfect prognosis approach, which is popular in climate change studies, is sensitive to the typical systematic biases in the predictor field. MOS, instead, is mainly popular in short- and medium-range weather prediction, requires long hindcasts but accounts for the systematic biases in the predictions. While these methods have been widely explored, the availability of appropriate and well-documented user-friendly tools/services to offer s2d predictions at a local scale was not systematically undertaken before SPECS. Besides, their relative merits not been properly described in an s2d prediction context. A key activity consisted in the development and documentation of a new open-source R package for s2d statistical downscaling including the state-of-the-art techniques and suitable for direct application by the stakeholders. In spite of the above, statistical and dynamical downscaling methodologies have also been compared using seasonal predictions and multiple regional models in two experiments over Europe and South America, respectively.

Climate predictions produced with either physically-based, or dynamical, and empirical/statistical forecast systems pose a challenge to a large number of users. These predictions usually come from a plethora of sources. Given the availability of all these imperfect forecast systems, there is also the need for developing novel and robust methods for combining and calibrating the predictions. This has been achieved by extracting the best information from each forecast system based on their past (retrospective) performance. The goal has been to ensure that reliable probabilistic predictions are produced by integrating the information from all available sources for users to have a consolidated source of climate information. The final combined and calibrated forecasts represented a fundamental contribution to the development of a climate service because all climate-sensitive sectors are extremely sensitive to the lack of reliability of the predictions. For this reason, a special effort has been made to transfer this information to the users involved in SPECS.

A new global empirical decadal prediction system has been developed and tested (Suckling et al., 2016). This system is part of a pair of systems targeting seasonal and decadal time scales. The model, based on a multiple linear regression approach consistent with the approach has been designed to produce probabilistic predictions. It uses observed and projected global forcings based on well-understood physical relationships, as well as large-scale SST patterns, as predictors for local (grid-scale) annual mean surface air temperature and precipitation over the globe. The system incorporates uncertainty information through the generation of ensembles, which are output in a similar format to those of dynamical models in order to aid robust comparisons. The selection of forcings and predictors is based on physical principles and well-understood observed relationships to the fullest extent, yet is as simple as possible, using as few predictors as necessary to minimise the risk of overfitting. The present study has focused on prediction of surface air temperature and precipitation using global mean radiative forcings and observed ENSO, AMV and interdecadal Pacific Oscillation (IPO) indices since there is a relative abundance of data with which to build and evaluate the system. The model performance has been evaluated over a set of historical hindcast experiments, in terms of both their deterministic and probabilistic skill.

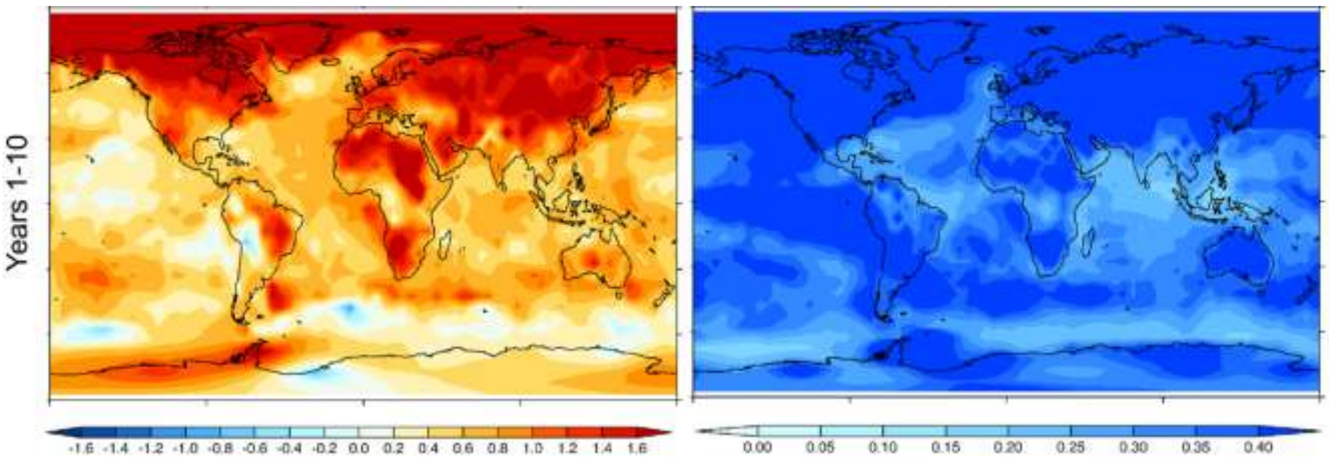


Figure 8: Empirical predictions of forecast years 1-10 for surface air temperature anomalies (relative to 1961-1990) for the period 2016-2025. The left-hand panels show the ensemble mean from 51 ensemble members and the right-hand panels show the standard deviation across the ensemble. The largest warming is predicted to be over land, particularly in the Northern Hemisphere, although these regions are also associated with the largest ensemble spread. Only a few regions, such as in the North Atlantic and in the Southern Ocean are predicted to cool compared to the 1961-1990 baseline. Figure 11 from Sucking et al. (2016).

An example of the empirical downscaling exercise, which has also been compared to the outcome of dynamical downscaling simulations, appears in Figure 9. This example is for the Northeast of Brazil, an interesting case study because of the high skill of the seasonal predictions in this region. The different statistical downscaling approaches developed in SPECS (more details in D52.2 “R-based package released for statistical downscaling s2d forecast, including documentation and capability for evaluation and estimation of confidence intervals”) were tested in this region, including MOS (Bias Correction and Multiple Linear Regression) and perfect prognosis methods (Generalized Linear Models and analogs). The use of downscaling improves seasonal forecast quality over the raw model output beyond improving the systematic bias (which is done by construction by all statistical downscaling methods). The perfect prognosis methods exhibit the best performance showing an improvement (10% on average) of the skill. As expected, the mean bias correction exhibits no added value (beyond reducing the systematic bias). The MOS deteriorates the skill of the global model in some regions (approximately in 50% of the territory). However, it is worth mentioning that the observation uncertainty could be a relevant factor in this case study (e.g. poor spatial coverage, different record lengths may also introduce inhomogeneity), which has not yet been taken into account for forecast validation. The joint results of these experiments together with the dynamical downscaling simulations over this region have been described in D52.4 “Scientific paper on the added value of the dynamical downscaling”.

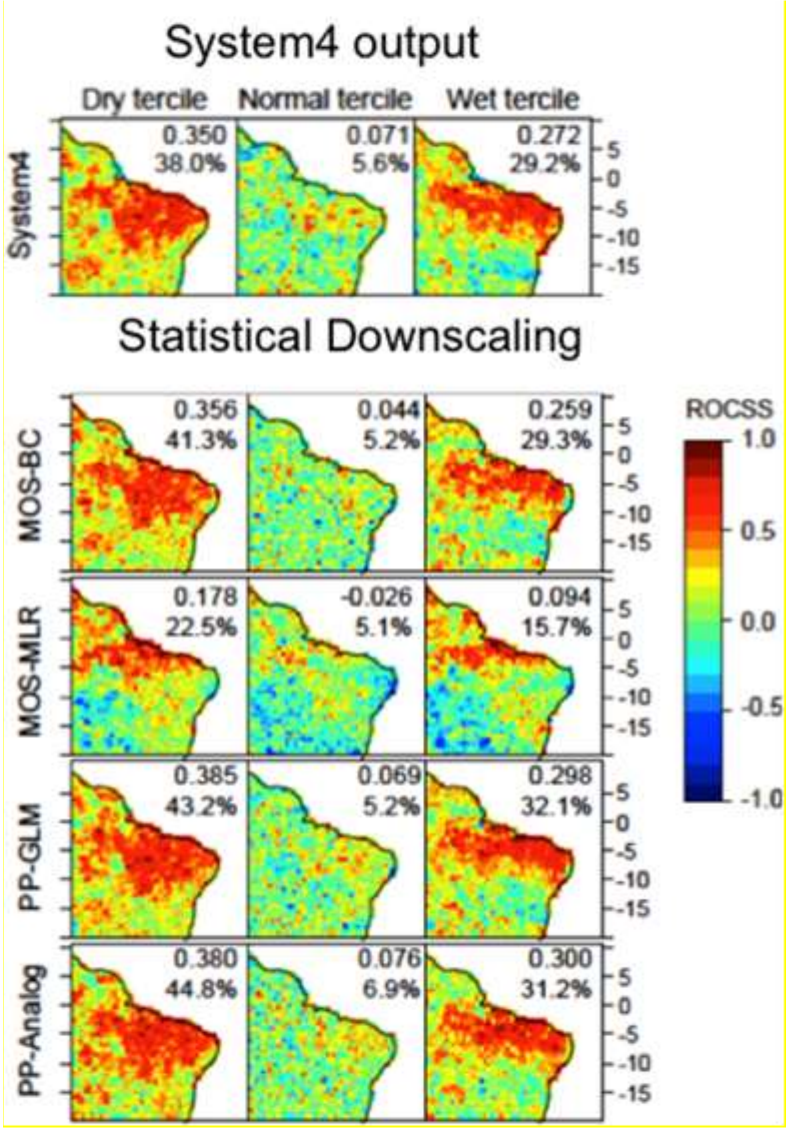


Figure 9: Relative operating characteristics skill score (ROCSS) of the different statistical downscaled predictions over Northeast Brazil. Global model output from ECMWF's System 4, empirical bias correction (MOS-BC), regression MOS (MOS-MLR) and Perfect Prog. Generalized Linear Model (GLM) and Analog techniques (in rows) are shown for the dry (left), normal (centre) and wet (right) tercile categories of precipitation.

Fundamental research on the statistical post-processing of multi-model forecasts has been done in SPECS. In decision-critical contexts, a number of s2d forecast systems produced by different institutions are usually available to a decision maker. An old, but still very relevant question in this context is how to combine all available predictions from different systems into a single forecast. An appropriate statistical post-processing method should correct systematic biases that might be different between different models. A method to combine different models should also be able to take into account that different models have different levels of skill at predicting the forecast target. As an example, seasonal multi-model predictions of North-Atlantic Oscillation (NAO) and Nino3.4 index have been used. The analysis revealed that different forecast systems show different levels of predictive skill measured by the sample correlation coefficient between the ensemble mean and the corresponding observations. Sample correlations for Niño3.4 and NAO from a range of seasonal forecasts from the North American multi-model and the DEMETER project are shown in Tables 1 and 2.

CFS	CMC	GFDL	MF	NASA	EC
0.81	0.89	0.85	0.87	0.88	0.91

Table 1: Niño3.4 correlation skill for the 1982-2010 hindcast period of six global forecast systems. Hindcasts initialised in August valid for December (five-month lead).

EC	LODYN	MF	MPI	UKMO
-0.07	0.03	0.18	0.019	-0.14

Table 2: NAO correlation skill for the 1975-2002 hindcast period of 5 coupled ocean-atmosphere models (data from the DEMETER project). Forecasts initialised in November valid for DJF (1-month lead)

The different levels of skill suggest that different models should be weighted differently when combined into a single prediction. Two different combination schemes were explored. The first method uses the same (equal) weights for all models to calculate the final prediction. This method is equivalent to linear regression of the observation on the multi-model ensemble mean, which is a simple and widely-used method. The second method constructs the final prediction by weighting the different model forecasts unequally, using multiple linear regression of the observation on the different model runs. This method for combining multiple pieces of model information into a single prediction is also a type of MOS. These two weighting schemes were compared in a cross-validation study, where the model weights were estimated on a small number on training data, and the resulting predictions by model combination were evaluated on test data that was not part of the training data. Table 3 shows the cross-validated logarithmic scores (for which lower scores indicate better skill) for the equal and unequal weighted Niño3.4 and NAO predictions.

LOG SCORES	Nino3.4	NAO
equal weighting	1.80	1.89
unequal weighting	1.97	2.17

Table 3: Logarithmic scores for equal and unequal weighting of Niño3.4 and NAO predictions.

The main result in these examples is that equal weighting performs generally better than unequal weighting. The difference is small, but coherent across both investigated data sets. To understand these results observations and different model runs were simulated from multivariate Normal distributions that generate forecasts that have different levels of skill. The differences in forecast skill can be adjusted by varying the entries in the correlation matrix of the Normal distribution. Based on theoretical arguments it can be shown that for such forecasts an unequal weighting strategy is optimal. The main problem that can make the unequal weighting strategy suboptimal is that the weights are unknown parameters of a statistical model that have to be estimated from a finite number of training data. Random parameter estimation errors degrade the quality of forecasts, and the more parameters that have to be estimated, the higher the resulting forecast error gets. The dependence of forecast error on parameter estimation error explains why the equal weighting strategy produces better forecasts than the unequal weighting strategy, despite the fact that the artificially generated forecasts have different levels of skill, which calls for an unequal weighting strategy.

Figure 10 shows two examples of how forecast skill of combined multi-model ensembles with equal and unequal weights depends on the number of training data available to estimate the weights. These examples illustrate that for hindcast sizes smaller than 50 years the unequal weighting scheme is unlikely to improve upon the equal weighting scheme. For larger sample sizes, the weights can be estimated more robustly, and the unequal weighting scheme can better unfold its power to produce more skilful forecasts. Furthermore, if all the individual models have low correlation skill (as in the NAO example) and if the hindcast sample size is small, it has been found that even the climatological mean forecast (where all models have weight zero) can be preferable to the unequal weighting scheme.

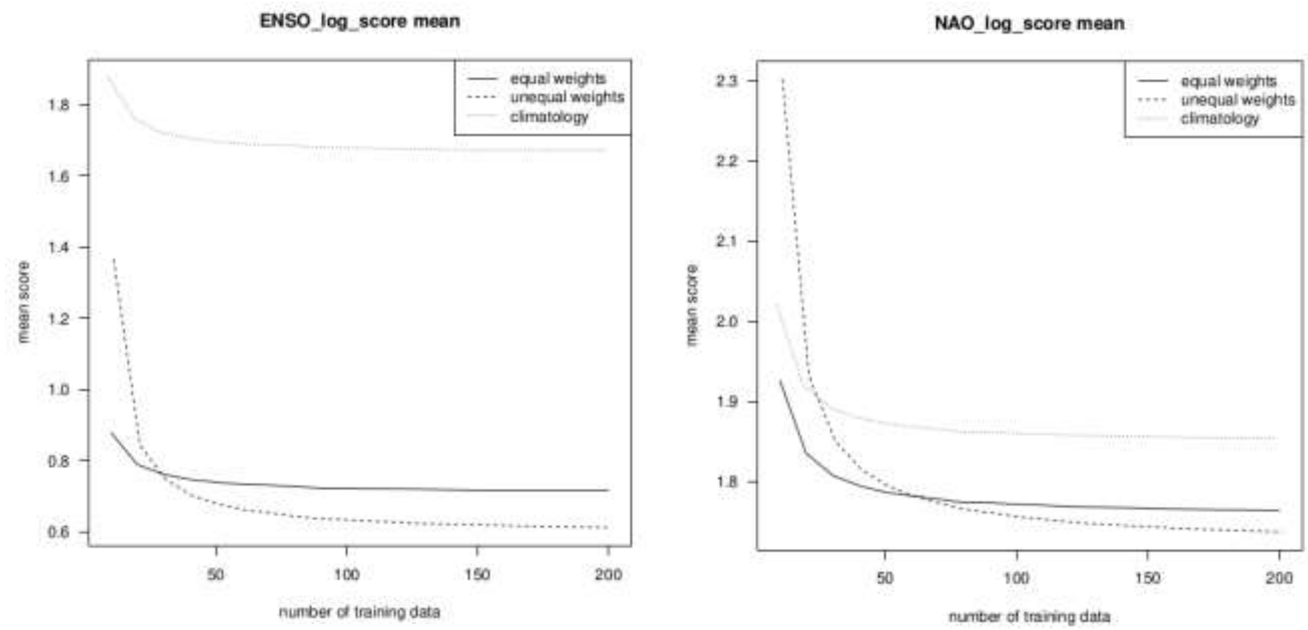


Figure 10: Forecast skill (logarithmic score; lower is better) for simulated data similar to the ENSO and NAO multi-model ensembles, as a function of the sample size available for estimating the post-processing weights.

1.5 Applications and international relevance

The variability of climate at s2d time scales is one of the factors that determine the evolution of many climate-sensitive sectors. The identification of the specific climate information (time scales, frequency of update, minimum spatial scale, variables, etc.) required by these sectors and their sensitivity to the intrinsic uncertainty of this climate information set the targets for the improvement of operational climate forecast systems. SPECS assessed the impact of the forecast system improvements obtained across the project on two impact sectors: agriculture and renewable energy generation. These are two sectors with a long history of collaboration with climate forecast developers. Besides, these sectors adapt to climate change on an interannual time scale instead of on the basis of very long-term climate change projections. Impact model simulations have been performed, sometimes using the downscaled climate predictions from already existing forecast datasets and improvements of forecast quality been evaluated taking into account the full chain of uncertainties in collaboration with relevant stakeholders. The interaction in this point with the EUPORIAS project has been particularly important.

SPECS has also paid attention to ensuring a coordinated European response to international climate prediction initiatives, taking the lead whenever that was possible. It recognised the central role of the Global Producing Centres (GPCs) and Regional Climate Centres (RCCs) in the creation of efficient, operational climate services and as an intermediary with private-sector stakeholders. Efforts were made to improve the visualisation of the predictions compared to what was available at the beginning of the project, particularly for Regional Climate Outlook Fora (RCOF) outlets. Identifying substantial gaps in the climate prediction field, the partners undertook the challenge of documenting the public dissemination of the climate information generated (for instance, setting standards that have been taken up internationally) and elaborate introductory, general-public information on the methodologies used to generate the predictions and the regions, variables and times of the year where useful forecasts can be expected, always responding to the feedback gathered from either affiliated partners, stakeholders and EUPORIAS. Central objectives have been the demonstration of usefulness of s2d forecasts and transfer the knowledge to the relevant operational platforms.

Partners have worked with different impact sectors, but wind energy has been paid a special attention. Figure 11 compares the seasonal predictions of 10-metre wind speed produced in two experiments, one with high resolution and another with low resolution. These experiments have been performed with the EC-Earth forecast system (version 3.0.1) initialised the first of November in the period 1993-2009. This assessment is interesting for the wind energy sector because sometimes the users are interested in high resolution but depending on the region, the prediction with lower resolution could provide better results. This kind of interaction with stakeholders are used as opportunities to describe the scientific results of climate predictions and gather user's feedback about relevant technical aspects as well as about the actual impact that forecast system improvements have for the users.

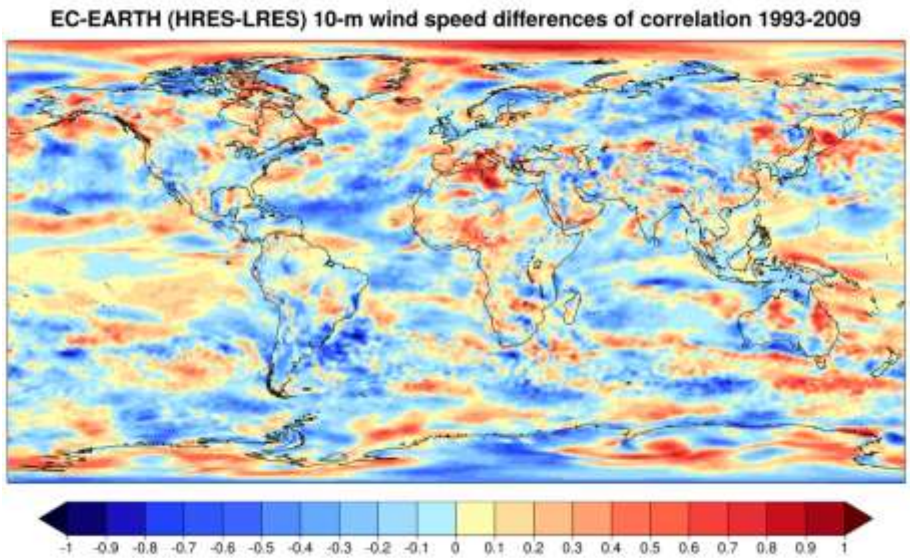


Figure 11: Difference in the correlation of the wind speed predictions from two sets of experiments with the EC-EARTH3.0.1 forecast system at two different horizontal resolutions (around 40 km and 70 km) initialised in the first of November over 1993-2009.

The exchange of decadal predictions, which started at the beginning of the project, has successfully been achieved again in 2016, with around ten prediction systems providing real-time forecasts out to five forecast years (Figure 12). An important advance in the decadal forecast exchange has been the expansion to other variables including precipitation and ocean diagnostics such as the Atlantic meridional overturning circulation. There is a notable and clear consensus across models for a decline in the Atlantic overturning over the coming years. Full details are available from the web page of this initiative led by SPECS.

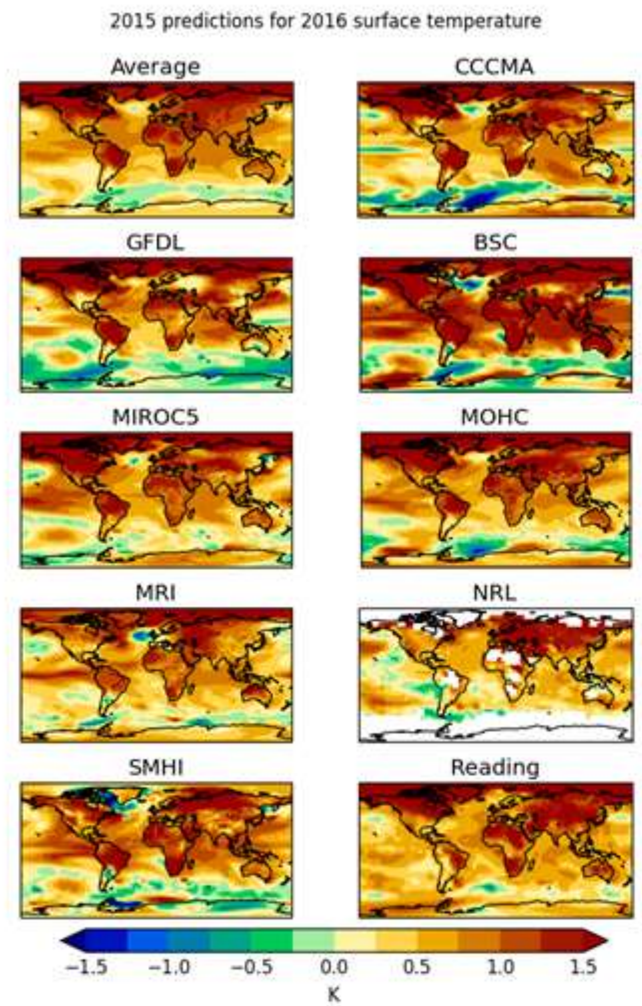


Figure 12: Decadal forecast information available for 2016 onwards on the SPECS funded decadal forecast exchange.

1.6 References

- Alessandri, A., F. Catalano, M. De Felice, B. van den Hurk, F.J. Doblas Reyes, S. Boussetta, G. Balsamo and P.A. Miller (2016). Multi-scale enhancement of climate prediction over land by increasing the model sensitivity to vegetation variability in EC-Earth. Climate Dyn., doi:10.1007/s00382-016-3372-4.
- Batté L. and M. Déqué (2016). Randomly correcting model errors in the ARPEGE-Climate v6.1 component of CNRM-CM: applications for seasonal forecasts. Geosci. Model Dev., 9, 2055-2076.
- Dunstone, N., D. Smith, A.A. Scaife, L. Hermanson, R. Eade, N. Robinson, M. Andrews and J. Knight (2016). Skilful predictions of the winter North Atlantic Oscillation one year ahead. Nature Geosci., 9, 809-814, doi:10.1038/ngeo2824.
- Massonnet, F., O. Bellprat, V. Guemas, and F.J. Doblas-Reyes (2016). Using climate models to estimate the quality of global observational data sets. Science, 354, 452-455, doi: 10.1126/science.aaf6369.

- Suckling, E.B., G.J. van Oldenborgh, J.M. Eden and E. Hawkins (2016). An empirical model for probabilistic decadal prediction: global attribution and regional hindcasts. Climate Dyn., doi:10.1007/s00382-016-3255-8.

2 Potential impact (including the socio-economic impact and the wider societal implications of the project so far) and the main dissemination activities and the exploitation of results

2.1 Dissemination

The dissemination has focused on the following activities:

- **Scientific events and workshops**: The results of the project have been shown in several of the most important conferences and workshops on climate variability and change. These activities can be consulted in the annex *A2 List of dissemination activities.*
- **Scientific publications, peer reviewed**: This project has been characterized by the large amount of scientific publications. The results of the deliverables submitted have become a good source for these publications. Whenever it was possible, the consortium has followed the principle of Open Access recommend by the EC. The list can be consulted in the annex *A1 List of publications*.
- **SPECS technical notes**: The research issues of the project that do not fit into deliverables or publications have been also distributed among the scientific community through these technical notes. During the last period around seven technical notes have been published. They are available in the SPECS website and also included in the annex A2. The SPECS scientists consider these notes a very interesting tool to raise scientific discussions and share them with the scientific community.
- **Project website**: It has become a very important dissemination tool since the twitter account of the project has been created. This twitter account is very active and connects scientists from different projects also linked to SPECS. The website is continuously updated to better adapt to the dissemination needs of the consortium. The data produced by the SPECS partners is formatted in a unified way. The project website also includes all the information related to publications and dissemination activities.
- **SPECS fact sheets**: They are provided by the scientific teams in the consortium and are focused on outstanding topics of interest to stakeholders and the progress of the project. The SPECS fact sheets are available in the project website. From the time being seven have been published within work package 6.2.

Data dissemination strategy

At the end of the project, all the data planned at the beginning of the project have been uploaded, quality controlled, archived on the STFC Jasmin server and published through the ESGF nodes of BADC and SMHI.

Around 400,000 files are available on the ESGF for the 9 experiments defined in the DoW and from the 10 partners involved. The total amount of data that has been made available is of more than 130TB. This is much more than what was initially planned.

The ESGF nodes that were down during the last reporting period have now been reinstalled and their security has been improved to be able to share the data from SPECS and other projects. The parallel access solutions installed during the downtime (ftp and OPeNDAP servers and web browser) have been maintained and will still be active after the official end of SPECS, offering a multi-faceted access to the SPECS data.

As described in the DoW, the data present in the CEDA archive will be kept for “at least five years after the end of the project”. In practice, CEDA rarely deletes old project data although it might be migrated to tape if the usage is very low. It is the same thing for the access through ESGF: as long as CEDA exists and is participating in ESGF, the data will be presented there.

In terms of the SPECS legacy, in addition to the strict data management part, it is worth noting that some software improvements have been made to be able to cope with the SPECS requirements. We can cite CDO (climate data operators, developed by DKRZ and one of the most used data transformation software in the climate community) that has been modified to be able to handle the SPECS double time axis and THREDDS (Unidata's Thematic Real-time Environmental Distributed Data Services) that has been improved to be able to cope with the very high number of directory levels to store the data generated by the inherent complexity of the seasonal-to-decadal data.

Finally, the SPECS metadata convention has been adopted by the World Climate Research Programme for the dissemination of its repository of climate predictions.

2.2 Impact

The project has achieved an unprecedented level of results published in the peer-review literature. At the same time, the attendance to conferences remained at a high level, as reported in the last periodic report, with sessions organised in main events (EGU) and workshops organised in collaboration with other initiatives, notably WMO. As SPECS is a climate-services research project, the links to the users have not been neglected. Participation in professional conferences was also substantial as this proved to be the best way (instead of individual meetings or general-purpose user workshops) to reach out a sufficient number of stakeholders. This has been complemented with a successful series of factsheets available from the project web site, that are used as an example by other projects (e.g. PRIMAVERA). SPECS has risen interest among those responsible of the Copernicus Climate Change Services. As a result, several partners have provided detailed information about the project structure and outcomes that has been used to design some of the calls published. The SPECS partners closely collaborated with those in the EUPORIAS project, gathering an unprecedented set of end users and stakeholders. EUPORIAS provides feedback on the user needs and interests, mainly through its prototypes, which are taken into account and addressed (for instance, providing model results and examples from the SPECS datasets) by the SPECS partners. This relationship was made obvious during the final general assembly and the conference on climate modelling, observations and services that followed. Finally, SPECS kept influencing international activities (mainly under the WMO umbrella) such as the MedCOF, the design of components of the CMIP6 experiment, the link to the IS-ENES Climate4impact portal, or the multi-model decadal forecast exchange.

SPECS influenced international activities (mainly under the WMO umbrella) such as the MedCOF (and other RCOFs), the design of components of the CMIP6 experiment, as an inspiration of the activities of the WCRP Working Group on Subseasonal-to-Interdecadal Prediction (WGSIP), the link to the IS-ENES Climate4impact portal, or the multi-model decadal forecast exchange. It has been central in the ECOMS activities, for instance editing the ECOMS newsletter, but also participating in the creation of the H2020 ECOMS2 project. Most importantly, many SPECS activities have contributed to the further development of the European operational climate prediction systems such as EUROSIP, DePreSys and MiKlip, not only from the modelling point of view, but also (and this is probably the most relevant) from aspects linked to ways to make a more efficient and traceable use of the forecasts by a wider range of users.