

This discussion paper is/has been under review for the journal Geoscientific Model Development (GMD). Please refer to the corresponding final paper in GMD if available.

A global empirical system for probabilistic seasonal climate prediction

J. M. Eden¹, G. J. van Oldenborgh¹, E. Hawkins², and E. B. Suckling²

¹Royal Netherlands Meteorological Institute (KNMI), De Bilt, the Netherlands

²NCAS-Climate, Department of Meteorology, University of Reading, Reading, UK

Received: 2 April 2015 – Accepted: 21 April 2015 – Published: 26 May 2015

Correspondence to: J. M. Eden (jonathan.eden@knmi.nl)

Published by Copernicus Publications on behalf of the European Geosciences Union.

3941

Abstract

Preparing for episodes with risks of anomalous weather a month to a year ahead is an important challenge for governments, NGOs and companies and relies on the availability of reliable forecasts. The majority of operational seasonal forecasts are made using process-based dynamical models, which are complex, computationally challenging and prone to biases. Empirical forecast approaches built on statistical models to represent physical processes offer an alternative to dynamical systems and can provide either a benchmark for comparison or independent supplementary forecasts. Here, we present a simple empirical system based on multiple linear regression for producing probabilistic forecasts of seasonal surface air temperature and precipitation across the globe. The global CO₂-equivalent concentration is taken as the primary predictor; subsequent predictors, including large-scale modes of variability in the climate system and local-scale information, are selected on the basis of their physical relationship with the predictand. The focus given to the climate change signal as a source of skill and the probabilistic nature of the forecasts produced constitute a novel approach to global empirical prediction.

Hindcasts for the period 1961–2013 are validated using correlation and skill scores. Good skill is found in many regions, particularly for surface air temperature and most notably in much of Europe during the spring and summer seasons. For precipitation, skill is generally limited to regions with known ENSO teleconnections. The system is used in a quasi-operational framework to generate empirical seasonal forecasts on a monthly basis.

1 Introduction

The provision of reliable seasonal forecasts is an important area in climate science and understanding the limitations and quantifying uncertainty remains a key challenge (Doblas-Reyes et al., 2013; Weisheimer and Palmer, 2014). Operational seasonal fore-

3942

casting, although once limited to a handful of research centres, is now a regular activity across the globe. Much recent focus has been given to the skill and reliability of seasonal climate predictions. Dynamical (process-based) forecast systems are arguably the most important tool in producing predictions of seasonal climate at continental and regional scales. Such systems are based on numerical models that represent dynamical processes in the atmospheric, ocean and land surface in addition to the linear and non-linear interactions between them. However, the development of dynamical systems is a continuous challenge; climate models are inherently complex and computationally demanding and often contain considerable systematic errors that limit model skill in particular regions and seasons.

As an alternative to dynamical forecast systems, empirical approaches aim to describe a known physical relationship between regional-scale anomalies in a target variable (the predictand), say, temperature or precipitation, and preceding climate phenomena (the predictors). In its simplest form, an empirical forecast may be based on persistence in which observations of the variable to be predicted are taken as the forecast for some lead time. Such forecasts have frequently performed better at short lead times than those simply prescribed by the long-term climatology, particularly so in the Tropics. More sophisticated statistical methods include analog forecasting (van den Dool, 2007; Suckling and Smith, 2013) and regression-based techniques, which may in turn take predictive information from spatial patterns using, for instance, empirical orthogonal functions (EOFs) (e.g. van Oldenborgh et al., 2005b), maximum covariance analysis (MCA) (e.g. Coelho et al., 2006) and linear inverse modelling (LIM) (e.g. Penland and Matrosova, 1998). Empirical predictions for the phase and strength of the El Niño Southern Oscillation (ENSO) have historically shown comparable skill to those produced by dynamical systems (e.g. Sardeshmukh et al., 2000; Peng et al., 2000; van Oldenborgh et al., 2005b). Additionally, an inherent advantage of empirical methods is the ease with which knowledge of climate variability gained from analysis of up-to-date observations can be incorporated into a prediction system (Doblas-Reyes et al.,

3943

2013), which in turn facilitates the development of new methodologies and statistical techniques (van den Dool, 2007).

Empirical forecasts serve both as a baseline for dynamical models and can be used to improve the forecasts by limiting the effects of model biases. However, differences in the development and output of dynamical and empirical-statistical approaches makes systematic comparison troublesome, and understanding the relative skill of each forecast type is challenging. Recent attempts have been made in developing empirical benchmark systems for multiple variables, such as land and sea surface temperature, on decadal time scales (e.g. Ho et al., 2013; Newman, 2013), concluding that the usefulness of such systems merits further development. While comparison of dynamical and empirical systems for seasonal forecasts is not novel, a systematic global comparison for multiple variables, including probabilistic measures, has been lacking. A key potential benefit of such comparison is the identification of regions where empirical models are skilful and may be able to provide useful forecast information to complement the output of dynamical systems. Supplementing dynamical forecasts with empirical forecasts is of great importance in situations where dynamical systems are known to have weaknesses. It has also been shown that combining the output of empirical and dynamical systems can produce marked improvement over single-system forecasts (e.g. Coelho et al., 2006; Schepen et al., 2012).

A fundamental criticism of empirical systems is the question of their applicability in a future, perturbed climate. In other words, to what extent will the predictor-predictand relationships underpinning a statistical model remain stationary under climate change? Sterl et al. (2007) noted that within the statistical uncertainties, no changes could be detected in ENSO teleconnections. Doblas-Reyes et al. (2013) recently noted that the temporal evolution of seasonal climate should be considered as forced not only by the internal variability of the climate system but also by changes in concentrations of greenhouse gas and aerosols as a result of anthropogenic activities. Such external forcings are considered in climate change simulations, and also to an increasing extent in the field of decadal prediction (e.g. Krueger and Von Storch, 2011). Current seasonal

3944

forecast systems now include these forcings (Doblas-Reyes et al., 2006; Liniger et al., 2007), but the resulting trends are sometimes not realistic.

Here we present and validate a simple empirical system for predicting seasonal climate across the globe. The prediction system, based on multiple linear regression, produces probabilistic forecasts for temperature and precipitation using a number of predictors based on well-understood physical relationships. In all forecasts, the global equivalent CO₂ concentration is used as the primary predictor as an indicator of the climate change signal. Additional predictors describing large-scale modes of variability in the climate system, starting with ENSO, and local-scale information are subsequently selected on the basis of their potential to provide additional predictive power. The system presented will have two purposes: (a) to serve as a benchmark for assessing and comparing the skill of dynamical forecast systems; and (b) to act as an independent forecast system in combination with predictions from dynamical systems. Key to achieving these goals will be the system's implementation in a quasi-operational framework with empirical forecasts made on a monthly basis and the availability of a set of hindcasts.

The method implemented here constitutes a relatively simple approach to empirical forecasting. The global and automated nature of the prediction system calls for the underlying empirical method to be parsimonious in terms of the predictive sources used to construct it. The statistical model and the selection of predictors will thus be based on physical principles and processes to the fullest extent so as to elicit the maximum predictive power of, first of all, the long-term trend associated with the climate change signal and, secondly, as few additional predictors as is necessary in order to minimise the risk of overfitting. The final system will also be sufficiently flexible to facilitate its future development. Such development may involve inclusion of additional predictors should more complete and reliable datasets become available, or the application of the system to alternative predictands including those relating to the magnitude and frequency of extreme events.

3945

Producing empirical forecast output in similar format to dynamical systems is crucial when designing a framework for robust comparison. A weakness of current dynamical-empirical system comparison is the general lack of a common set of validation measures. Whereas dynamical systems inherently provide output in the form of ensemble forecasts, which may be validated in probabilistic terms, validation of empirical systems does not always extend beyond deterministic measures, such as bias, RMS error and correlation (Mason and Mimmack, 2002). Here, the uncertainties are explicitly parametrised as an ensemble of forecasts and we employ a rigorous validation framework designed to assess both the deterministic and probabilistic aspects of the forecast system.

The remainder of the paper is structured as follows. Section 2 describes the prediction system in full, including the observational data used for empirical model fitting and validation. An analysis of the potential usefulness of the predictors is given in Sect. 3. The skill of the prediction system is then assessed in Sect. 4 with a discussion and outlook given in Sect. 5.

2 Model fitting and validation measures

Key to achieving the goals set out in Sect. 1 is the development of an automated forecast system that can be applied globally and, in principle, for any number of predictands. For these reasons, the regression-based prediction system developed here is relatively simple in comparison with more sophisticated statistical models, with emphasis given to a basis of physical processes and the avoidance of overfitting.

Our system incorporates a two-step regression approach for estimating seasonal (three-month) surface air temperature (SAT) and precipitation (PREC) as a function of global and local atmospheric and oceanic fields. The first step is to utilise the predictive information in externally forced low frequency variability associated with anthropogenic activity; this is represented by CO₂-equivalent concentration (CO₂EQV), based on historical estimates until 2005 and according to Representative Concentration Pathway

3946

(RCP) 4.5 thereafter, which constitutes the net forcing of greenhouse gases, aerosols and other anthropogenic emissions (Meinshausen et al., 2011). The second step is to identify predictive potential of a set of predictors representing internal variability following removal of the trend associated with external forcing. The final product is a season- and location-specific multiple linear regression using the selected predictors, with an independent regression model calibrated at each grid point. Analysing the degree of additional predictive skill offered by each predictor will form an important precursor to the implementation of the system.

As additional predictors, we consider first of all variables that describe large-scale modes of variability. ENSO is the most important of these in terms of its contribution to the skill of seasonal predictions, particularly in the tropics (van Oldenborgh et al., 2005a; Balmaseda and Anderson, 2009; Weisheimer et al., 2009; Doblas-Reyes et al., 2013). Circulation and precipitation patterns in the tropical Pacific associated with ENSO SST anomalies are subsequently linked to climate variability in other parts of the globe (Alexander et al., 2002). In addition, modes of variability in other tropical oceans, including the tropical Atlantic and Indian basins, are known to contribute substantially to variability in SAT and PREC, particularly in surrounding regions (Doblas-Reyes et al., 2013). Many such phenomena are linked in some way to ENSO, although variability in the Indian Ocean Dipole (IOD) is known to occur independently (Zhao and Hendon, 2009). Similarly, the Pacific Decadal Oscillation (PDO), defined as the leading empirical orthogonal function (EOF) of North Pacific monthly SST anomalies, is considered as a representation of variability on interdecadal time scales that is not otherwise apparent in interannual ENSO variability (Liu and Alexander, 2007). Drought occurrence in the US is known to be linked to the phase of both PDO and the Atlantic Multidecadal Oscillation (AMO). Atmospheric anomalies, including troposphere–stratosphere interactions, are also known to have predictive potential. The Quasi-Biennial Oscillation (QBO) (Ebdon and Verryard, 1961; Baldwin et al., 2001) has recently been considered in a multiple regression model for predicting European winter climate (Folland et al., 2012). With this in mind, the following indices are considered as predictors: NINO3.4

3947

(representative of ENSO), PDO, AMO, IOD and QBO. The system is designed to be flexible enough for the inclusion of additional predictors in the future.

External forcing and global modes of variability are not the only source of skill in seasonal forecasts. Many studies, including those based on dynamical systems, have found links between local climate and variations in preceding nearby climate phenomena (e.g. van den Hurk et al., 2012; Quesada et al., 2012). The most simple of these is persistence; that is, the value of the predictand (either SAT or PREC) for the same location at some lead time. Here, we seek to elicit predictive information from persistence (PERS) and other variables that vary from grid point to grid point in addition to the set of large-scale modes of variability described above. For coastal locations in particular, we seek to maximise the potential of short-term memory contained within neighbouring sea surface temperatures to provide greater predictability than PERS at the specified lead time. We derive a local sea surface temperature (LSST) index for each predictand grid cell, defined as the mean of the k nearest grid cells containing SST information. Here, $k = 5$ throughout the analysis although this value could of course be altered or optimised for region-specific analysis. Additionally, soil moisture has been shown to impact on local temperature (e.g. van den Hurk et al., 2012). As a proxy, we consider as a predictor the accumulated rainfall (CPREC) during the predictor period.

Global observational datasets provide the predictand (SAT and PREC) and predictor fields required for model calibration and validation. SAT is taken from the Cowtan and Way (2014) reconstruction of the Hadley Centre–Climatic Research Unit Version 4 (HadCRUT4) (Morice et al., 2012), which uses kriging to account for missing data in unsampled regions. PREC is taken from the Global Precipitation Climatology Centre (GPCC) Full Data Reanalysis version 6 (Schneider et al., 2011) for the period 1901–2010 combined with additional data for the period 2011–2013 taken from the GPCC monitoring product following bias correction. Further details of the sources of predictor data are given in Table 1. Our list of predictors is not exhaustive. Much recent work has sought to identify predictability arising from the extent of sea ice and snow covered land, the reflective and insulative attributes of which are relevant for SAT and

3948

PREC in several regions of the extra-tropics (e.g. Shongwe et al., 2007; Dutra et al., 2011; Brands et al., 2012; Chevallier and Salas-Méla, 2012). However, these variables are not considered for the present system due to the absence of sufficiently long and reliable datasets, although some effects are effectively captured by persistence. The design of the prediction system facilitates inclusion of additional predictors should high quality observational or reanalysis data become available.

Global maps of linear correlation between predictand-predictor pairs form a basis for a physical understanding of the factors governing variability and provide a first step in determining predictive potential. Predictors that show good potential and do not exhibit colinearity with other predictors are included in the selection procedure for the empirical forecast model itself. The model is calibrated and validated in a hindcast framework using a causal approach: hindcasts are produced for 1961–2010 using data since 1901 prior to the forecast year. The causal approach was chosen instead of a leave-one-out framework in order to replicate the set of observational data that would have been available for each hindcast were it produced in real time. Separate models are calibrated for each three-month season with a one month lead time (the season March–April–May is thus predicted by November–December–January). CO2EQV is included as a predictor for all seasons at all locations. In order to identify subsequent informative predictors, the linear regression onto CO2EQV is removed from the predictand and each predictor. Correlation is performed between the detrended predictand and predictor pairs. Predictor selection is season- and location-specific; for each season and grid point, the predictors that produce significant (at the 90 % level) correlation with the predictand are entered into a multiple linear regression along with CO2EQV. This strategy is designed to avoid overfitting, which would lower skill scores, and to ensure that the empirical model is built on physical principles to the fullest extent. Predictor inclusion is determined independently for each hindcast, allowing for the maximum value to be taken from predictor information in the fairest way.

A key component of the empirical prediction system is the provision of probabilistic output. The residuals from the regression fit for each hindcast are used to generate

3949

an ensemble of predictions. The set of residuals is randomly sampled 51 times (with replacement), reflecting the number of members in a typical dynamical forecast ensemble. The ensemble allows for the calculation of probabilistic skill scores and will provide a basis for full comparison with the output of dynamical systems. It is anticipated that future development of the system will consider more complex methods of ensemble generation.

Both the deterministic and probabilistic aspects of the prediction system must be systematically validated using a number of measures. Global maps of correlation between hindcast estimates and observations provide a view on the degree of representation of temporal variability. Verification scores originally developed in the context of numerical weather prediction, including the root mean squared error skill score (RMSESS) and the continuous rank probability skill score (CRPSS), provide a quantification of the degree of bias and the skill of the probability distribution produced by the ensemble respectively. Such verification measures are also used to determine skill scores that describe forecast skill against a reference ensemble forecast. The reference forecast is produced by random sampling of the climatology, i.e. the observations for each year in the fitting period.

3 Analysis of potential predictors

3.1 Surface air temperature

The surface air temperature (SAT) shows a clear trend almost everywhere. We parametrise this trend as proportional to the forcing of greenhouse gases, described by the equivalent CO₂ concentrations, which describes the global mean trend quite well ($r \sim 0.93$). Separate spatially varying aerosol forcings have not yet been implemented. As mentioned in Sect. 2, this trend is treated differently from the other predictors in the sense it is always included in the empirical model; all other predictors are included only in cases where they appear to add value. Figure 1 shows seasonal correlation between

3950

SAT and the previous year's CO2EQV along the top row of panels. Subsequent rows show the correlation derived from predictor-predictand pairs following removal of the trend associated with CO2EQV. Correlation between SAT and CO2EQV is in general strongly positive across the majority of the globe, and particularly so when natural variability is small compared to the forced signal, for example in the Northern Hemisphere during spring (MAM) and summer (JJA) and throughout the tropics at all times of year.

NINO3.4 shows the second strongest relationship with SAT; the importance of ENSO in governing variability in temperatures across the tropics is highlighted by correlation stronger than ± 0.5 in parts of South America, Africa and northern Australia in addition to the tropical Pacific and Indian Oceans. ENSO-based relationships in extra-tropical land regions are less apparent, although positive correlation in the northern half of the North American continent and negative ones around the Gulf of Mexico show the well-known influence on winter (DJF) and spring (MAM) SAT (Ropelewski and Halpert, 1987; Kiladis and Diaz, 1989). Very low correlations are found across Europe.

The PDO and IOD correlation patterns are very similar to those for NINO3.4. Much of the signal associated with PDO is captured by NINO3.4; additional skill is confined to the northern Pacific, which is likely to be associated with the region of enhanced cyclonic circulation around the deepened Aleutian low associated with a positive, warm PDO phase (Liu and Alexander, 2007). Other areas of stronger correlation include small areas of central North America during summer, which supports the association of PDO with multidecadal drought frequency in the US (McCabe et al., 2004). The AMO correlation patterns clearly act independently of ENSO and feature correlations throughout the high northern latitudes and the North Atlantic, but curiously not so much in Western Europe (van Oldenborgh et al., 2009b). The PDO, IOD and AMO indices are all included for selection in the prediction system.

Correlation associated with the QBO is poor with the notable exception of northern and central Russia during the Boreal autumn. In agreement with Folland et al. (2012) we found no significant correlation for winter in Europe with a one month lead time. This is surprising given the link found in previous work between the QBO and the Arctic

3951

Oscillation (AO), and thus on European surface climate, although the authors suggest that predictability requires a shorter optimal lead time than that used here (Marshall and Scaife, 2009). We thus omit QBO from the SAT prediction system.

Persistence (PERS) shows strong correlations in some key regions and is particularly important for high latitude seas in the Northern Hemisphere during winter, reflecting the latent heat of melting of the sea ice. Over land however, there are relatively few regions associated with strong correlation outside of the tropics. Correlation is greater than 0.4 in parts of western Europe (MAM), south-east Europe (JJA), central North America (JJA) and parts of central Asia (JJA). However, aside from these examples, the memory of land surface temperature outside of the tropics does not appear to extend to the predictor period.

Unsurprisingly, including local SST (LSST) produces higher correlation than persistence over the oceans but offers no skill over most continental regions. However, LSST is clearly beneficial in coastal regions, including northern and western Europe. We thus make both predictors available for selection in the SAT forecast system. The relationship between antecedent precipitation (CPREC) and SAT is in general quite poor but correlation is around 0.4 in northern Europe during spring (MAM), most likely representing the connection between a mild, wet winter to a mild spring. The negative correlation during summer (JJA), significant over France, suggests that CPREC is reasonably able to represent the link between soil moisture and SAT at this time of year shown in previous work (van den Hurk et al., 2012). The correlation is also strong in parts of Australia and south-east Asia, in addition to southern Africa (MAM) and northern South America (DJF and MAM).

3.2 Precipitation

Lagged correlation between PREC and the predictors is shown in Fig. 2. As expected, the response of PREC to the trend in CO2EQV is not as strong as that of global temperature. Increased PREC in northern high latitudes during the Boreal winter has a known association with global warming (Hartmann et al., 2013). However, the response of

3952

precipitation to global warming is not yet visible above the noise in much of the mid-latitudes and these regions are associated with low correlation at all times of the year.

The strong lagged correlation exhibited between NINO3.4 and PREC in many parts of the world provides the most important basis for predictability. In addition to ENSO-related changes in tropical precipitation patterns, there are a number of known links with precipitation in the extra-tropics (Alexander et al., 2002; Doblas-Reyes et al., 2013), although only a weak one in MAM is found in Europe (van Oldenborgh et al., 2000). Correlation patterns for the PDO are again similar for NINO3.4. For the IOD, correlations of around 0.5 exists in eastern Africa during autumn (SON) and winter (DJF) but again these patterns are very similar to those for NINO3.4. Correlation of IOD and PREC following removal of the NINO3.4 signal (not shown) indicates an ENSO-independent relationship, particularly during DJF in East Africa, which is supported by the findings of previous work (Goddard and Graham, 1999), and also parts of Europe. In the absence of known links between the phase of PDO and precipitation anomalies that are independent of ENSO, PDO is omitted from the prediction system. QBO is also omitted on the basis that there are few areas of correlation of statistical significance. AMO on the other hand produces significant correlation in regions influenced by the Atlantic where NINO3.4 does not, including the Sahel (JAS, visible in JJA and SON), eastern South America (JJA). The AMO-PREC relationship does not appear to extend to extra-tropical regions; there are no discernible areas of strong correlation in Europe or eastern North America. This contrasts with the strong link previously identified between the AMO and JJA precipitation in Europe during the 1990s (Sutton and Dong, 2012). The use of long-term time series, correlations rather than composites and an absence of temporal filtering here results in lower correlations.

For PERS, there are a number of regions, particularly in the extra-tropics, where significant correlation offer potential for predictability. The most obvious of such correlation is during DJF in the mid- to high-latitudes of the Northern Hemisphere; the persistence of dry (wet) conditions during autumn in much of central Eurasia is an indicator for similar conditions during winter. In Europe, significant negative correlation during summer

3953

(JJA) suggests evidence for dry (wet) springs followed by wet (dry) summers. By contrast, there are relatively few regions where LSST is significantly correlated with PREC. These include the western US (MAM) and south-east Asia where SST has variability that is independent from ENSO and adds to the skill in dynamical systems (van Oldenborgh et al., 2005a). It remains to be seen to what extent LSST may offer additional value to this empirical prediction system.

4 Prediction system development and validation

For each hindcast between 1961–2010, and for each season and grid point, predictors are selected on the basis of the significance of the (detrended) correlation with the predictand for the fitting period. For validation, causal hindcast estimates are compared with observations to determine the skill of the deterministic and probabilistic aspects of the prediction system.

4.1 Surface air temperature

Following the assessment of potential predictors, the following were chosen in addition to CO₂EQV for inclusion in the prediction system: NINO3.4, PDO, AMO, IOD, PERS, LSST and CPREC. Hindcasts were produced with each predictor added in turn and verified against observations. Figure 3 shows the correlation between observations and a hindcast constructed using CO₂-equivalent only (top line), the incremental correlation attained by including additional predictors (second to eighth lines), and the correlation following the inclusion of all predictors. Note that these are the correlations of a causal system that only uses information from before the hindcast date, the values are therefore much lower than the full correlations of Fig. 1. If the correlations are spurious, i.e., there was no physical connection, but the predictor was included because the correlation exceeded the 90% significance criterion (this happens by chance on 10% of the grid points without connection) the hindcast skill is degraded by the inclusion of

3954

the hindcast produces a positive CRPSS, which would indicate an improvement on the ensemble forecast derived from the climatology (Fig. 6; right panels). This leads us to conclude that, while the deterministic component of the system is able to reproduce some components of seasonal precipitation variability, probabilistically the system does not perform well outside limited areas in its present guise.

5 Discussion and outlook

A global empirical system for seasonal climate prediction has been developed and validated. Multiple linear regression was chosen as the basis of the system; a simple predictor selection scheme sought to maximise the predictive skill of a number of predictors describing global-scale modes of variability and local-scale information alongside that of the climate change signal. Probabilistic hindcasts of surface air temperature (SAT) and precipitation (PREC) have been produced using prediction models based on multiple linear regression and validated against observations using correlation and skill scores. The prediction system shows good skill in many regions. For SAT, the trend and interannual variability are well-represented throughout the tropics and in a number of extra-tropical regions, including parts of Europe, particularly during spring and summer, southern Africa and eastern Australia. Skill associated with the probabilistic component of the seasonal predictions shows similar spatial patterns. For PREC, few areas of notable skill are found outside of regions with known ENSO teleconnections and, probabilistically, the system does not perform better than a climatological ensemble throughout most of the world.

As outlined in Sect. 1, the system presented here has been designed to serve both as a benchmark for dynamical prediction systems and as an independent forecast system to be combined with dynamical output to produce more robust forecasts. Concerning the second purpose, it is important to identify seasons and regions where dynamical systems lack skill and whether our system may potentially add value in such instances. In general, dynamical system skill is limited to regions that are strongly linked to ENSO;

3957

in extra-tropical regions, where seasonal variability in the atmospheric state is governed to a greater extent by random internal variability, skill is inevitably lower than in the tropics (Kumar et al., 2007; Arribas et al., 2011). The good skill in many parts of Europe, particularly for forecasts of SAT, is an encouraging property of our system and a detailed comparison with dynamical European forecasts is forthcoming. The inclusion of locally-varying predictors, in combination with predictors describing large-scale modes of variability provides a basis to elicit more skill than can be attained using global indices alone.

An important outcome of this work is the system's implementation in a quasi-operational framework and the provision of regular forecasts. Monthly forecasts are generated for each forthcoming three-month season and made publicly available through the KNMI Climate Explorer along with uncertainty parameters and updated hindcast validation. The system's framework permits the potential to test empirical prediction methods other than linear regression, such as neural networks that potentially capture non-linear aspects of the climate system. Additionally, as mentioned in Sect. 2, the current list of predictors considered for inclusion is not exhaustive and there is scope to better exploit the predictive information in other locally-varying predictors. Further avenues for system development include region-specific and case-based analysis and application to alternative predictands from century-long reanalyses or those describing extreme events. Focus will also be given to alternative methods of ensemble generation using, for instance, derived uncertainty in regression parameters and spatial patterns.

Acknowledgements. This work has been undertaken within the SPECS project, funded by the Seventh Framework Programme (FP7) of the European Commission (Grant Agreement No.: GA 308378). We also acknowledge the helpful comments and suggestions of A. Sterl.

References

- Alexander, M. A., Bladé, I., Newman, M., Lanzante, J. R., Lau, N.-C., and Scott, J. D.: The atmospheric bridge: the influence of ENSO teleconnections on air-sea interaction over the global oceans, *J. Climate*, 15, 2205–2231, 2002. 3947, 3953
- 5 Arribas, A., Glover, M., Maidens, A., Peterson, K., Gordon, M., MacLachlan, C., Graham, R., Fereday, D., Camp, J., Scaife, A. A., Xavier, P., McLean, P., Colman, A., and Cusack, S.: The GloSea4 ensemble prediction system for seasonal forecasting, *Mon. Weather Rev.*, 139, 1891–1910, 2011. 3958
- 10 Baldwin, M. P., Gray, L. J., Dunkerton, T. J., Hamilton, K., Haynes, P. H., Randel, W. J., Holton, J. R., Alexander, M. J., Hirota, I., Horinouchi, T., Jones, D. B. A., Kinnerson, J. S., Marquardt, C., Sato, K., and Takahashi, M.: The quasi-biennial oscillation, *Rev. Geophys.*, 39, 179–229, 2001. 3947
- Balmaseda, M. and Anderson, D.: Impact of initialization strategies and observations on seasonal forecast skill, *Geophys. Res. Lett.*, 36, L01701, doi:10.1029/2008GL035561, 2009. 3947
- 15 Brands, S., Manzananas, R., Gutiérrez, J. M., and Cohen, J.: Seasonal predictability of wintertime precipitation in Europe using the snow advance index, *J. Climate*, 25, 4023–4028, 2012. 3949
- Brönnimann, S., Annis, J. L., Vogler, C., and Jones, P. D.: Reconstructing the quasi-biennial oscillation back to the early 1900s, *Geophys. Res. Lett.*, 34, L22805, doi:10.1029/2007GL031354, 2007. 3964
- 20 Chevallier, M. and Salas-Méila, D.: The role of sea ice thickness distribution in the Arctic sea ice potential predictability: a diagnostic approach with a coupled GCM, *J. Climate*, 25, 3025–3038, 2012. 3949
- 25 Coelho, C. A. S., Stephenson, D. B., Balmaseda, M., Doblas-Reyes, F. J., and van Oldenborgh, G. J.: Toward an integrated seasonal forecasting system for South America, *J. Climate*, 19, 3704–3721, doi:10.1175/JCLI3801.1, 2006. 3943, 3944
- Cowtan, K. and Way, R. G.: Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends, *Q. J. Roy. Meteor. Soc.*, 140, 1935–1944, 2014. 3948
- 30 Doblas-Reyes, F. J., Hagedorn, R., Palmer, T. N., and Morcrette, J.-J.: Impact of increasing greenhouse gas concentrations in seasonal ensemble forecasts, *Geophys. Res. Lett.*, 33, L07708, doi:10.1029/2005GL025061, 2006. 3945

3959

- Doblas-Reyes, F. J., Garcia-Serrano, J., Lienert, F., Pinto Biescas, A., and Rodrigues, L. R. L.: Seasonal climate predictability and forecasting: status and prospects, *Wiley Interdisciplinary Reviews-Climate Change*, 4, 245–268, doi:10.1002/wcc.217, 2013. 3942, 3943, 3944, 3947, 3953
- 5 Dutra, E., Schär, C., Viterbo, P., and Miranda, P.: Land-atmosphere coupling associated with snow cover, *Geophys. Res. Lett.*, 38, L15707, doi:10.1029/2011GL048435, 2011. 3949
- Ebdon, R. A. and Verryard, R. G.: Fluctuations in equatorial stratospheric winds, *Nature*, 189, 791–793, 1961. 3947
- Folland, C. K., Scaife, A. A., Lindesay, J., and Stephenson, D. B.: How potentially predictable is northern European winter climate a season ahead?, *Int. J. Climatol.*, 32, 801–818, 2012. 3947, 3951
- 10 Goddard, L. and Graham, N. E.: Importance of the Indian Ocean for simulating rainfall anomalies over eastern and southern Africa, *J. Geophys. Res.-Atmos.*, 104, 19099–19116, 1999. 3953
- 15 Hartmann, D. L., Klein Tank, A. M. G., Rusticucci, M., Alexander, L. V., Brönnimann, S., Charabi, Y., Dentener, F. J., Dlugokencky, E. J., Easterling, D. R., Kaplan, A., Soden, B. J., Thorne, P. W., Wild, M., and Zhai, P. M.: Observations: atmosphere and surface, in: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by: Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, A., and Midgley, P. M., Cambridge University Press, Cambridge, UK and New York, NY, USA., 2013. 3952
- 20 Ho, C. K., Hawkins, E., Shaffrey, L., and Underwood, F. M.: Statistical decadal predictions for sea surface temperatures: a benchmark for dynamical GCM predictions, *Clim. Dynam.*, 41, 917–935, doi:10.1007/s00382-012-1531-9, 2013. 3944
- 25 Kennedy, J. J., Rayner, N. A., Smith, R. O., Parker, D. E., and Saunby, M.: Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 1. Measurement and sampling uncertainties, *J. Geophys. Res.-Atmos.*, 116, D14103, doi:10.1029/2010JD015220, 2011a. 3964
- 30 Kennedy, J. J., Rayner, N. A., Smith, R. O., Parker, D. E., and Saunby, M.: Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 2. Biases and homogenization, *J. Geophys. Res.-Atmos.*, 116, D14104, doi:10.1029/2010JD015218, 2011b. 3964

3960

- Kiladis, G. N. and Diaz, H. F.: Global climatic anomalies associated with extremes in the Southern Oscillation, *J. Climate*, 2, 1069–1090, 1989. 3951
- Kim, H.-M., Webster, P. J., and Curry, J. A.: Seasonal prediction skill of ECMWF System 4 and NCEP CFSv2 retrospective forecast for the Northern Hemisphere Winter, *Clim. Dynam.*, 39, 2957–2973, 2012. 3956
- 5 Krueger, O. and Von Storch, J.-S.: A simple empirical model for decadal climate prediction, *J. Climate*, 24, 1276–1283, 2011. 3944
- Kumar, A., Jha, B., Zhang, Q., and Bounoua, L.: A new methodology for estimating the unpredictable component of seasonal atmospheric variability, *J. Climate*, 20, 3888–3901, 2007. 3958
- 10 Liniger, M. A., Mathis, H., Appenzeller, C., and Doblas-Reyes, F. J.: Realistic greenhouse gas forcing and seasonal forecasts, *Geophys. Res. Lett.*, 34, L04705, doi:10.1029/2006GL028335, 2007. 3945
- Liu, Z. and Alexander, M.: Atmospheric bridge, oceanic tunnel, and global climatic teleconnections, *Rev. Geophys.*, 45, RG2005, doi:10.1029/2005RG000172, 2007. 3947
- 15 Marshall, A. G. and Scaife, A. A.: Impact of the QBO on surface winter climate, *J. Geophys. Res.-Atmos.*, 114, D18110, doi:10.1029/2009JD011737, 2009. 3952
- Mason, S. J. and Mimmack, G. M.: Comparison of some statistical methods of probabilistic forecasting of ENSO, *J. Climate*, 15, 8–29, 2002. 3946
- 20 McCabe, G. J., Palecki, M. A., and Betancourt, J. L.: Pacific and Atlantic Ocean influences on multidecadal drought frequency in the United States, *P. Natl. Acad. Sci. USA*, 101, 4136–4141, 2004. 3951
- Meinshausen, M., Smith, S. J., Calvin, K., Daniel, J. S., Kainuma, M. L. T., Lamarque, J. F., Matsumoto, K., Montzka, S. A., Raper, S. C. B., Riahi, K., Thomson, A., Velders, G. J. M., and van Vuuren, D. P. P.: The RCP greenhouse gas concentrations and their extensions from 1765 to 2300, *Climatic Change*, 109, 213–241, 2011. 3947, 3964
- 25 Morice, C. P., Kennedy, J. J., Rayner, N. A., and Jones, P. D.: Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: the HadCRUT4 data set, *J. Geophys. Res.-Atmos.*, 117, D08101, doi:10.1029/2011JD017187, 2012. 3948
- 30 Newman, M.: An empirical benchmark for decadal forecasts of global surface temperature anomalies, *J. Climate*, 26, 5260–5269, 2013. 3944

3961

- Peng, P., Kumar, A., Barnston, A. G., and Goddard, L.: Simulation skills of the SST-forced global climate variability of the NCEP-MRF9 and the Scripps-MPI ECHAM3 models, *J. Climate*, 13, 3657–3679, 2000. 3943
- Penland, C. and Matrosova, L.: Prediction of tropical Atlantic sea surface temperatures using linear inverse modeling, *J. Climate*, 11, 483–496, 1998. 3943
- 5 Quesada, B., Vautard, R., Yiou, P., Hirschi, M., and Seneviratne, S. I.: Asymmetric European summer heat predictability from wet and dry southern winters and springs, *Nature Climate Change*, 2, 736–741, 2012. 3948
- Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., Kent, E. C., and Kaplan, A.: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century, *J. Geophys. Res.-Atmos.*, 108, 4407, doi:10.1029/2002JD002670, 2003. 3964
- 10 Ropelewski, C. F. and Halpert, M. S.: Global and regional scale precipitation patterns associated with the El Niño/southern oscillation, *Mon. Weather Rev.*, 115, 1606–1626, 1987. 3951
- 15 Sardeshmukh, P. D., Compo, G. P., and Penland, C.: Changes of probability associated with El Niño, *J. Climate*, 13, 4268–4286, 2000. 3943
- Schepen, A., Wang, Q. J., and Robertson, D. E.: Combining the strengths of statistical and dynamical modeling approaches for forecasting Australian seasonal rainfall, *J. Geophys. Res.-Atmos.*, 117, D20107, doi:10.1029/2012JD018011, 2012. 3944
- 20 Schneider, U., Becker, A., Finger, P., Meyer-Christoffer, A., Rudolf, B., and Ziese, M.: GPCC Full Data Reanalysis Version 6.0 at 2.5°: Monthly Land-Surface Precipitation from Rain-Gauges built on GTS-based and Historic Data, doi:10.5676/DWD_GPCC/FD_M_V6_250, 2011. 3948, 3964
- Shongwe, M. E., Ferro, C. A. T., Coelho, C. A. S., and van Oldenborgh, G. J.: Predictability of cold spring seasons in Europe, *Mon. Weather Rev.*, 135, 4185–4201, 2007. 3949
- 25 Sterl, A., van Oldenborgh, G. J., Hazeleger, W., and Burgers, G.: On the robustness of ENSO teleconnections, *Clim. Dynam.*, 29, 469–485, 2007. 3944
- Suckling, E. B. and Smith, L. A.: An evaluation of decadal probability forecasts from state-of-the-art climate models, *J. Climate*, 26, 9334–9347, 2013. 3943
- 30 Sutton, R. T. and Dong, B.: Atlantic Ocean influence on a shift in European climate in the 1990s, *Nat. Geosci.*, 5, 788–792, 2012. 3953
- van den Dool, H.: *Empirical Methods in Short-Term Climate Prediction*, Oxford University Press, 2007. 3943, 3944

3962

- van den Hurk, B., Doblas-Reyes, F., Balsamo, G., Koster, R. D., Seneviratne, S. I., and Camargo Jr., H.: Soil moisture effects on seasonal temperature and precipitation forecast scores in Europe, *Clim. Dynam.*, 38, 349–362, 2012. 3948, 3952
- van Oldenborgh, G. J., Burgers, G., and Klein Tank, A.: On the El Niño teleconnection to spring precipitation in Europe, *Int. J. Climatol.*, 20, 565–574, doi:10.1002/(SICI)1097-0088(200004)20:5<565::AID-JOC488>3.0.CO;2-5, 2000. 3953
- van Oldenborgh, G. J., Balmaseda, M. A., Ferranti, L., Stockdale, T. N., and Anderson, D. L. T.: Evaluation of atmospheric fields from the ECMWF seasonal forecasts over a 15-year period, *J. Climate*, 18, 3250–3269, 2005a. 3947, 3954
- van Oldenborgh, G. J., Balmaseda, M. A., Ferranti, L., Stockdale, T. N., and Anderson, D. L. T.: Did the ECMWF seasonal forecast model outperform statistical ENSO forecast models over the last 15 years?, *J. Climate*, 18, 3240–3249, 2005b. 3943
- van Oldenborgh, G. J., te Raa, L. A., Dijkstra, H. A., and Philip, S. Y.: Frequency- or amplitude-dependent effects of the Atlantic meridional overturning on the tropical Pacific Ocean, *Ocean Sci.*, 5, 293–301, doi:10.5194/os-5-293-2009, 2009a. 3964
- van Oldenborgh, G. J., te Raa, L. A., Dijkstra, H. A., and Philip, S. Y.: Frequency- or amplitude-dependent effects of the Atlantic meridional overturning on the tropical Pacific Ocean, *Ocean Sci.*, 5, 293–301, doi:10.5194/os-5-293-2009, 2009b. 3951
- Weisheimer, A. and Palmer, T. N.: On the reliability of seasonal climate forecasts, *J. Roy. Soc. Interface*, 11, 20131162, doi:10.1098/rsif.2013.1162, 2014. 3942
- Weisheimer, A., Doblas-Reyes, F. J., Palmer, T. N., Alessandri, A., Arribas, A., Déqué, M., Keenlyside, N., MacVean, M., Navarra, A., and Rogel, P.: ENSEMBLES: A new multi-model ensemble for seasonal-to-annual predictions Skill and progress beyond DEMETER in forecasting tropical Pacific SSTs, *Geophys. Res. Lett.*, 36, L21711, doi:10.1029/2009GL040896, 2009. 3947
- Zhao, M. and Hendon, H. H.: Representation and prediction of the Indian Ocean dipole in the POAMA seasonal forecast model, *Q. J. Roy. Meteor. Soc.*, 135, 337–352, 2009. 3947

3963

Table 1. Description of predictor variables and their sources.

Predictor	Source
CO2EQV	CO ₂ -equivalent concentrations (Meinshausen et al., 2011)
NINO3.4	Calculated from SST fields from HadISST (Rayner et al., 2003)
PDO	University of Washington (http://research.jisao.washington.edu/pdo/)
QBO	At 30 hPa from the reconstruction of Brönnimann et al. (2007)
AMO	Calculated by van Oldenborgh et al. (2009a); based on HadSST (Kennedy et al., 2011a, b)
IOD	Calculated from SST fields from HadISST (Rayner et al., 2003)
LSST	HadSST3 (Kennedy et al., 2011a, b)
CPREC	GPCC Full Data Reanalysis version 6 (Schneider et al., 2011)

3964

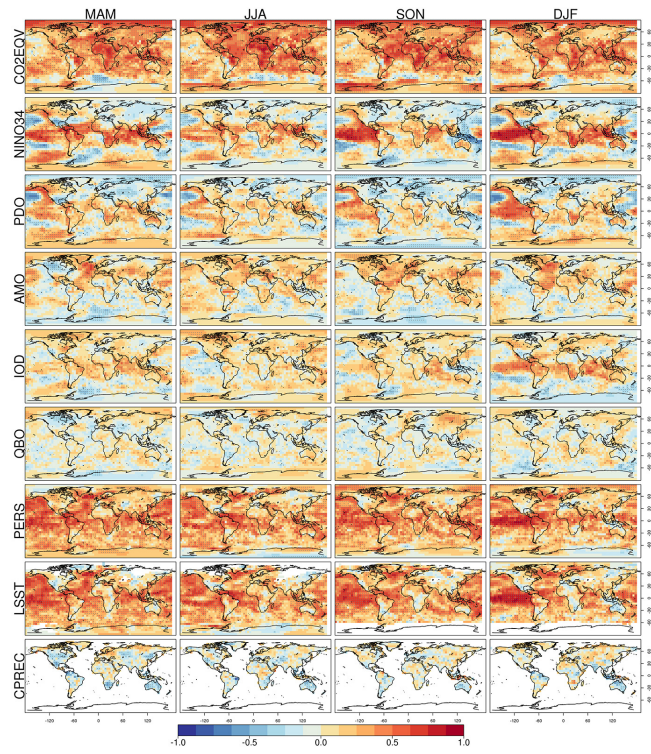


Figure 1. Correlation between seasonal SAT and the set of predictors with a one month lead time. Correlation between CO2EQV is shown in the top line; subsequent lines show correlation between predictand-predictor pairs following removal of the CO2EQV trend.

3965

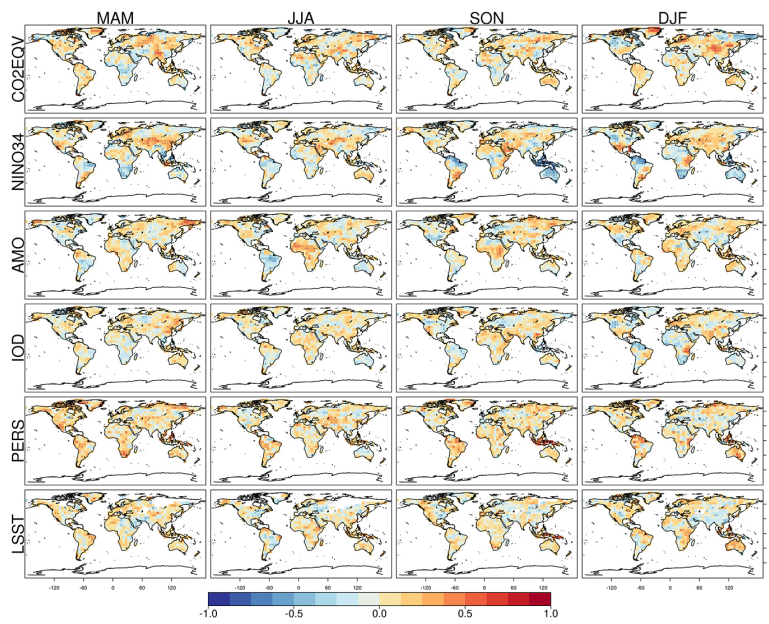


Figure 2. Correlation between seasonal PREC and the set of predictors with a one month lead time. As in Fig. 1, correlation between CO2EQV is shown in the top line; subsequent lines show correlation between predictand-predictor pairs following removal of the CO2EQV trend.

3966

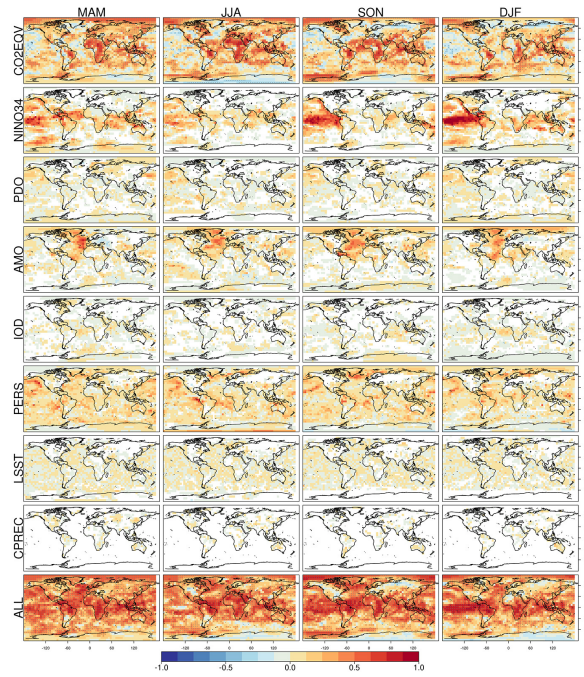


Figure 3. Correlation between SAT hindcasts and observations. The top line shows correlation between observations and SAT hindcasts constructed using CO₂-equivalent as the sole predictor. Subsequent lines show the difference in correlation following the inclusion of additional predictors. The bottom line shows the correlation for the full model. For the top and bottom lines, stippling is used to indicate significance at the 95 % level.

3967

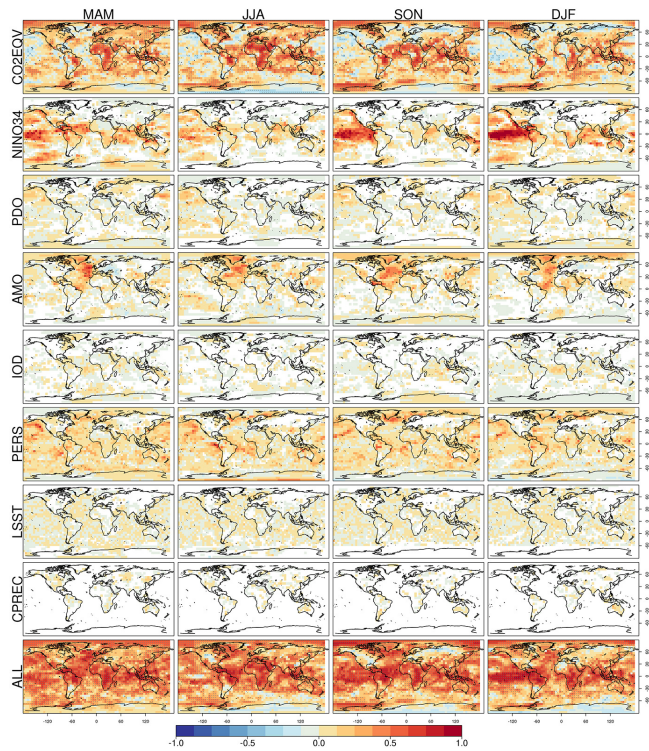


Figure 4. RMSESS and CRPSS of the SAT hindcasts expressed as a skill score against a climatology ensemble forecast. For CRPSS, stippling is used to indicate significance at the 95 % level following a one sided *t* test.

3968

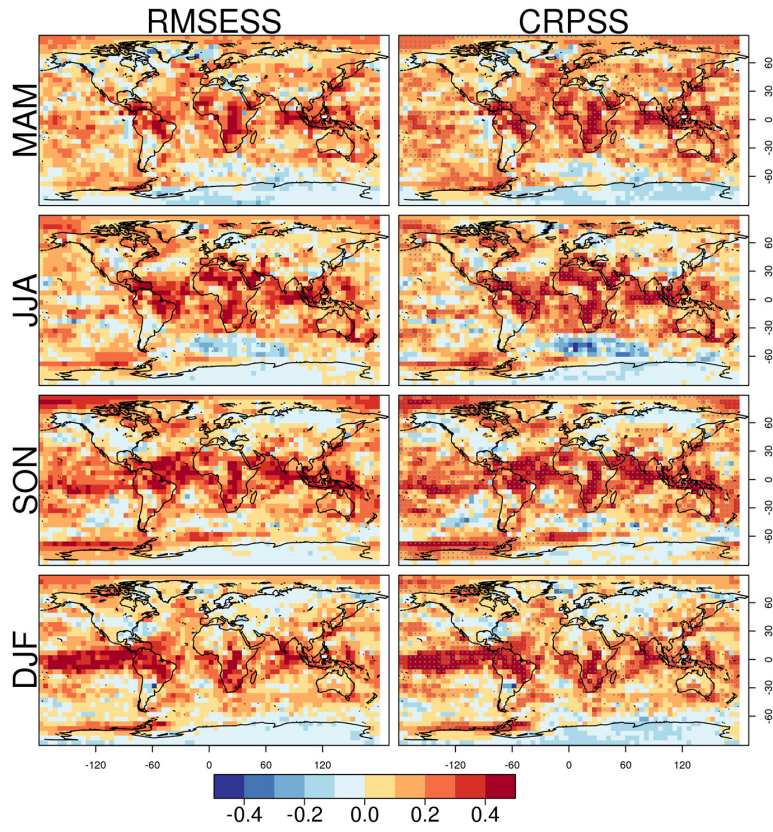


Figure 5. As Fig. 3 but for PREC.

3969

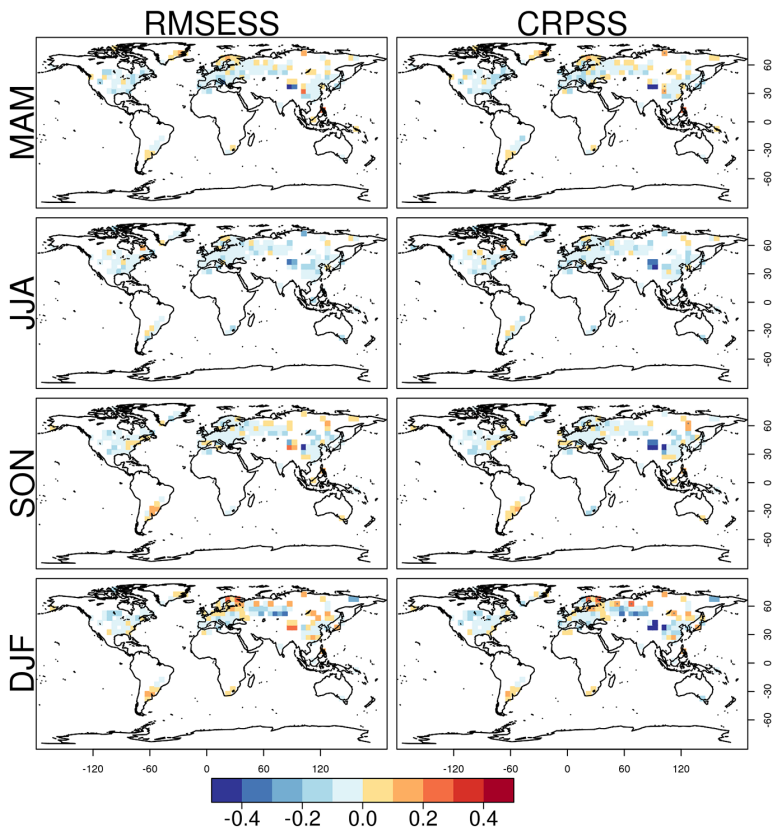


Figure 6. As Fig. 4 but for PREC.

3970