Multivariate

– inter-variable, spatial and temporal –

bias correction

MathieuVrac *

Laboratoire des Sciences du Climat et de l'Environnement (LSCE-IPSL, CNRS),

Centre d'Etudes de Sacaly, Orme des Merisiers, 91190 Gif-sur-Yvette, France

Petra Friederichs

Meteorological Institute University of Bonn, Auf dem Hügel 20, 53121 Bonn, Germany

* *Corresponding author address:*

Dr. Mathieu Vrac, LSCE, Centre d'étude de Saclay, Orme des Merisiers, 91191, Gif-sur Yvette, France.

E-mail: mathieu.vrac@lsce.ipsl.fr

ABSTRACT

Statistical methods to bias correct global or regional climate model output are now common to get data closer to observations in distribution. However, most bias correction (BC) methods work for one variable and one location at a time and basically reproduce the temporal structure of the models. The inter-variable, spatial and temporal dependencies of the corrected data are usually poor compared to observations. Here, we propose a novel method for multivariate BC. The empirical copula - bias correction (EC-BC) combines a 1-dimensional BC with a shuffling technique that restores an empirical multidimensional copula. Several BC methods are investigated and compared to high-resolution reference data over the French Mediterranean basin, notably: (i) a 1d-BC method applied independently to precipitation and temperature fields, (ii) a recent conditional correction approach developed for producing correct two-dimensional inter-variable structures, (iii) the EC-BC method.

Assessments are realized in terms of inter-variable, spatial and temporal dependencies and an objective evaluation using the integrated quadratic distance (IQD) is presented. As expected, the 1d methods cannot produce correct multidimensional properties. The conditional technique appears efficient for inter-variable properties but not for spatial and temporal dependencies. EC-BC provides realistic dependencies in all respects, inter-variable, spatial and temporal. The IQD results are clearly in favour of EC-BC. As many BC methods, EC-BC relies on a stationarity assumption and is only able to reproduce patterns inherited from historical data. However, due to its easiness of coding, its speed of application and the quality of its results, the EC-BC method is a very good candidate for all needs in multivariate bias correction.

1. Introduction

The use of simulations from climate or meteorological models at large or regional scales is now common in many impact studies, such as hydrological, environmental, or economic studies among others, or more generally in studies on consequences of climate change and adaptation. Although those simulations provide much useful information, they are in general not directly comparable to observations: for example, many observations are point measurements, whereas simulated data represent volume integrated dynamical variables. Moreover, simulated data are associated with potential biases in the sense their statistical distribution differs from the distribution of the observations. This is partly due to the fact that Global Climate Models (GCMs) have too low a spatial resolution to be employed directly in most of the impact models (e.g., Meehl 2007; Christensen et al. 2008). Regional Climate Models (RCMs) reduce some of the biases but not those unrelated to spatial resolution (Maraun 2013; White and Toumi 2013). Statistical bias correction methods – correcting the distribution (e.g., the cumulative distribution function) – are then commonly applied to transform the simulated data into new data with no, or at least fewer, statistical biases with respect to reference, generally observed, time series. In general, there is no clear distinction between a change of support problem (i.e. down- or upscaling), and bias correction.

In all the following, capital letters – e.g., X – represent random variables, while small letters – e.g., x – are used for realizations or values of a random variable. The most employed bias correction (BC) methods are based on quantile-association. The most famous is certainly the so-called quantile-mapping approach (Panofsky and Brier 1958; Haddad and Rosenfeld 1997; Wood et al. 2004; Déqué 2007; Piani et al. 2010; Gudmundsson et al. 2012), trying to map a modeled value x (with a cumulative distribution function – CDF – F_X) to an observed value y (with a CDF F_Y) through a function f, such that their distributions are equivalent (Piani et al. 2010b):

$$y = h(x) \text{ such that } F_Y(y) = F_X(x). \quad (1)$$

This mapping function h can be derived from distributions, regression-like transformations, in both cases either parametric or non-parametric (see, e.g., Gudmundsson et al. 2012, for some details). A very popular distribution-derived non-parametric approach (e.g., Déqué 2007) directly uses the constraint $F_X(x) = F_Y(y)$ to derive the corrected value y from the modeled value x through the so-called "Empirical Quantile Mapping" (EQM):

$$y = F_Y^{-1}(F_X(x)) \quad (2)$$

where F^{-1} is the inverse function of the CDF F, both modeled non-parametrically.

One major issue of such quantile mapping methods and their variants (e.g., Michelangeli et al. 2009) is that they are essentially univariate: they work only for one variable at a time, one location at a time and basically reproduce the temporal structure of the climate models. Hence, although the resulting marginal (i.e., one-dimensional) statistical distributions of the corrected data are improved, those one-dimensional techniques suffer from various limitations. Among the latter, one major limitation for many impact studies is that, because they are applied to one location at a time, the spatial and temporal structures of the corrected time series are misrepresented (Colette et al. 2012; Maraun 2013) and basically correspond to the structures of the model to be corrected. This leads to potentially significant inadequacies when used as forcing, for example in a hydrological model where spatialization and chronology of the input rainfall are of importance.

Moreover, as most of the BC methods correct one variable at a time – e.g., temperature is corrected separately and independently from precipitation –, the corrected variables can be inconsistent between each other and then generate unrealistic situations (e.g., Chen et al. 2011; Muerth et al. 2013).

Such (spatial, temporal and inter-variable) issues also appear when BC is applied to de-bias GCM outputs prior to downscaling with regional climate models. Although Colette et al. (2012) and White and Toumi (2013) showed that such a "prior" correction of the large-scale inputs for RCMs with a quantile-association based method clearly improves the quality of the RCM simulations, White and Toumi (2013) found that it can nevertheless produce undesirable results in the RCM simulations.

Recently, efforts have been made to improve or create BC models that solve (some of) those issues. Piani and Haerter (2012) developed a BC methodology to bypass the problem of physical consistency between two variables (e.g., temperature and precipitation) to be corrected. Their approach consists in applying a univariate BC to the time series of one variable (e.g., precipitation) conditionally on the bias corrected values of the time series for the other variable (e.g., temperature). Their results show the clear improvement of the temperature-precipitation dependence representation with respect to the traditional separate univariate temperature and precipitation bias corrections.

Furthermore, to overcome the lack of realistic spatial variability and temporal persistence in precipitation and temperature fields simulated by a numerical weather prediction (NWP) model, Clark et al. (2004) presented a method for reordering NWP outputs to recover the space-time variability. In this approach, each time series is ranked and matched with observation data. The element of the time series are then shuffled to match the original order of the historical dataset. Based on this shuffling technique, Clark et al. (2004) correctly reconstructed the space-time variability of forecasted precipitation and temperature fields. This technique has seen great success in hydrological applications, e.g. for flood forecasts (Voisin et al. 2010, 2011), to construct ensemble forecasts from single-value forecasts of precipitation and temperature (Schaake et al. 2007), or for ensemble post-processing (Verkade et al. 2013; Robertson et al. 2013). The ensemble copula coupling (ECC) is an adaptation thereof to multivariate ensemble postprocessing (Schefzik et al. 2013a; Möller et al. 2012; Schuhen et al. 2012; Thorarinsdottir et al. 2013b). Related methods are also described in Johnson and Bowler (2009), Pinson (2012) and Roulin and Vannitsem (2012). A recent article by Wilks (2014) compares the Schaake shuffle and the ECC in the context of ensemble post-processing. To the best of our knowledge, the shuffle technique has not yet been applied for the purpose of multivariate bias correction or downscaling of climate simulations.

The main objective of this article is to promote a technique that is readily available and easy to apply. This technique will be referred to as the "empirical copula - bias correction" (EC-BC) approach and combines a univariate BC method with the shuffling technique presented by Clark et al. (2004). We further provide an intercomparison of this method with a one-dimensional BC method and the conditional approach of Piani and Haerter (2012).

This article is organized as follows: In the next section, the data to be corrected and the reference data are first presented, as well as the experimental cross-validation setup. In section 3, a short description of the 1d-bias correction method used as a benchmark in this study is provided. Then, theoretical and technical details are given concerning the bivariate and multivariate bias correction methods compared in this article in Section 4, namely: the "conditional" technique, the shuffling-based method and the EC-BC approach. Section 5 contains the results of the intercomparison in terms of inter-variable, spatial and temporal analyses. Finally, general conclusions are given in section 6 as well as a discussion concerning the underlying assumptions and some potential adaptations of the various approaches.

2. Reference and model data

In this article, the reference data are daily temperature and precipitation time series from the SAFRAN reanalysis data (Quintana-Segui et al. 2008) over the south-west region of France $[2^oE, 7.5^oE] \times [42^oN, 45^oN]$ corresponding to 1506 continental grid-cells with an approximate 8×8 km spatial resolution. Fig. 1(a) displays the map of France with the region of interest in a box, as well as the mean cumulated annual precipitation (fig. 1(b)) and the mean daily temperature (fig. 1(c)). The SAFRAN dataset allows one to avoid gaps in the time series. It has been employed as a reference for evaluation of different statistical or dynamical downscaling approaches in various studies (e.g., Lavaysse et al. 2012; Vrac et al. 2012). A detailed description of SAFRAN, its validation and its application over France is given by Quintana-Segui et al. (2008).

Model data to be corrected are the ERA-Interim (hereafter ERA-I) daily reanalysis temperature and precipitation data with a 0.75^o by 0.75^o spatial resolution. Using an improved atmospheric model and assimilation system from those used in ERA-40 (Simmons and Gibson 2000), ERA-Interim represents a third generation reanalysis system (Dee et al. 2011). ERA-Interim reanalyses are now widely employed (e.g., Vautard and 25 authors 2013) and serve as meteorological forcing of the downscaling models involved in the "COordinated

Regional Downscaling EXperiment" (CORDEX) initiative[1].

For both model and reference datasets, data have been extracted from Jan., 1, 1980 to Dec., 31, 2009. Then, each ERA-I grid-cell has been Co-located with the SAFRAN grid-cell the closest to its center. Hence, each ERA-I grid-cell time series to be corrected has a unique reference SAFRAN grid-cell.

Moreover, in the following, all bias correction methods are applied separately to two periods of the year : from October 15th to April 14th (hereafter referred to as "winter") and from April 15th to October 14th (hereafter referred to as "summer").

The calibration of the following BC methods is performed over the period 1980-1994 and all evaluations are performed over the period 1995-2009.

3. Univariate bias correction

A variant of EQM has been recently developed by Michelangeli et al. (2009) and applied in many climate-related studies (e.g., Oettli et al. 2011; Colette et al. 2012; Tisseuil et al. 2012; Vrac et al. 2012; Vigaud et al. 2013, among others). This variant first estimates the distributions F_{Yp} and F_{Xp} for the random variables Y and X over the *projection* time period (either future, or simply evaluation time period) before applying a distribution-derived quantile mapping as defined in Eq. (2) in replacing X and Y by Xp and Yp respectively. If F_{Xp} can be directly modeled – parametrically or not – from the data to be corrected in the projection period, the modeling of F_{Yp} is based on the assumption that a mathematical

[1] http://wcrp-cordex.ipsl.jussieu.fr/

transformation T allows to go from F_X to F_Y in the calibration period:

$$T(F_X(z)) = F_Y(z) \tag{3}$$

for any z in the domain of X and Y; and that T is still valid in the projection period, i.e.:

$$T(F_{Xp}(z)) = F_{Yp}(z). \tag{4}$$

Replacing z by $F_X^{-1}(u)$ in (3), where u is any probability in $[0, 1]$, we obtain

$$T(u) = F_Y(F_X^{-1}(u)), \tag{5}$$

corresponding to a simple definition for T. Inserting (5) in (4) leads to a modeling of F_{Yp}:

$$F_{Yp}(z) = F_Y(F_X^{-1}(F_{Xp}(z))). \tag{6}$$

Once F_{Xp} and then F_{Yp} are modeled, a distribution-based quantile-mapping is applied as in (2). Hence, this so-called "Cumulative Distribution Function - transform" (CDFt) approach – as named by Michelangeli et al. (2009) – includes the information about the distributions over the projection time period in the quantile-mapping technique. Some more details about CDFt can be found in Vrac et al. (2012).

In the following, only the CDFt univariate bias correction approach will be applied. Indeed, preliminary analyses showed that EQM and CDFt display equivalent results in the context of the present study. Although this has not been tested, we strongly expect other univariate bias correction techniques (parametric or not, distribution-based or not) to behave relatively similarly. Hence, the univariate BC method CDFt is first applied independently to precipitation (PR) and 2m-temperature (T2) from ERA-I. This will provide the benchmark bias corrected ERA-I dataset to which some bivariate or multivariate correction procedures will be compared.

4. Bivariate / Multivariate bias correction

a. A short reminder on statistical dependence and copulas

The notion of (spatial, temporal or inter-variable) dependence structure is in close relationship with the so-called copula functions (e.g., Nelsen 2006). An introduction of the copula approach for climate research is given, e.g., in Schoelzel and Friederichs (2008). The basis of the copula approach is Sklar's theorem (Sklar 1959) which states that every multivariate or joint CDF can be expressed by the marginal CDFs of the univariate components of the multivariate random variable and the copula. The copula is a joint CDF that describes the statistical dependence of the transformed random variables $U_j = F_{X_j}(X_j)$, where X_j is the j-th component of the multivariate random variable $\mathbf{X} = (X_1, \ldots, X_d)^T$ and F_{X_j} the respective marginal CDF. Sklar's theorem states that every joint CDF $F_{\mathbf{X}}$ can be expressed as

$$F_{\mathbf{X}} = C_{\mathbf{X}}\left(F_{X_1}, \ldots, F_{X_d}\right), \tag{7}$$

where $C_{\mathbf{X}}$ is the copula of $\mathbf{X}$. Both bivariate and multivariate BC methods presented next are designed to restore the dependence structure and therefore the underlying copula function.

b. The bivariate "Conditional" approach

Piani and Haerter (2012) developed a bivariate BC method whose the main idea is to apply a univariate BC to precipitation time series conditionally on the bias corrected values of temperature classified into binned temperature values. This "conditional" approach works in three steps: First, a standard 1d-BC method is applied separately to model temperature.

Then, the (temperature, precipitation) pairs are grouped into temperature quantile bins. Finally, a standard 1d-BC method is applied for precipitation within each temperature bin. They concluded that this approach improved the 2d temperature-precipitation copula and that even a relatively small number of temperature bins allows to significantly improve the dependence structure (i.e., the copula) between the two physical variables. Technical details can be found in Piani and Haerter (2012). In the following, this conditional approach is applied both ways to our data: to bias correct precipitation time series conditionally on the bias corrected values of temperature, and to bias correct temperature time series conditionally on the bias corrected values of precipitation. For precipitation given temperature, five quantile bins have been used. Higher numbers of bins have also been tested but the quality of the results did not change significantly (not shown). For temperature given precipitation, only three quantile bins have been used (with the first interval bin including all zeros) to avoid the size of the bins to be too much different due to a larger number of dry days. Note that this 2D approach is relatively independent of the 1d-BC method since the conditional correction can be performed with most of the classical 1d-BC techniques. This is a very interesting feature that makes the procedure flexible.

However, this conditional approach reproduces only the 2D inter-variable dependences: we may want to correct the spatial and or temporal structures as well. Then, other techniques have to be employed.

c. The "Schaake Shuffle" method

Clark et al. (2004) highlighted another shuffling technique – sometimes named as the

"Schaake shuffle" after Dr. J. Schaake (National Weather Service Office of Hydrologic Development) – in the context of correcting forecasts from NWP models. This method was adapted by Schefzik et al. (2013b) and by Möller et al. (2012) in the context of ensemble postprocessing. Here, the Schaake approach is adapted and presented in the context of bias correction of time series generated by (global or regional) climate models – potentially previously dynamically or statistically downscaled – whose spatial, temporal and/or inter-variable properties have to be corrected.

The "Schaake shuffle" as illustrated in Table 1 is very simple to implement. Assume we have a reference sample of length 4 for the variable Z. The reference sample has a certain rank structure given by the rank k of an element in the sample with respect to the other data in the sample. When new samples arrive – e.g., from model output or from 1d-bias corrected data –, the main idea is to reorder the new samples such that their rank structure is identical to that of the reference sample. Let's take the example of the variable Z with reference sample $Z_R = (0.3, 0.5, 0.9, 0.8)$ and prediction sample (i.e., the sample data to be corrected) $Z_P = (0.7, 0.5, 0.2, 0.9)$. The associated ranks of Z_R are $k(Z_R) = (k(0.3) = 1,$ $k(0.5) = 2,$ $k(0.9) = 4,$ $k(0.8) = 3)$ – noted as $k(Z_R) = (1, 2, 4, 3)$ – and those of Z_P are $k(Z_P) = (k(0.7) = 3,$ $k(0.5) = 2,$ $k(0.2) = 1,$ $k(0.9) = 4)$ – noted as $k(Z_P) = (3, 2, 1, 4)$. The shuffling procedure consists in reordering the elements of Z_P into a new sample $Z_{shuffled}$ such that the rank of this new sample is identical to the rank of the training sample: $k(Z_{shuffled}) = k(Z_R) = (1, 2, 4, 3)$. Hence, based on the present example, the first element of $Z_{shuffled}$ must be the element of Z_P with rank 1, that is 0.2; the second element of $Z_{shuffled}$ must be the element of Z_P with rank 2, that is 0.5; the third element of $Z_{shuffled}$ must the element of Z_P with rank 4, that is 0.9; and the four element of $Z_{shuffled}$ must the element of Z_P with rank

3, that is 0.7. Therefore, $Z_{shuffled} = (0.2, 0.5, 0.9, 0.7)$ and satisfies $k(Z_{shuffled}) = k(Z_R)$. See Clark et al. (2004) for a more technical and mathematical formulation of the shuffling procedure. Note that Z_R represents the dataset from which the dependence structure is "learned". In our case it represents one time series in the SAFRAN reference dataset during the training period. Z_P represents the prediction, which in our study is the corresponding ERA-I time series, either bias corrected or not. The main difference between the shuffling methods mentioned in the introduction, namely the Schaake shuffle and the ECC, is the dataset that determines the dependence structure (i.e. the ranks).

In the present work, for practical reasons, the rank associated with exact same values (such as zeros for precipitation) is supposed to be increasing with time. In other words, if $z_{t_1} = z_{t_2} = 0$ are precipitation values at time t_1 and t_2 respectively, with $t_1 < t_2$, then $rank(z_{t_1}) < rank(z_{t_2})$. In the context of a 3-dimensional data matrix (say, n time steps, s gridcells or stations, p physical variables), the Schaake method is applied separately to the n-component vector resulting from each combination "one gridcell $\times$ one variable" (i.e., it is applied $s \times p$ times). The remarkable effect is that simply by reordering the data independently in time, not only the temporal, but also intervariable and spatial dependencies are restored. How powerful the "Schaake shuffle" is will be shown in section 5.

Why is Sklar's theorem (Eq. 7) of relevance for the shuffling method presented here? An important property of the transformed random variables U_j is that if Z_j has the CDF F_{Z_j} then $U_j = F_{Z_j}(Z_j) \sim Unif(0,1)$, i.e. U_j are uniformly distributed on the interval $[0,1]$. Lets now assume we have a sample $z_j^{(i)}, i = 1, \ldots, N$ of Z_j without knowing F_{Z_j}, then $u_j^{(i)} = F_{Z_j}(z_j^{(i)})$ is generally estimated as the rank $k_j^{(i)}$ of $z_j^{(i)}$ with respect to the sample $z_j^{(i)}, i = 1, \ldots, N$ divided by $N+1$, i.e. $\hat{u}_j^{(i)} = k_j^{(i)}/(N+1)$. Hence, re-shuffling of the multivariate data with respect to their ranks $k_j^{(i)}$ has the potential to restore (parts of) the dependence structure, namely the copula $C_{\mathbf{Z}}$. It is the same reason why rank correlation is an adequate measure to assess dependence between random variables. An important consequence of Sklar's theorem (Eq. 7) is that the BC of the marginals and the restoration of the dependence structure can be performed independently, at least as long as the BC of the marginals does not affect the ranks of the data (this is generally given since transfer functions are usually monotonic functions). In the following, the application of the Schaake shuffling technique to previously 1d-bias corrected time series will be referred to as "Empirical Copula - Bias Correction" (EC-BC).

d. Raw and shuffled ERA-I reanalyses

For comparison purposes, the "raw" ERA-I data (i.e., without any correction) as well as the Schaake shuffling technique are directly applied to ERA-I without any preliminary univariate bias correction are also evaluated. Hence, in the "Results" section, the following BC methods are intercompared:

- the **independent univariate bias corrections** (CDFt) of the ERA-I reanalyses of precipitation and temperature;
- CDFt on ERA-I followed by the Schaake shuffle method, i.e., **the EC-BC approach**;
- the **conditional approach** based on CDFt on ERA-I (with precipitation corrected conditionally on temperature and the other way around);
- the **raw ERA-I data** (i.e., without any correction);

- and the **Schaake shuffling technique directly applied to ERA-I** without any preliminary univariate bias correction.

5. Results

The various BC methods are evaluated according to three different angles: How do the corrected data reproduce the inter-variable statistical properties? How do they reproduce the spatial properties? How do they reproduce the temporal properties? In the following, due to the large number of figures available, only winter evaluations are shown. However, summer plots are fairly equivalent or provide equivalent conclusions and are provided as auxiliary material.

a. Inter-variable correlations

For many impact models (e.g., hydrology, agriculture), the correlation between variables – here precipitation and temperature – is an important feature that must be accurately modeled by the meteorological input data. Hence, fig. 2 shows maps of inter-variable Spearman correlation coefficients between PR and T in winter over the evaluation period for the various BC models as well as for the SAFRAN dataset. While the Pearson correlation coefficient is the most widely used, the Spearman correlation is employed here. Indeed, the Pearson coefficient measures the strength of the linear relationship between normally distributed variables. However, precipitation is not normally distributed and, besides, the relationship between temperature and precipitation is not supposed to be linear. Hence, in that context, it is more appropriate to use the Spearman correlation that does not require a linear relationship, neither the variables to be normally distributed (e.g., Hauke and Kossowski 2011). In fig. 2, only correlations that are statistically equivalent to the SAFRAN correlation (i.e., not significantly different at 95%) are shown in those plots. A bootstrap technique (Efron and Tibshirani 1993) with block-replacement of 10-day blocks has been applied to determine if the correlations were significantly different or not at 95%. The procedure was the following for each grid-cell: (i) Take the N daily observations in the verification period; (ii) Generate 1000 times N-day long bootstrapped samples with replacement (i.e., each sample is constituted of ($N/10$) 10-day blocks); (iii) Compute the 2.5% and 97.5% percentiles from the 1000 correlations as the 95% uncertainty interval: if the correlations of the BC data are outside this range, they are considered as significantly different. The length of the blocks (10 days) has been chosen to account for temporal correlations, i.e., that the effective number of degrees of freedom in the daily time series is significantly smaller than N. In each panel of fig. 2, the percentage of grid-points with correlation significantly different (%GPCSD) from that of SAFRAN is also indicated. As expected, ERA-I correlations appear clearly as inappropriate (%GPCSD is more than 66%). This is true also for the correlations from the univariate BC method that roughly reproduce the ERA-I pattern (%GPCSD$\simeq$61%). Interestingly, the "conditional" approach does not give the same correlations when applied to correct temperature given the precipitation (fig. 2(d)) or to correct precipitation given the temperature (fig. 2(e)): the former provides much better correlations in the present setting (about 30% vs. 75% for the %GPCSD). One explanation is that while a given PR interval provides useful constraints on the possible range of associated temperatures, the opposite is not true: temperature is not a good predictor of precipitation that remains relatively highly variable even for a given small interval of temperatures. The EC-BC method generates equivalently good results in terms of inter-variable dependence and provides satisfactory correlations (%GPCSD≃27%). This is true also when the Schaake shuffle is applied directly to ERA-I reanalyses (%GPCSD≃38%). It is interesting to note that some "not-significantly equivalent correlations" regions are different from one model to another. Some additional analyses and experiments (not shown) illustrate that the EC-BC method is not sensitive to the choice of the univariate BC method (CFt or EQM) as preliminary step. This is not exactly the case for the conditional approach where some differences appear between "Cond. CDFt" and "Cond. EQM" (not shown) and one must be cautious to this point when applying the bivariate conditional approach.

b. Spatial correlations

The statistical spatial properties are also very important in many impact studies. A very common way to investigate spatially coherent variability is a principal component analysis (PCA). It has first to be noted, that the dominant empirical orthogonal function (EOF) for both temperature and precipitation represents almost constant changes over the entire region. This is due to the small spatial extent of the region, where day to day weather variability is large and affects the whole domain in a very similar way. We thus first investigate the variability of the area-mean temperature and precipitation times series, which is then removed from the data for the PCA.

Figure 3 represents bivariate histograms of area-mean 2m temperatures. We here consider the complete verification period taking summer and winter data together. In order to also show equivalent figures for the reference data, we generated a perturbed series of observed area-mean temperatures by randomly changing the order of the years while preserving the order of the day in the year. The reference data reveal a distinct seasonal cycle with an amplitude of more than 15 K. The seasonal cycle seems well reproduced in the bias corrected temperature, whereas it is largely underestimated in ERA-Interim (Fig. 3 (b) and (f)). Thus, univariate BC is helpful to correct the amplitude of the seasonal cycle. EC-BC or conditional BC seem not to significantly improve the distribution of the area-mean values.

Since the distribution of precipitation is highly skewed, we set the zero precipitation values to a small value different from zero (0.00033) and work in the following on the logarithm of precipitation. For area-mean precipitation (Fig. 4) no obvious seasonal cycle exists. The effect of the Schaake shuffle (e.g., comparing Figs. 4 (b) and (f)) seems to concentrate the area-mean precipitation values, presumably because the Schaake shuffle increases the spatial variability of ERA-Interim precipitation (i.e., induces more small scale structures by shuffling). The conditional BC seems to shift the modus of the precipitation values to lower values. None of the area-mean precipitation series seems superior from this analysis.

For the PCA we now removed the area-mean from the data at each time step. We concentrate on winter data, but the results are similar for summer. Figs. 5 and 6 show the eigenvalues and explained variance fractions of the leading EOF for temperature and log-precipitation, respectively. Zero precipitation values were again set to a small value of 0.00033. Note that a principal component analysis for the still non-Gaussian log-precipitation fields should be interpreted with caution. We think, however, that in our case it is a valuable tool to compare spatially coherent modes of variability.

The eigenvalue spectra for temperature in Fig. 5 show, that the total variance, i.e. the sum of the eigenvalues, is generally largest for the reference data, and smallest for ERA-Interim, either shuffled or not. Thus, one important effect of BC is to correct for total variance of the data. The conditional BC approach has a realistic variance spectrum, whereas the EC-BC provides an eigenvalue spectrum very close to that of the reference data. The explained variance spectra in Fig. 5 in turn give an indication of the relative importance of the leading EOF. A flat spectrum indicates weak coherence in the spatial patterns, whereas a steep spectrum generally indicates the presence of large-scale coherent structures. Since independent BC inherits the spatial dependence of ERA-Interim, they both have a very dominant first EOF. The explained variance spectra for the conditional and the EC-BC approaches are very realistic.

Similar results are obtained for precipitation (Fig. 6). The total variance of ERA-Interim and the shuffled ERA-Interim data is much too small, whereas BC has a very positive effect even for the independent BC. The conditional BC seems to underestimate the variance of the first EOF. The explained variance spectra show only small differences. Precipitation generally has much more small-scale variability which is reflected in the small explained variance fraction of the leading EOF. ERA-Interim and independently bias-corrected ERA-Interim exhibit slightly larger scale dominant patterns.

The differences become even more evident in the structure of the leading EOFs. The leading EOF for temperature (Fig. 7) in the reference data represents a dipole pattern with higher than normal temperatures near the Mediterranean coast and colder temperatures in the northern and north-eastern part of the region. All BC methods except those that apply the Schaake shuffle, reproduce the checked pattern imposed by the ERA-Interim grid structure and an east-west dipole. The conditional approach only slightly modifies the large-scale pattern. This effect also pervades higher order EOF (not shown). In contrast, the EC-BC has a very realistic leading EOF, and albeit with a smaller amplitude, the first EOF is also well reproduced in the shuffled ERA-I dataset.

For log-precipitation (Fig. 8) the results are similar. Again, the leading EOF of the EC-BC data set is very close to that of the reference data. The conditional BC introduces some noise, but besides this its first EOF is very close to the first EOF of ERA-I. Note that the conditional approach has been applied here to model temperature conditionally on precipitation (Figs. 5 and 7) or the other way around (Figs. 6 and 8), i.e., in an inter-variable context and not a spatial one. One can expect this conditional technique to work better if applied in a spatial one, e.g., if the station i is modeled according to the station j. Nevertheless, one could get as many references as stations j. Hence, the correction is then not unique and therefore may be quite complicated to interpret. Besides, the combinatory of BC to be applied can quickly increase and make the practical implementation intractable.

Globally, EC-BC shows the most satisfying spatial variance pattern, whenever designed with CDFt or EQM (not shown for EQM). The results are also satisfactory – to a lesser extent – for the Schaake shuffle directly applied to the raw ERA-I data. In order to assess the similarity of spatial variance patterns more objectively, we performed a reduction of spatial degrees of freedom. To this end, we calculate the EOF of the reference data and project all data onto the leading 10 EOF of the reference data. We thus obtain 10 times series (i.e. expansion coefficients) for each dataset. The analysis is now performed within the 10 dimensional subspace spanned by the 10 leading EOF.

We first examine the covariance matrices of the reduced data sets for 2m temperature (Fig. 9). By construction, the expansion coefficients of the reference data show a diagonal covariance matrix. The covariances between the expansion coefficients are zero since the eigenvectors of the covariance matrix are statistically orthogonal. This is not anymore the case for the other datasets. Here, the covariances between the expansion coefficients are generally non-zero. The degree to which the off-diagonal are different from zero indicates how different the respective variation patterns are. EC-BC seems to project very well on the EOF of the reference data, all other methods show substantial differences. For precipitation (Fig. 10) results are similar. The similarity of the covariance matrix obtained from EC-BC with that of the reference data is again striking.

We finally want to quantify the quality of each of the approaches by using a distance function between the empirical (multivariate) distribution of the reference data and each of the BC methods. As distance measure we use the integrated quadratic distance (IQD), which is a proper divergence function Thorarinsdottir et al. (2013a). It measures the distance between two distribution functions. The IQD between two distribution functions F and G is defined as the integral

$$d(F, G) = \int_{\Omega} (F(\omega) - G(\omega))^2 d\omega, \tag{8}$$

where Ω represents the sample space. The IQD is closely related to the energy score used in forecast verification (Gneiting and Raftery 2007). It may be empirically estimated using the equivalent formulation

$$d(F, G) = E||\mathbf{X} - \mathbf{Y}|| - \frac{1}{2}E||\mathbf{X} - \mathbf{X}'|| - \frac{1}{2}E||\mathbf{Y} - \mathbf{Y}'||, \tag{9}$$

where $\mathbf{X}$ and $\mathbf{X}'$, and $\mathbf{Y}$ and $\mathbf{Y}'$ represent independent draws from multivariate distribution functions F and G, respectively. The vector norm used here is the Euclidian norm.

In our application, $\mathbf{X}$ and $\mathbf{Y}$ are the expansion coefficients of the reference and the BC data, respectively. To get independent random realizations of the differences we randomly draw 50.000 vectors with replacement for $\mathbf{X}$, $\mathbf{X}'$, $\mathbf{Y}$ and $\mathbf{Y}'$ out of the available data sets of length 2.734 winter days, respectively, and calculate the IQD using (9). In order to assess the uncertainty of the IQD, we additionally apply a bootstrap method with replacement (Efron and Tibshirani 1993). Repeating this 200 times provides estimates of the uncertainty of the IQD. Fig. 11 shows the IQD estimates together with the 95% bootstrap sampling uncertainty. The IQD in Fig. 11 is evaluated hierarchically, first in the subspace of the leading, then the first 2 leading upto the first 10 leading EOF of the reference data.

The IQD quantitatively confirms the superiority of EC-BC. For 2m temperature (Fig. 11 (a)) the IQD for the EC-BC data varies closely above zero throughout the hierarchy. It only slightly increases with a higher dimensionality. There is a rather clear ranking between the different approaches, with EC-BC performing best, conditional BC second best when using more that 2 EOF, followed by ERA-Interim with Schaake shuffle, and independent BC. The raw ERA-Interim data have the largest IQD, so any approach provides improvements in terms of IQD. For temperature, large improvements of the spatial covariances are obtained solely by the Schaake shuffle. Its effect on the IQD is stronger that that of the independent BC of the marginals. In comparison to independent BC, the conditional BC only slightly improves the spatial covariances.

For precipitation (Fig. 11 (b)) again EC-BC is clearly superior, but the ranking is not the same as for temperature and less distinct. EC-BC performs best, followed by independent BC. Interestingly, the Schaake shuffle applied without BC seems to degrade the IQD. The most important correction here is the BC of the marginals, whereas the correction for the dependence structure is less important for precipitation. The conditional approach seems to work less well for precipitation under this respect.

c. Temporal correlations

We finally investigate the temporal structure of the time series, which used as input in impact models may also have great consequences. Its accurate modeling may then be crucial.

To this end, n-day lag autocorrelations have been studied for n between 1 and 5. Figures 12 and 13 display the lag-1 auto-correlations for the different BC models in winter for temperature and precipitation respectively.

For temperature, the conditional approach (Fig. 12(d)) clearly underestimates lag-1 auto-correlations, while, in that temporal context, the independent BC (Fig. 12(c)) gives relatively consistent results, although strongly imperfect due to the structure in "squares" already present in non-corrected ERA-I auto-correlations (Fig. 12(b)). The shuffling procedure provides the best temporal dependencies either applied to CDFt results (i.e., the EC-BC approach, Fig. 12(e)) or directly to ERA-I data (Fig. 12(f)). Most of the conclusions from lag-1 temperature auto-correlation are still valid for lag-5 auto-correlations (not shown): The results of the shuffling procedure (on 1d-BC or non-corrected data) are still very close to the reference, while the conditional approach provides too low correlations and ERA-I data continue to have too high correlations. However, the independent BC method is not as consistent as for lag-1 results, with too low lag-n auto-correlations for $n \geq 2$.

In terms of precipitation, (Fig. 13), contrary to temperature, lag-1 correlations from independent BC (Fig. 13(c)) are not acceptable, showing a pronounced underestimation. On the opposite, the direct shuffling of ERA-I (Fig. 13(f)) globally overestimates the 1-day auto-correlation, especially on the North-East part of the domain. This was already true (with a smaller magnitude) for uncorrected ERA-I (Fig. 13(b)). The EC-BC approach (Fig. 13(e)) provides the precipitation lag-1 auto-correlation structures and intensities the closest to those of the SAFRAN dataset (Fig. 13(a)), while the conditional approach (Fig. 13(d)) gives correct auto-correlation magnitudes but with relatively inappropriate structures. For lags longer than one day, the precipitation auto-correlation drops very quickly close to zero for observations and almost all models (not shown), except for the direct shuffling of ERA-I data that continues to provide very high (unobserved and unrealistic) auto-correlations of about 0.8 – at least until a 5-day lag – for the north-east region. This is somehow surprising since the "raw" ERA-I data (i.e., without any bias correction) do not show such a strong feature although with a very slight overestimation of the auto-correlation for this region.

Moreover, to describe more specifically the rainfall occurrence temporal structure obtained from the BC methods, the maps of the probability of a dry day given that the previous day was wet – i.e., $Proba(dry|wet)$ noted as Pdw –, as well as the opposite – i.e., $Proba(wet|dry)$ noted as Pwd – have been computed and are displayed in Figs. 14 and 15 respectively. For the maps of $Proba(dry|wet)$, it is quickly seen that ERA-I (Fig. 14(b)) and the conditional approach (Fig. 14(d)) globally overestimate the probability of a dry day given that the previous day was wet. However, all the other BC methods provide satisfying Pdw values, close to those of SAFRAN.

Interestingly, the Pwd maps (Fig. 15) are not completely the "symmetric" of the Pdw maps. Here, the conditional approach (Fig. 15(d)), ERA-I (Fig. 15(b)), as well as its direct shuffling

(Fig. 15(f)) underestimate the dry day probabilities (particularly strongly for the latter two datasets). The independent BC (Fig. 15(c)) shows better Pwd values, although too high in the north-east region. However, the Pwd results the closest to those of the reference dataset are obtained from the EC-BC model (Fig. 15(e)), which shows quite similar values and spatial structures.

6. Conclusions and Discussion

a. Conclusions

In this paper, we have compared several univariate, bivariate and multivariate bias correction (BC) methods designed for specific multivariate properties:

- One univariate "independent BC" based on the CDFt approach;
- The "Conditional approach" (Piani and Haerter 2012) (here, based on CDFt) developed specifically for producing a correct two-dimensional inter-variable structure;
- The 'Schaake Shuffle"method (Clark et al. 2004) applied directly to raw (i.e., uncorrected) ERA-I reanalyses precipitation and temperature time series;
- The "Empirical Copula - Bias Correction" (EC-BC) approach constituted with the Schaake Shuffle method applied to previously 1d-bias corrected time series (here, through the CDFt method) of precipitation and temperature.

The Schaake method is based on temporal shuffling of the elements in each time series such that the temporal rank structure is reconstructed.

Globally, on those datasets and with this experimental setting, although it is quite useful for correction of the marginal distributions, the one-dimension CDFt bias correction alone is not good at reproducing any of the inter-variable, spatial or temporal properties of the observed data. This is true also for the univariate EQM method (not shown). In contrast, the application of the EC-BC techniques clearly improves those properties. The conditional and the shuffling methods improve the inter-variable properties – often, even when applied directly to ERA-I data. This is not the case for the spatial structure where the conditional technique – which was initially designed only for inter-variable structures – is not suitable, whereas the EC-BC approach is quite efficient in general. This inappropriateness of the conditional method is also visible in the temporal properties where auto-correlations are underestimated. Again, in this temporal context, the EC-BC technique is relatively satisfying for both temperature and precipitation.

As global conclusions:

- The one-dimensional BC method CDFt is not able to produce correct multidimensional properties (similar results were obtained with the EQM method, not shown);
- The conditional technique – at least as applied in this experimental setup – is only good for inter-variable properties reproduction;
- The EC-BC approach is good for both, inter-variable, spatial and temporal correlations. The preliminary of 1d-BC before the shuffling procedure is nevertheless an important requisite for precipitation since the combination "1d BC/shuffling" generally provides the most satisfying results;
- Due to its easiness of coding, its speed of application and the good quality of its results

for both inter-variable, spatial and temporal properties, the Schaake Shuffle method applied after a 1d BC method (i.e., the EC-BC approach) is a very good candidate for all needs in multivariate bias correction.

Although not tested, the application of these BC methods to correct GCM outputs instead of reanalysis data is expected to slightly degrade the results but to produce equivalent rankings: the simpler methods should perform worse when based on GCM data due to the GCM weather sequence that generally needs additional corrections, while the EC-BC approach should continue to work well. More precisely, the 1d methods (CDF-t and EQM) will basically reproduce the inter-variable, spatial and temporal properties of the input data. So if those properties are wrong from the GCMs, they will be wrong as well for the 1d corrected data. The conditional approach, by construction, should work fine to reconstruct an inter-variable dependence close to that of the observations, even when driven by GCM outputs. However, the spatial and temporal properties of the data corrected in following this approach should stay relatively close to those of the GCM data. Moreover, although the use of the Schaake shuffle directly to the GCM simulations should improve those, it is expected that EC-BC will provide the best results in terms of the three types of properties studied in this paper. Hence, by construction of the EC-BC approach, results similar to those presented in this article can be expected on different regions or with different reference or model datasets.

b. Discussion

The general idea of the EC-BC and shuffling methods presented here is to re-shuffle the predictive multivariate spatio-temporal data according to some rank structure derived from training data. In doing so the data in the evaluation set receives a dependence structure close to the dependence structure of the training dataset. More concretely, lets assume we have training and test datasets, each with a multivariate spatio-temporal structure. With the Schaake shuffle method, simply by shuffling the test data set in time such that the ranks of the data in time is identical to that of the training data, we restore at least partly the inter-variable, spatial and temporal dependencies of the training data set. Since the univariate BC as presented above is a monotonic transformation of the data and is applied to each variable and point in space independently, it has no influence on the copula function. Shuffling can be performed prior or after univariate BC.

Note also that, if the CDFt method has been employed as univariate BC within the EC-BC approach, other techniques can be used. This shuffling post-processing can be performed based on most of the standard 1d-BC techniques. This interesting feature makes the procedure flexible and easily applicable. Note that the shuffling can even be applied to most (if not all) of the 1d statistical downscaling (SD) approaches. This application of the Schaake shuffle to 1d-SD outputs should improve their temporal, spatial and inter-variable properties as much as it has been shown for the BC methods in the present article. Therefore, it would be interesting to compare such a multivariate SD based on shuffling post-processing to statistical downscaling models taking explicitly into account the multi-dimensional structure of the data to be downscaled (e.g., Yang et al. 2005; Flecher et al. 2010; Vrac et al. 2007, for multi-site, multi-variable and temporal dependences modeling respectively).

Moreover, other re-ordering of data might be applied to restore and preserve some specific structures. The Schaake method shuffles elements *in time*. In other words, one value associated to a given location (grid-cell or station) stays associated to this location but is placed at another time. However, one may want to allow shuffling values *both in time and in space*. This could improve the reproduction of the spatial dependences. To do so, it is easy to extend the Schaake approach: instead of computing ranks and shuffling values within vectors, this is made within 2-dimensional matrices. Hence, one value initially associated to a given location at a given time may be placed at another time and another location after this "full" shuffling. This technique has been tested and the results (not shown) both in terms of inter-variable, spatial and temporal properties are very similar to those of the Schaake shuffling presented all along the present study, except for precipitation where this full Schaake shuffle applied directly to ERA-I is not as efficient as the "regular" Schaake shuffling. Note that this full shuffling could also be performed for different physical variables at once. If the variables have the same units (e.g., all variables are temperature values), this can make sense. However, if the variables are different (e.g., precipitation and temperature), the shuffling of values between the two variables can be strongly inappropriate and quite difficult to interpret afterward.

Finally, there are essential assumptions to BC and EC-BC. Univariate BC estimates a transfer function (TF) between model and observations from the training data, and applies this TF to the evaluation (or projection) dataset. The main assumption is that the relation between model and observations remains unchanged during the projection period. However, if the distribution of the model data changes in the projection period so does the distribution of the projected values. EC-BC (through the Schaake shuffle) represents a method to restore the dependence structure within the projected values, which is inherited from the dependence structure of the observations in the training dataset. The dependence structure of the model data is completely ignored. This is absolutely reasonable in our context of downscaling, since we know that the dependence structure in the large-scale model is erroneous. However, EC-BC also ignores potential changes in the dependence structure suggested by the model data. This is an important assumption: the (spatial, temporal and/or inter-variable) dependence structures do not change between the training period and the projection period. Although this conservative assumption is reasonable and simplifies the bias corrections, it may not be valid in a climate change context where the multivariate properties to be corrected may evolve as well. Hence, if changes in the dependence properties or its temporal evolutions are of interest, the development of models allowing to make the dependence structures change in time or in function of some atmospheric covariates would be of great interest for both the climate and impacts communities.

Acknowledgments.

This work has been partially supported by the ANR-project StaRMIP, the GICC-project REMedHE, the VW-project PLEIADES and the ANR-project REMEMBER. All computations have been made in R. An R package containing functions for the shuffling procedure as

well as for the EC-BC approach (with CDFt) should soon be made available on the CRAN website[2] or upon request to the authors.

[2]http://cran.r-project.org/

REFERENCES

Chen, C., J. Haerter, S. Hagemann, and C. Piani, 2011: On the contribution of statistical bias correction to the uncertainty in the projected hydrological cycle. *Geophys. Res. Lett.*, **38, L20403**, doi:10.1029/2011GL049 318.

Christensen, J., F. Boberg, O. Christensen, and P. Lucas-Picher, 2008: On the need for bias correction of regional climate change projections of temperature and precipitation. *Geophysical Research Letters*, **35 (20)**.

Clark, M., S. Gangopadhyay, L. Hay, B. Rajagopalan, and R. Wilby, 2004: The Schaake shuffle: A method for reconstructing space-time variability in forecasted precipitation and temperature fields. *J. Hydrometeor.*, **5**, 243–262.

Colette, A., R. Vautard, and M. Vrac, 2012: Regional climate downscaling with prior statistical correction of the global climate forcing. *Geophysical Research Letters*, **39**, L13 707, doi:10.1029/2012GL052 258.

Dee, D. P., et al., 2011: The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q.J.R. Meteorol. Soc.*, **137**, 553–597. doi: 10.1002/qj.828.

Déqué, M., 2007: Frequency of precipitation and temperature extremes over France in an anthropogenic scenario: Model results and statistical correction according to observed values. *Global Planet. Change*, **57**, 16 – 26.

Efron, B. and R. J. Tibshirani, 1993: *An Introduction to the Bootstrap*. Chapman & Hall, 436 pp.

Flecher, C., P. Naveau, D. Allard, and N. Brisson, 2010: A stochastic daily weather generator for skewed data. *Water Resour. Res.*, **46 (W07519, doi:10.1029/2009WR008098)**.

Gneiting, T. and A. E. Raftery, 2007: Strictly proper scoring rules, prediction, and estimation. *J. Amer. Stat. Assoc.*, **102 (477)**, 359–378, doi:10.1198/016214506000001437.

Gudmundsson, L., J. B. Bremnes, J. E. Haugen, and T. Engen-Skaugen, 2012: Technical Note: Downscaling RCM precipitation to the station scale using statistical transformations – a comparison of methods. *Hydrology and Earth System Sciences*, **16 (9)**, 3383–3390, doi: 10.5194/hess-16-3383-2012.

Haddad, Z. and D. Rosenfeld, 1997: Optimality of empirical z-r relations. *Q. J. R. Meteorol. Soc.*, **123**, 1283–1293.

Hauke, J. and T. Kossowski, 2011: Comparison of values of pearson's and spearman's correlation coefficients on the same sets of data. *Quaestiones geographicae*, **30 (2)**, 87–93.

Johnson, C. and N. Bowler, 2009: On the reliability and calibration of ensemble forecasts. *Mon. Wea. Rev.*, **137**, 1717–1720.

Lavaysse, C., M. Vrac, P. Drobinski, M. Lengaigne, and T. Vischel, 2012: Statistical downscaling of the French Mediterranean climate: assessment for present and projection in an anthropogenic scenario. *Nat. Hazards Earth Syst. Sci.*, **12**, 651–670, doi:10.5194/nhess–12–651–2012.

Maraun, D., 2013: Bias correction, quantile mapping, and downscaling: Revisiting the inflation issue. *Journal of Climate*, **26**, 2137–2143. doi: http://dx.doi.org/10.1175/JCLI–D–12–00 821.1.

Meehl, G., 2007: *Global Climate Projections. In: Climate Change 2007: The physical basis. Contribution of Working Group 1 to the fourth Assessment report of the Intergovernmental Panel on Climate Change*. Solomon et al., Cambridge University Press, Cambridge, UK.

Michelangeli, P., M. Vrac, and H. Loukos, 2009: Probabilistic downscaling approaches: application to wind cumulative distribution functions. *Geophys. Res. Lett.*, **36**, L11 708, doi:10.1029/2009GL038 401.

Möller, A., A. Lenkoski, and T. L. Thorarinsdottir, 2012: Multivariate probabilistic forecasting using ensemble Bayesian model averaging and copulas. *Q. J. R. Meteorol. Soc.*, doi:10.1002/qj.2009.

Muerth, M., et al., 2013: On the need for bias correction in regional climate scenarios to assess climate change impacts on river runoff. *Hydrol. Earth Syst. Sci.*, **17**, 1189–1204, doi:10.5194/hess–17–1189–2013.

Nelsen, R. B., 2006: *An Introduction to Copulas, 2nd edition*. Springer, New-York, 269 pp.

Oettli, P., B. Sultan, C. Baron, and M. Vrac, 2011: Are regional climate models relevant for crop yield prediction in West Africa? *Environ. Res. Lett*, **6**, doi: 10.1088/1748–9326/6/1/014 008.

Panofsky, H. and G. Brier, 1958: Some applications of statistics to meteorology. Tech. rep., University Park, Penn. State Univ., 224 pp.

Piani, C., J. Haerter, and E. Coppola, 2010: Statistical bias correction for daily precipitation in regional climate models over Europe. *Theoretical and Applied Climatology*, **99**, 187–192, doi:10.1007/s00704-009-0134-9.

Piani, C. and J. O. Haerter, 2012: Two dimensional bias correction of temperature and precipitation copulas in climate models. *Geophys. Res. Lett.*, doi:10.1029/2012GL053839.

Piani, C., G. Weedon, M. Best, S. Gomes, P. Viterbo, S. Hagemann, and J. Haerter, 2010b: Statistical bias correction of global simulated daily precipitation and temperature for the application of hydrological models. *J. Hydrol.*, **395**, 199–215, doi:10.1016/j.jhydrol.2010.10.024.

Pinson, P., 2012: Adaptive calibration of (u, v)-wind ensemble forecasts. *Q. J. R. Meteorol. Soc.*, **138**, 1273–1284.

Quintana-Segui, P., et al., 2008: Analysis of near surface atmospheric variables: validation of the SAFRAN analysis over france. *J. Appl. Meteorol. and Climatol.*, **47**, 92–107.

Robertson, D. E., D. L. Shrestha, and Q. J. Wang, 2013: Post-processing rainfall forecasts from numerical weather prediction models for short-term streamflow forecasting. *Hydrology and Earth System Sciences*, **17 (9)**, 3587–3603, doi:10.5194/hess–17–3587–2013.

Roulin, E. and S. Vannitsem, 2012: Postprocessing of ensemble precipitation predictions with extended logistic regression based on hindcasts. *Mon. Wea. Rev.*, **140**, 874–888.

Schaake, J., et al., 2007: Precipitation and temperature ensemble forecasts from single-valued forecasts. *Hydrology and Earth System Sciences Discussions*, **4**, 655–717.

Schefzik, R., T. L. Thorarinsdottir, and T. Gneiting, 2013a: Uncertainty quantification in complex simulation models using ensemble copula coupling. *Statistical Science*, **28 (4)**, 616–640.

Schefzik, R., T. L. Thorarinsdottir, and T. Gneiting, 2013b: Uncertainty quantification in complex simulation models using ensemble copula coupling. *Statistical Science*, **28**, 616–640.

Schoelzel, C. and P. Friederichs, 2008: Multivariate non-normally distributed random variables in climate research - introduction to the copula approach. *Nonlin. Processes Geophys.*, **15**, 761–772.

Schuhen, N., T. L. Thorarinsdottir, and T. Gneiting, 2012: Ensemble model output statistics for wind vectors. *Mon. Wea. Rev.*, **140**, 3204–3219.

Simmons, A. J. and J. K. Gibson, 2000: The ERA-40 project plan. Era- 40 project rep. series 1, ECMWF, Shinfield Park, Reading, United Kingdom.

Sklar, A., 1959: Fonctions de répartition à n dimensions et leurs marges. Tech. Rep. 229-231, Publ. Inst. Statist. Univ. Paris 8.

Thorarinsdottir, T. L., T. Gneiting, and N. Gissibl, 2013a: Calibration diagnostics for point process models via the probability integral transform. *SIAM/ASA Journal on Uncertainty Quantification*, **1**, 150–158, doi:10.1137/130907550.

Thorarinsdottir, T. L., M. Scheuerer, and C. Heinz, 2013b: Assessing the calibration of high-dimensional ensemble forecasts using rank histograms. *arXiv:1310.0236*.

Tisseuil, C., M. Vrac, G. Grenouillet, M. Gevrey, T. Oberdorff, A. Wade, and S. Lek, 2012: Strengthening the link between hydro-climatic downscaling and species distribution modelling: Climate change impacts on freshwater biodiversity. *Science of the Total Environment (STOTEN)*, **424**, 193–201, doi: 10.1016/j.scitotenv.2012.02.035.

Vautard, R. and 25 authors, 2013: The simulation of European heat waves from an ensemble of regional climate models within the EURO-CORDEX project. *Climate Dynamics*, DOI 10.1007/s00 382–013–1714–z.

Verkade, J., J. Brown, A. Weerts, and A. H. Reggiani, P. Weerts, 2013: Post-processing ECMWF precipitation and temperature ensemble reforecasts for operational hydrologic forecasting at various spatial scales. *Journal of Hydrology*, **501**, 73–91, doi:10.1016/j.jhydrol.2013.07.039.

Vigaud, N., M. Vrac, and Y. Caballero, 2013: Probabilistic downscaling of GCM scenarios over southern India. *International Journal of Climatology*, **33 (5)**, 1248–1263, DOI: 10.1002/joc.3509.

Voisin, N., F. Pappenberger, D. P. Lettenmaier, R. Buizza, and J. C. Schaake, 2011: Application of a medium-range global hydrologic probabilistic forecast scheme to the ohio river basin. *Wea. Forecasting*, **26**, 425–446.

Voisin, N., J. C. Schaake, and D. P. Lettenmaier, 2010: Calibration and downscaling methods for quantitative ensemble precipitation forecasts. *Wea. Forecasting*, **25**, 1603–1627.

Vrac, M., P. Drobinski, A. Merlo, M. Herrmann, C. Lavaysse, L. Li, and S. Somot, 2012: Dynamical and statistical downscaling of the French Mediterranean climate: uncertainty assessment. *Nat. Hazards Earth Syst. Sci.*, **12**, 2769–2784, doi:10.5194/nhess–12–2769–2012.

Vrac, M., M. Stein, and K. Hayhoe, 2007: Statistical downscaling of precipitation through nonhomogeneous stochastic weather typing. *Clim. Res.*, **34**, 169–184.

White, R. and R. Toumi, 2013: The limitations of bias correcting regional climate model inputs. *Geophysical Research Letters*, **12 (40)**, 2907–2912, DOI: 10.1002/grl.50 612.

Wilks, D. S., 2014: Multivariate ensemble-MOS using empirical copula. *Submitted to "Quart. J. Roy. Meteor. Soc."*.

Wood, A., L. Leung, V. Sridhar, and D. Lettenmaier, 2004: Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs. *Clim. Change*, **62 (189–216)**.

Yang, C., R. E. Chandler, V. S. Isham, and H. S. Wheater, 2005: Spatialtemporal rainfall simulation using generalized linear models. *Water Resour. Res.*, **41 (W11415, doi:10.1029/2004WR003739)**.

List of Tables

Table 1. Reference data of sample size 4 for the illustration of the "Schaake shuffle". $k()$ indicates the rank within the sample.

Training				Prediction				Schaake Shuffle			
$x_T^{(i)}$	$k(x_T^{(i)})$	$y_T^{(i)}$	$k(y_T^{(i)})$	$x_P^{(i)}$	$k(x_P^{(i)})$	$y_P^{(i)}$	$k(y_T^{(i)})$	$x_P^{(i)}$	$k(x_{T_{SS}}^{(i)})$	$y_P^{(i)}$	$k(y_{P_{SS}}^{(i)})$
0.3	1	1.1	1	0.7	3	1.3	2	0.2	1	1.1	1
0.5	2	1.7	3	0.5	2	1.8	4	0.5	2	1.4	3
0.9	4	1.2	2	0.2	1	1.1	1	0.9	4	1.3	2
0.8	3	1.9	4	0.9	4	1.4	3	0.7	3	1.8	4

List of Figures

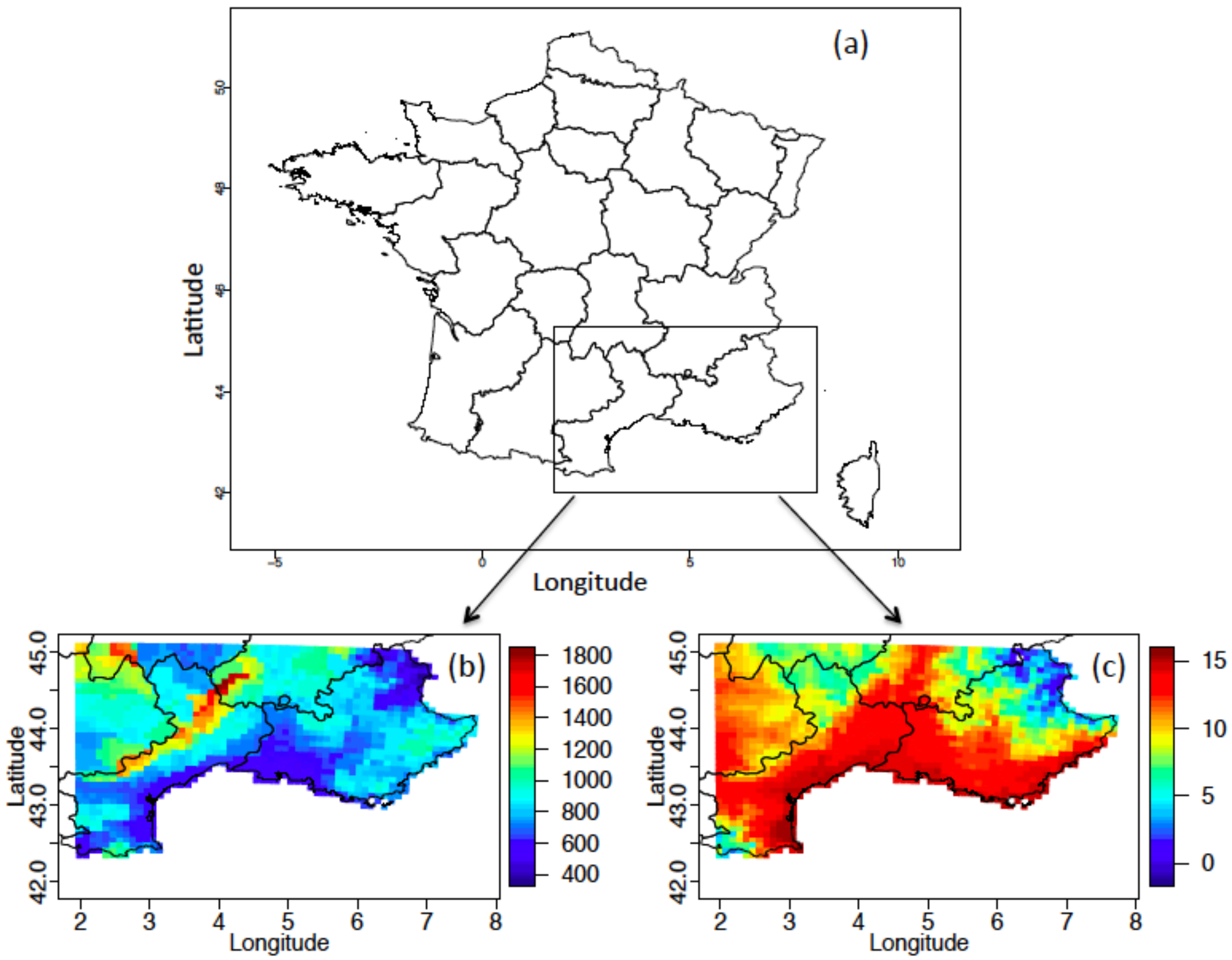

FIG. 1. (a) Map of France with the region of interest in a box, as well as (b) the mean cumulated annual precipitation and (c) the mean daily temperature.

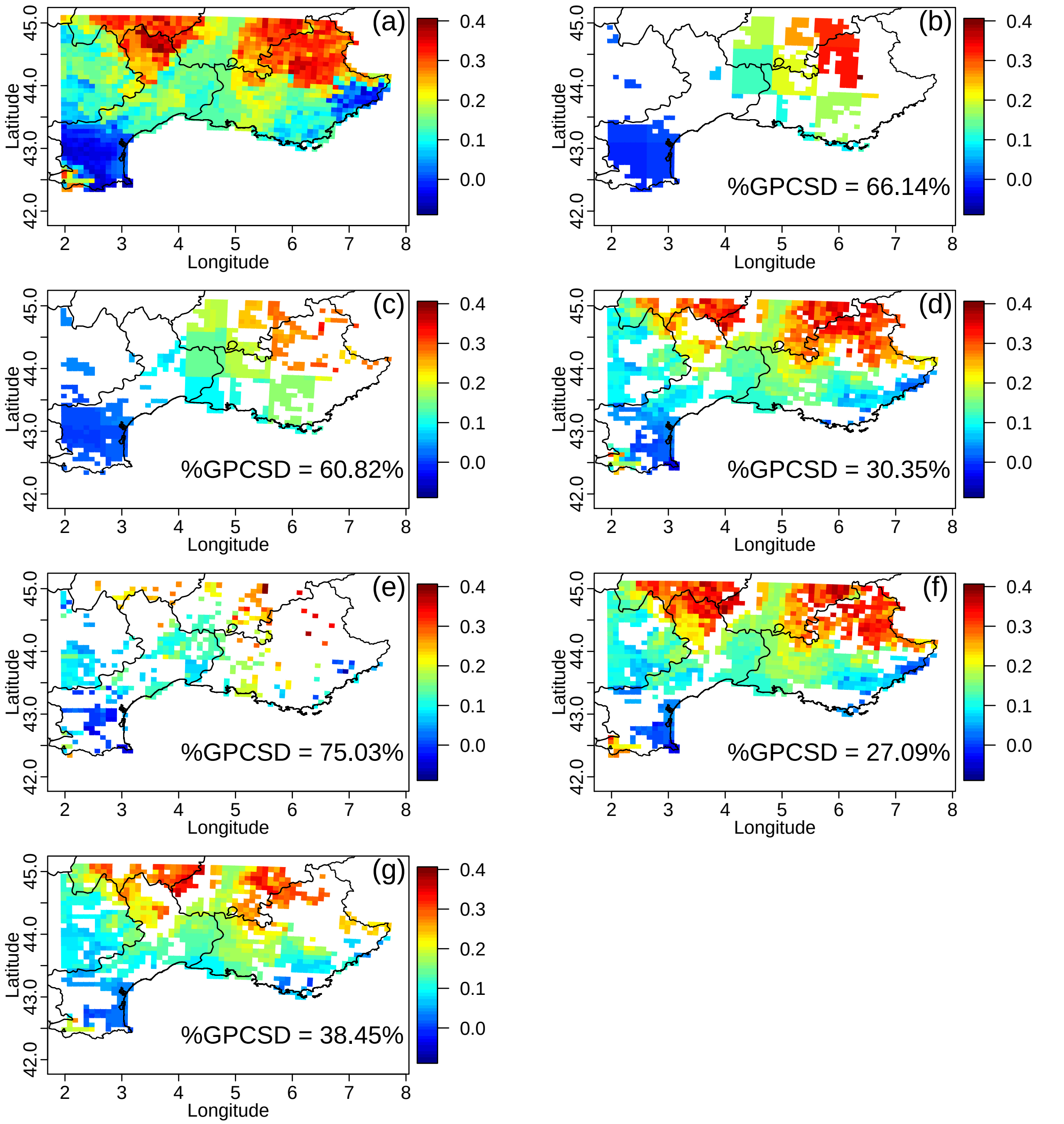

Fig. 2. Maps of inter-variable (PR, T) spearman correlations for the different approaches in winter: (a) SAFRAN; (b) ERA-I; (c) independent BC (through CDFt); (d) conditional BC of T2 given PR; (e) conditional BC of PR given T2; (f) EC-BC; (g) Schaake shuffle on ERA-I. The percentage of grid-points with correlation significantly different (%GPCSD) from that of SAFRAN is indicated on each panel.

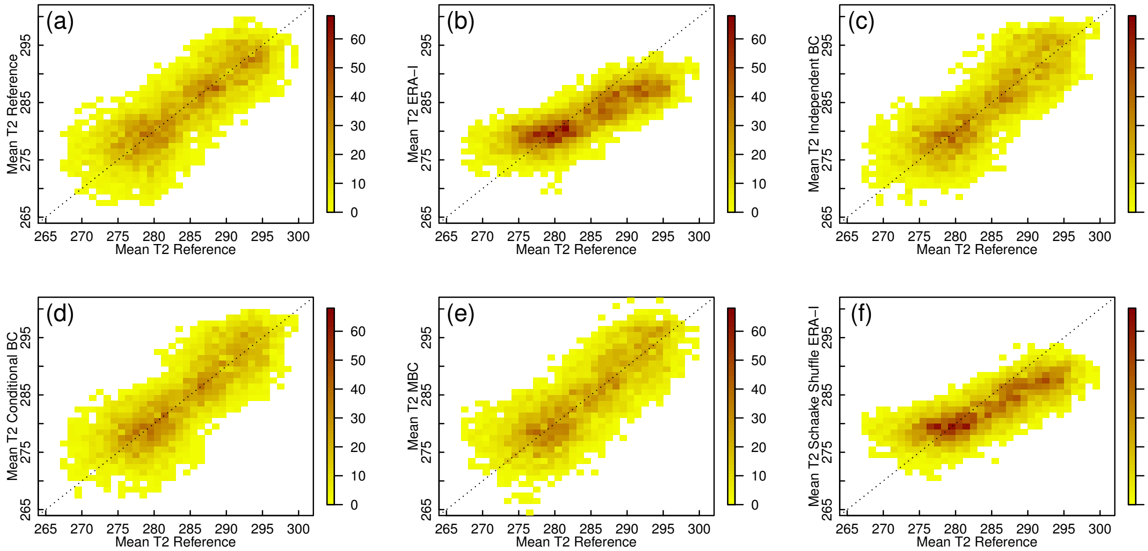

Fig. 3. Bivariate histogram between area-mean 2m temperature of reference data and (a) annually exchanged reference data, (b) ERA-I without bias correction, (c) bias corrected ERA-I data using independent bias correction, (d) conditional approach with T2 given PR, (e) EC-BC (solid line) and (f) Schaake shuffle on ERA-I.

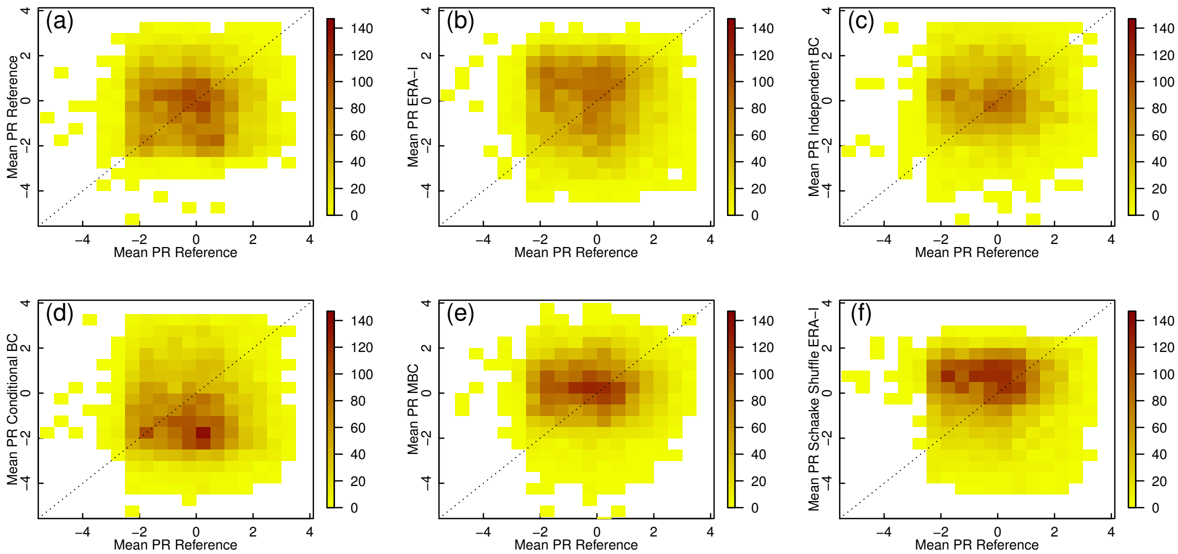

Fig. 4. Same as Fig. 3 but for precipitation.

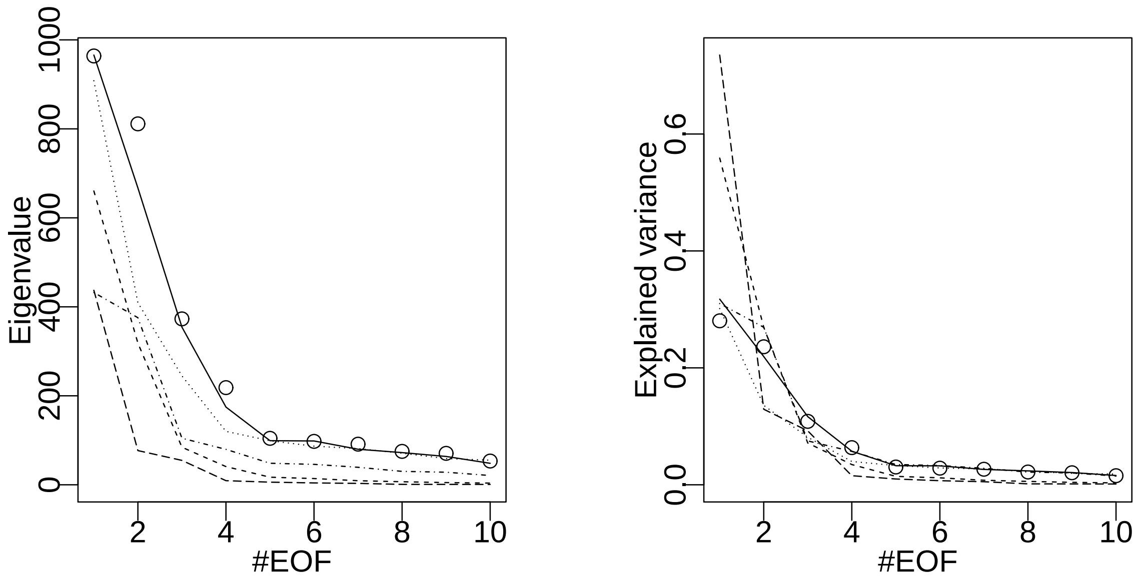

Fig. 5. Eigenvalues (left) and explained variance (right) of leading EOF of the 2m temperature reference data (circles), bias corrected ERA-I data using independent bias correction (dashed line), ERA-I without bias correction (long dashed line), conditional approach with T2 given PR (dotted line), EC-BC (solid line) and Schaake shuffle on ERA-I (dot-dashed line).

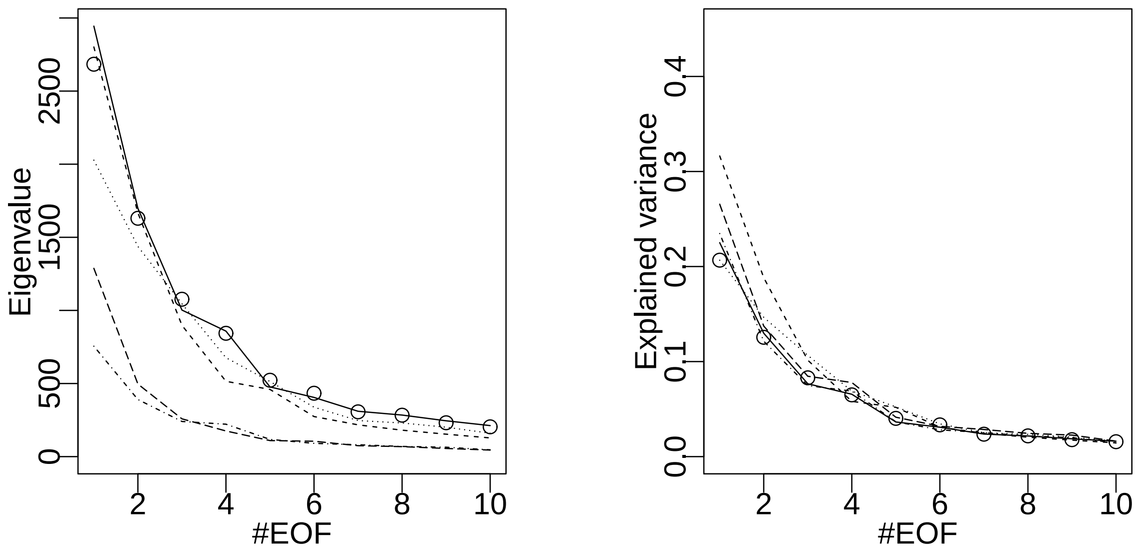

Fig. 6. Eigenvalues (left) and explained variance (right) of leading EOF of the precipitation reference data (circles), bias corrected ERA-I data using independent bias correction (dashed line), ERA-I without bias correction (long dashed line), conditional approach with PR given T2 (dotted line), EC-BC (solid line) and Schaake shuffle on ERA-I (dot-dashed line).

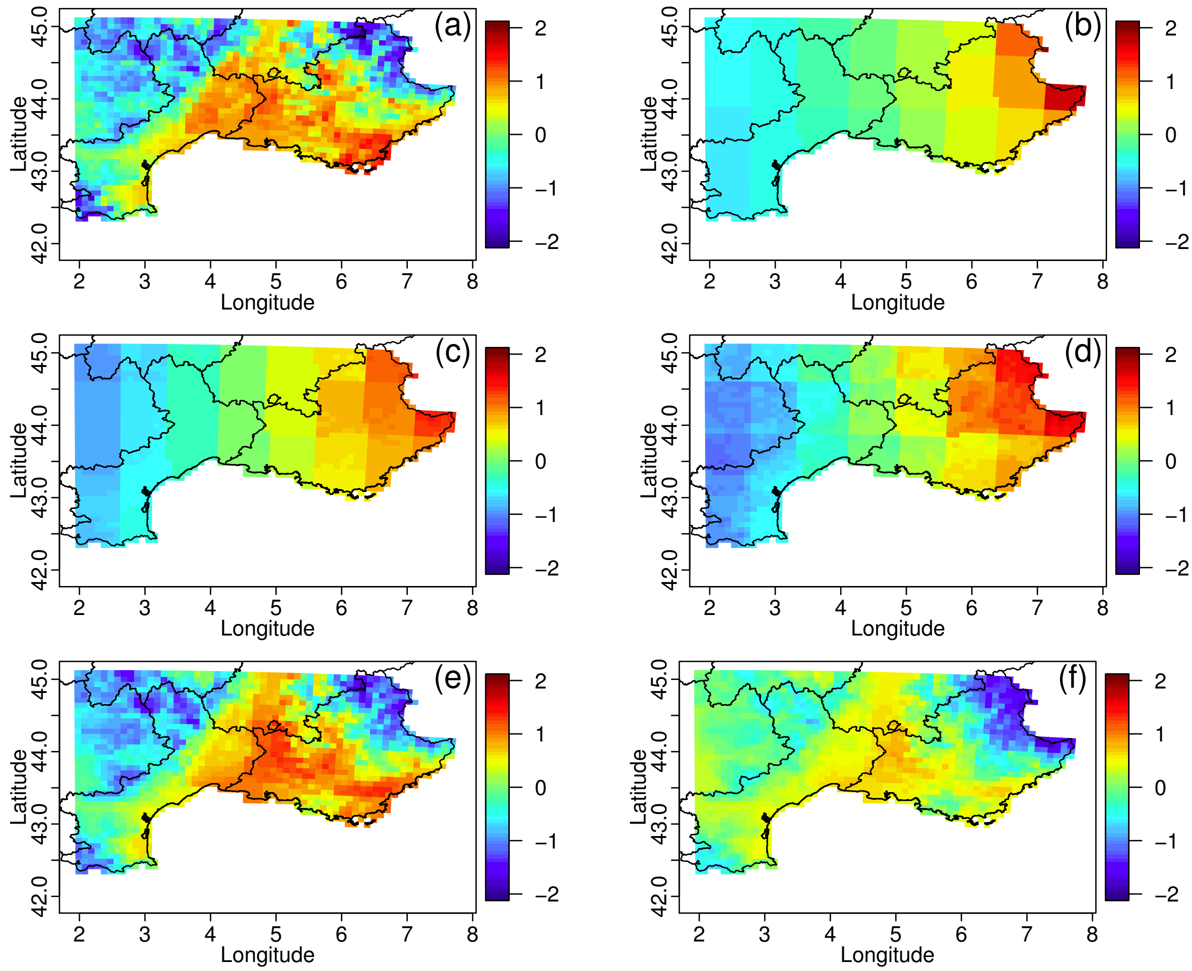

FIG. 7. First EOF of 2m temperature for (a) reference, (b) ERA-I, (c) independent bias correction, (d) conditional approach with T2 given PR, (e) EC-BC and (f) Schaake shuffle on ERA-I without BC.

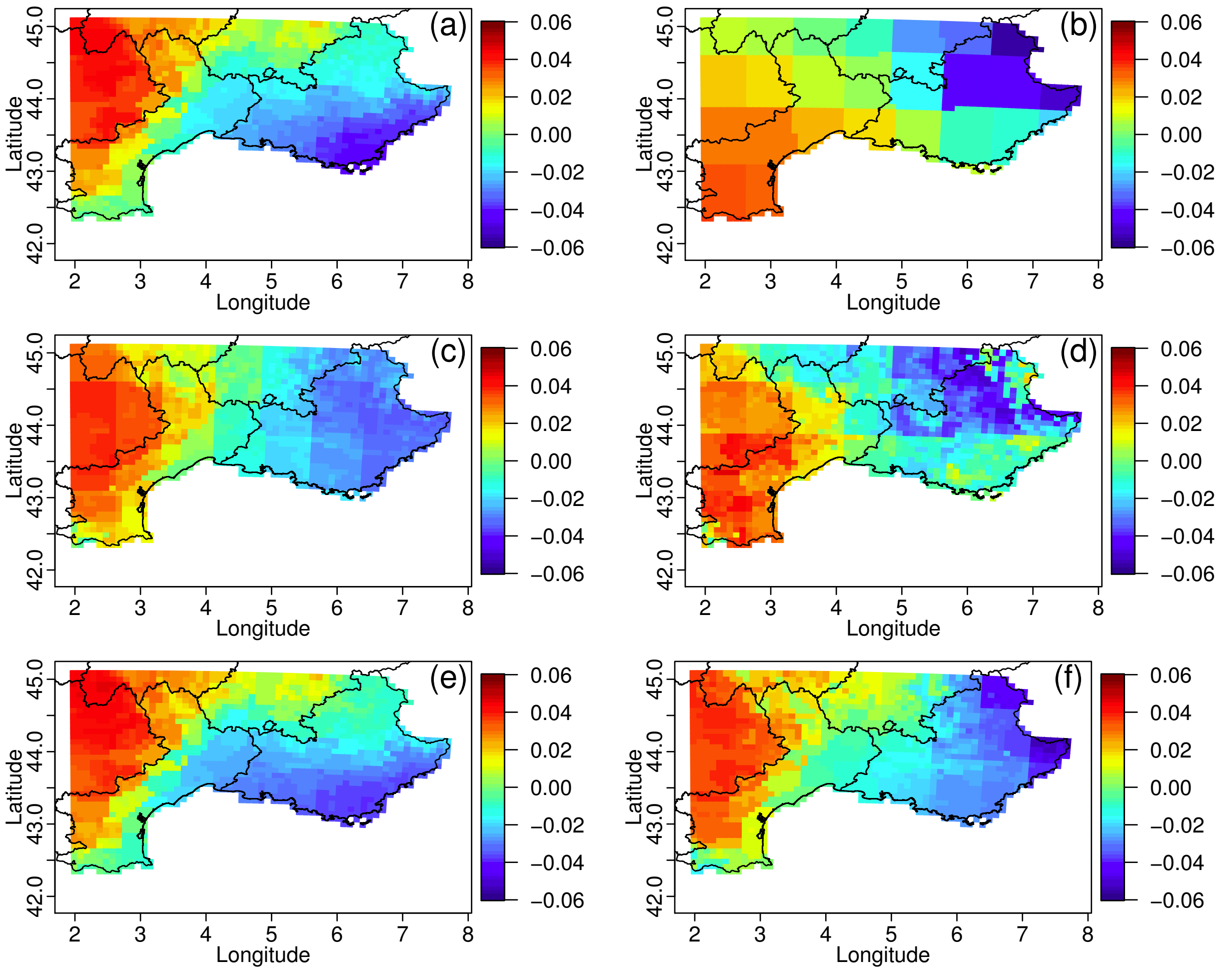

FIG. 8. First EOF of log-precipitation (zeros set to 0.00033) for (a) reference, (b) ERA-I, (c) independent bias correction, (d) conditional approach with PR given T2, (e) EC-BC and (f) Schaake shuffle on ERA-I without BC.

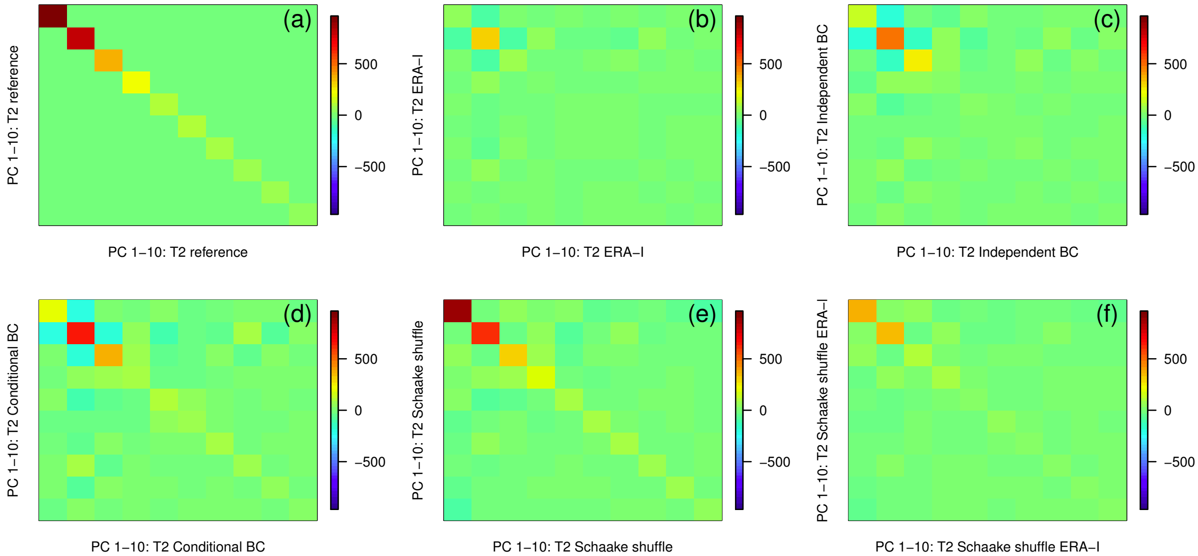

FIG. 9. Covariance matrix of leading 10 PC of 2m temperature for (a) reference, (b) ERA-I, (c) independent bias correction, (d) conditional approach with T2 given PR, (e) EC-BC and (f) Schaake shuffle on ERA-I.

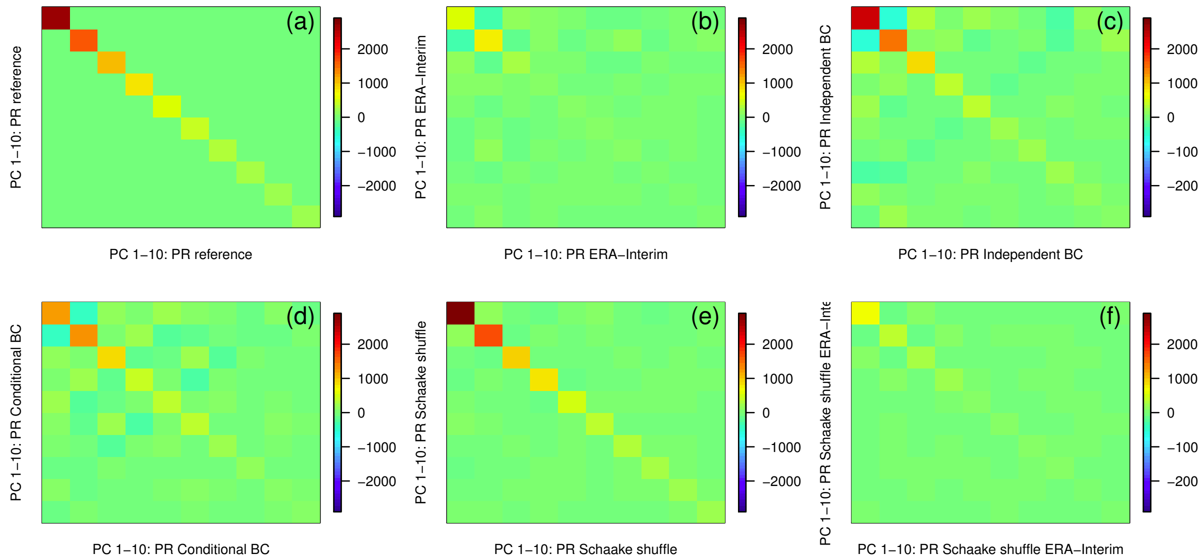

FIG. 10. Covariance matrix of leading 10 PC of log-precipitation (zeros set to 0.00033) for (a) reference, (b) ERA-I, (c) independent bias correction, (d) conditional approach with PR given T2, (e) EC-BC and (f) Schaake shuffle on ERA-I.

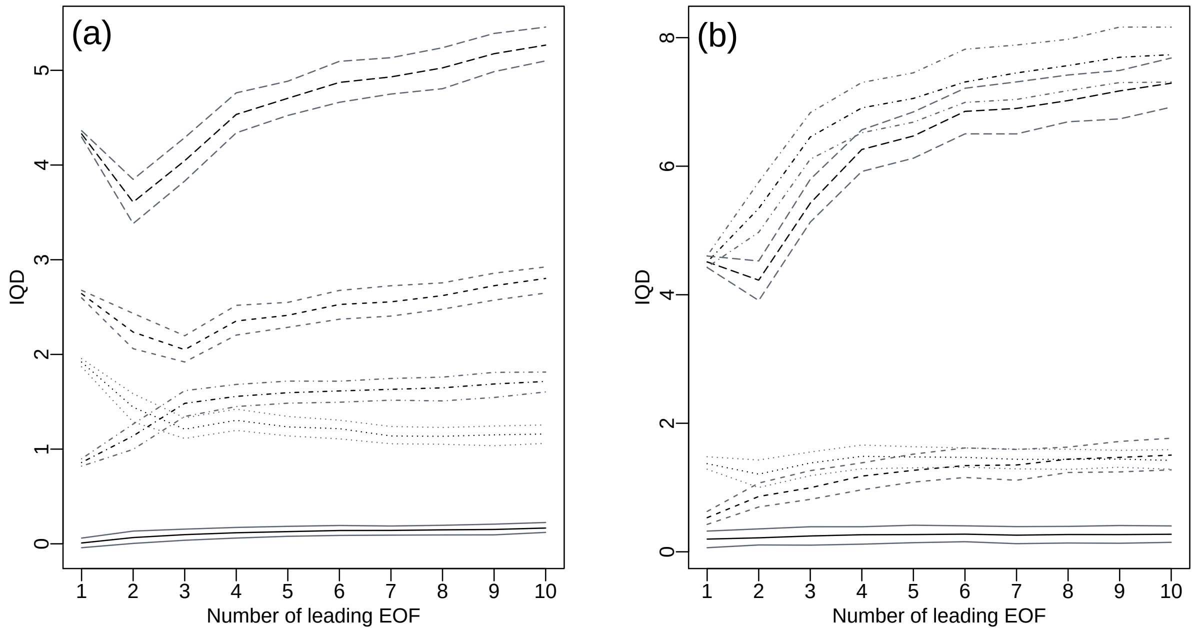

Fig. 11. IQD for (a) 2m temperature and (b) log-precipitation (zeros set to 0.00033) for bias corrected ERA-I data using independent bias correction (dashed line), ERA-I without bias correction (long dashed line), conditional approach (a) with T2 given PR and (c) with PR given T2 (dotted line), EC-BC (solid line) and Schaake shuffle on ERA-I (dot-dashed line).

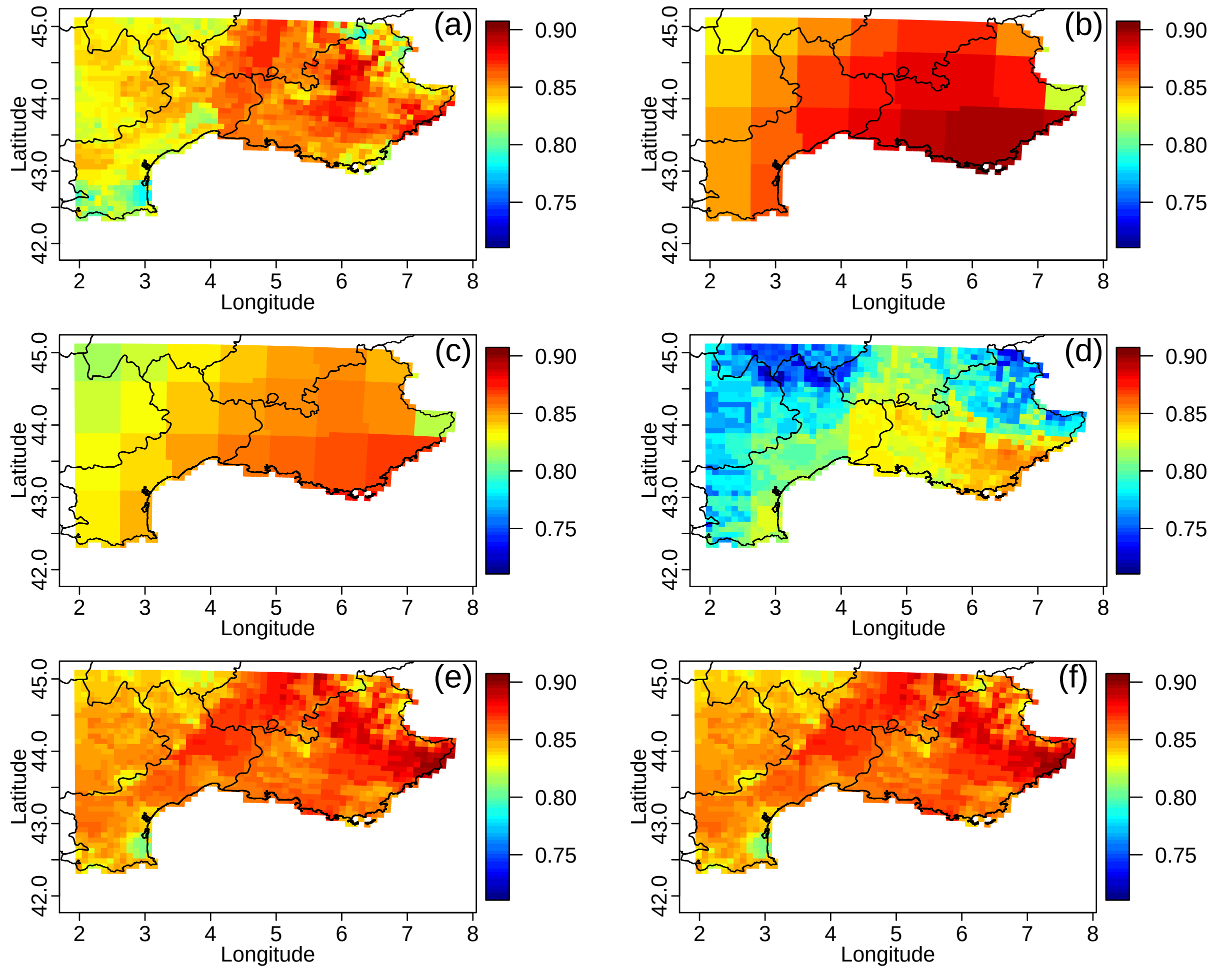

FIG. 12. Maps of 1-day lag temperature auto-correlations in winter for (a) reference, (b) ERA-I, (c) independent bias correction, (d) conditional approach with T2 given PR, (e) EC-BC and (f) Schaake shuffle on ERA-I without BC.

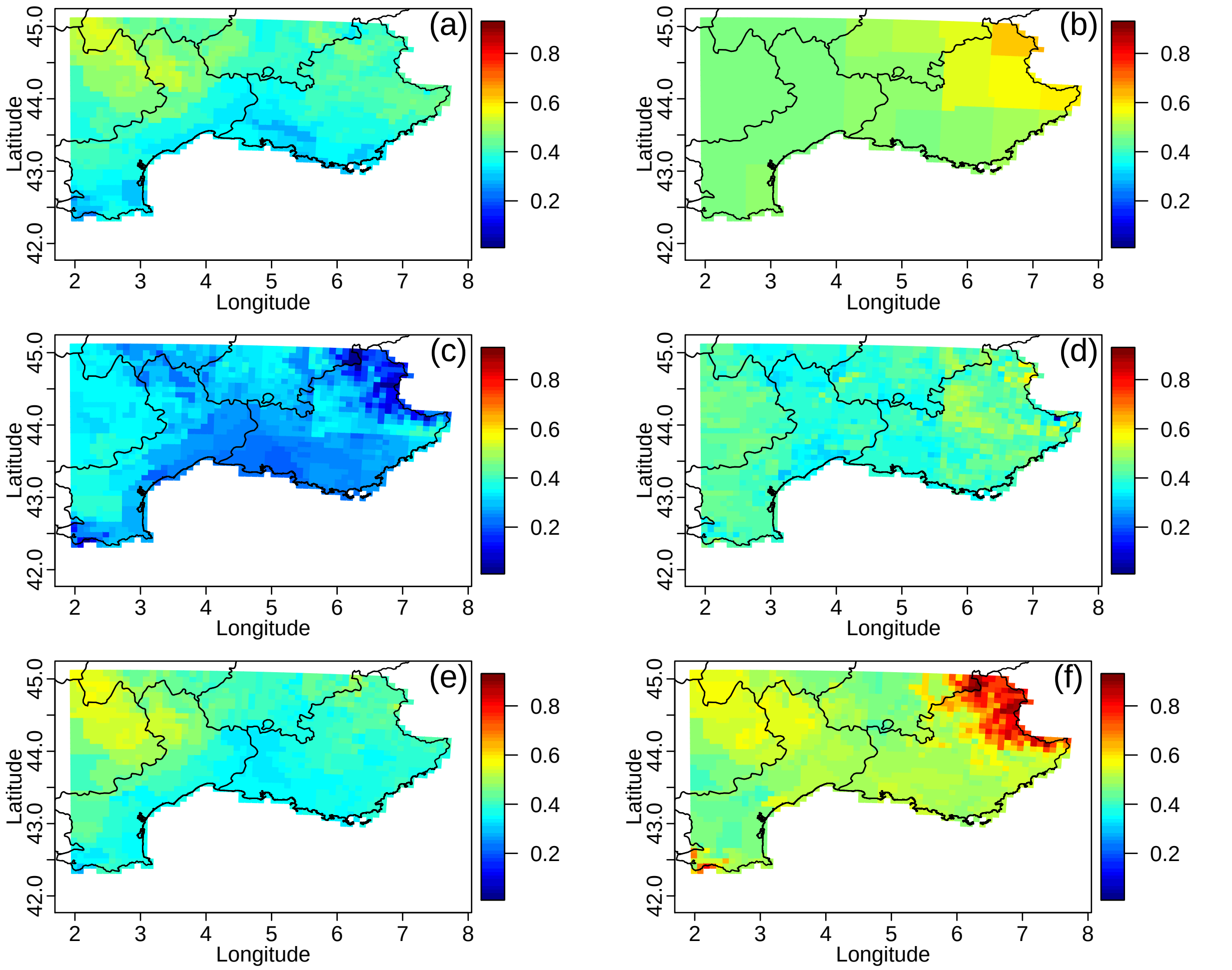

FIG. 13. Maps of 1-day lag precipitation auto-correlations in winter for (a) reference, (b) ERA-I, (c) independent bias correction, (d) conditional approach with PR given T2, (e) EC-BC and (f) Schaake shuffle on ERA-I without BC.

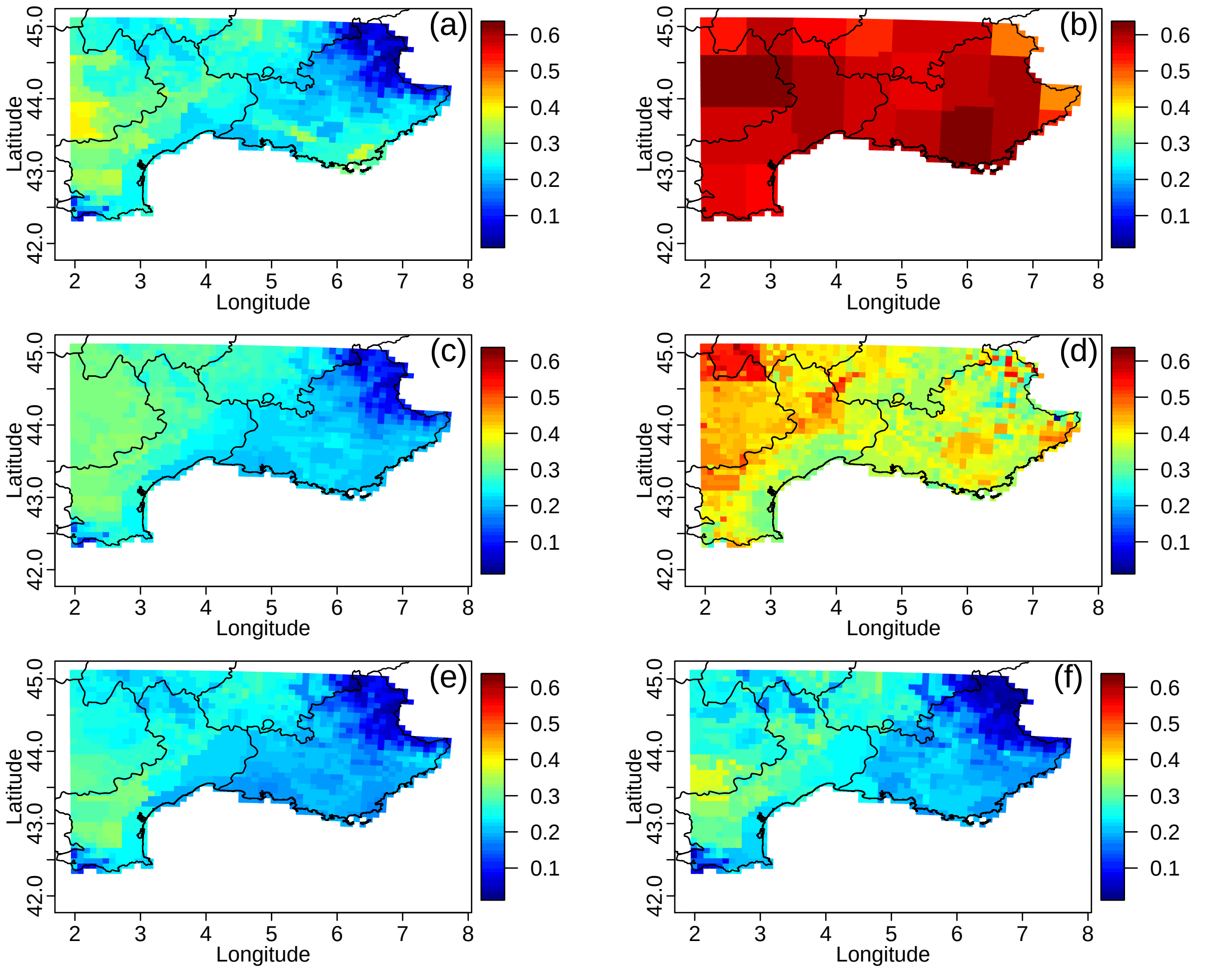

FIG. 14. Maps of daily probability of a dry rain given that the previous day was wet – i.e., $Proba(dry|wet)$ – in winter for (a) reference, (b) ERA-I, (c) independent bias correction, (d) conditional approach with PR given T2, (e) EC-BC and (f) Schaake shuffle on ERA-I without BC.

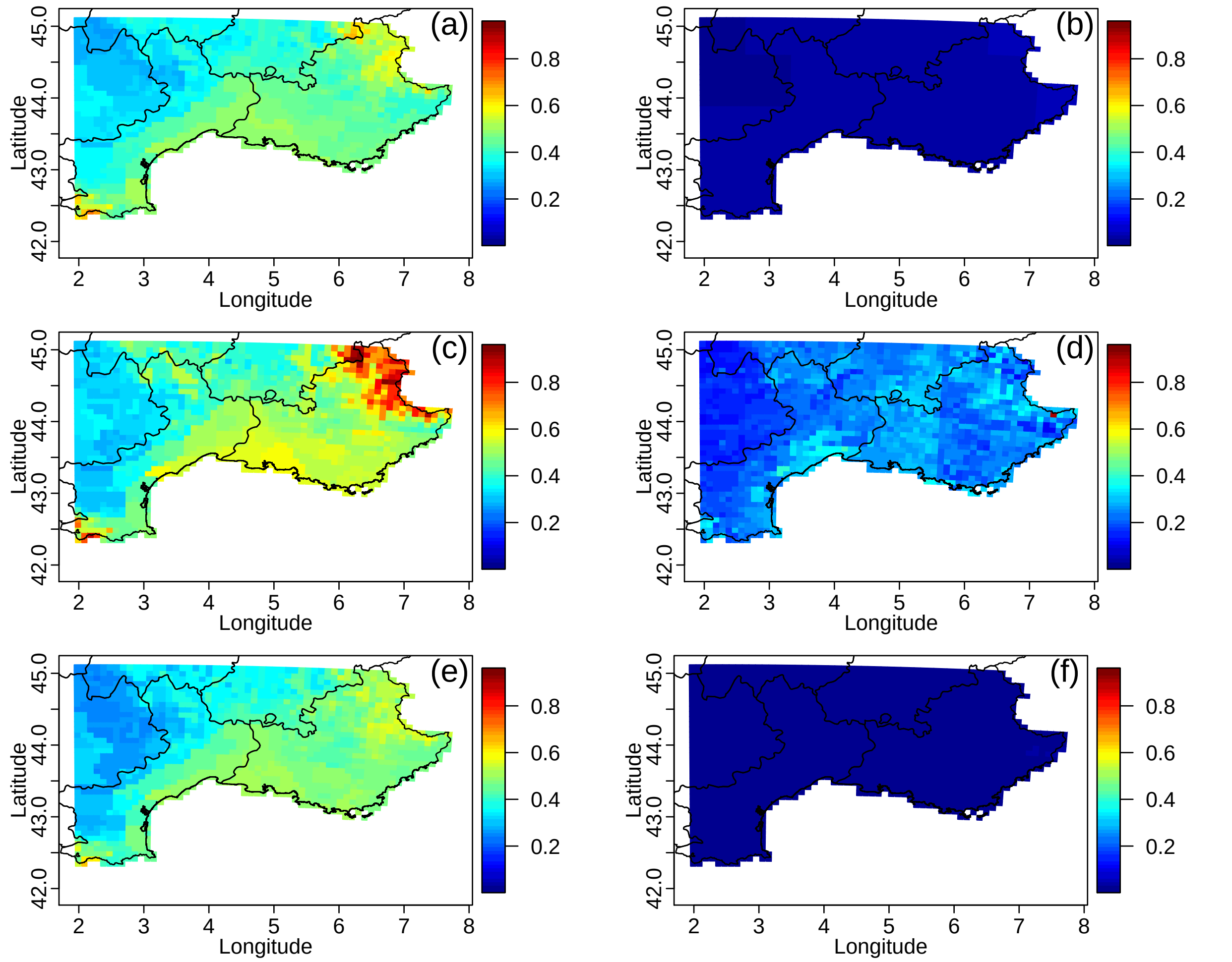

FIG. 15. Maps of daily probability of rain occurrence (i.e., wet day) given that the previous day was dry – i.e., $Proba(wet|dry)$ – in winter for (a) reference, (b) ERA-I, (c) independent bias correction, (d) conditional approach with PR given T2, (e) EC-BC and (f) Schaake shuffle on ERA-I without BC.