Intercomparison of statistical and dynamical downscaling models under the EURO- and MED-CORDEX initiative framework: present climate evaluations

Pradeebane Vaittinada Ayar, Mathieu Vrac, Sophie Bastin, Julie Carreau, Michel Déqué, Clemente Gallardo

Intercomparison of statistical and dynamical downscaling models under the EURO- and MED-CORDEX initiative framework: present climate evaluations

Pradeebane Vaittinada Ayar[1] · Mathieu Vrac[1] · Sophie Bastin[2,3,4] · Julie Carreau[5] · Michel Déqué[6] · Clemente Gallardo[7]

Abstract Given the coarse spatial resolution of General Circulation Models, finer scale projections of variables affected by local-scale processes such as precipitation are often needed to drive impacts models, for example in hydrology or ecology among other fields. This need for high-resolution data leads to apply projection techniques called downscaling. Downscaling can be performed according to two approaches: dynamical and statistical models. The latter approach is constituted by various statistical families conceptually different. If several studies have made some intercomparisons of existing downscaling models, none of them included all those families and approaches in a manner that all the models are equally considered. To this end, the present study conducts an intercomparison exercise under the EURO- and MED-CORDEX initiative hindcast framework. Six Statistical Downscaling Models (SDMs) and five Regional Climate Models (RCMs) are compared in terms of precipitation outputs. The downscaled simulations are driven by the ERAinterim reanalyses over the 1989–2008 period over a common area at 0.44° of resolution. The 11 models are evaluated according to four aspects of the precipitation: occurrence, intensity, as well as spatial and temporal properties. For each aspect, one or several indicators are computed to discriminate the models. The results indicate that marginal properties of rain occurrence and intensity are better modelled by stochastic and resampling-based SDMs, while spatial and temporal variability are better modelled by RCMs and resampling-based SDM. These general conclusions have to be considered with caution because they rely on the chosen indicators and could change when considering other specific criteria. The indicators suit specific purpose and therefore the model evaluation results depend on the end-users point of view and how they intend to use with model outputs. Nevertheless, building on previous intercomparison exercises, this study provides a consistent intercomparison framework, including both SDMs and RCMs, which is designed to be flexible, i.e., other models and indicators can easily be added. More generally, this framework provides a tool to select the downscaling model to be used according to the statistical properties of the local-scale climate data to drive properly specific impact models.

Keywords Statistical downscaling · Dynamical downscaling · CORDEX · Precipitation · Intercomparison

✉ Pradeebane Vaittinada Ayar
pvait@lsce.ipsl.fr

[1] Laboratoire des Sciences du Climat et de l'Environnement (LSCE-IPSL), CNRS/CEA/UVSQ, Centre d'Etudes de Saclay, Orme des Merisiers, 91191 Gif-sur-Yvette, France

[2] Université Versailles St-Quentin, Versailles, France

[3] Sorbonne Universités, UPMC Univ. Paris 06, Paris, France

[4] CNRS/INSU, LATMOS-IPSL, 11 bd d'Alembert, 78280 Guyancourt, France

[5] HydroSciences Montpellier (HSM), CNRS/IRD/UM1/UM2, Place Eugène Bataillon, 34095 Montpellier, France

[6] Météo-France, Centre National de Recherches Météorologiques, 42 Av. Coriolis, 31057 Toulouse, France

[7] Instituto de Ciencias Ambientales, Universidad de Castilla-La Mancha, Toledo, Spain

1 Introduction

The study of the many environmental and socio-economic impacts of meteorological phenomena and climate change

implies to improve our knowledge of climate at a local scale. Indeed, studying climate change impacts on agriculture, water resources, pollution, and many other environmental features at a human scale makes high-resolution model simulations essential. However, General Circulation Model (GCM) simulations of the different future climate scenarios prescribed by the Intergovernmental Panel on Climate Change (Vuuren et al. 2011) have generally a coarse spatial resolution (about 250 km) and are thus not adapted as inputs into the impacts models that need much finer scale climate information. Hence, it is required to bring GCMs climate simulations information to more regional or local scales, i.e., to generate high-resolution simulations based on (reanalyses or GCM) large-scale information. This is the aim of downscaling. Downscaling models can be dynamical or statistical, both approaches being driven by GCMs or reanalysis data.

Dynamical downscaling models correspond to the so-called "Regional Climate Models" (RCM), which simulate high-resolution physical processes consistent with the prescribed large-scale dynamics. RCMs can be a GCM with grid refinement over a specific region (e.g., Déqué and Piedelievre 1995; Hourdin et al. 2006) or a limited area model (LAM) constrained at its lateral boundaries by GCMs (WRF, Skamarock et al. 2008). Both GCMs and LAMs are sensitive to the resolution and the physical package which regroups all the model parametrization used in the model to take into account sub-grid scale processes. While the use of LAMs presents some advantages, for instance the fact that they are non-hydrostatic allows very high-resolution downscaling or also the possibility to set a region-specific parametrization, it also creates discontinuities at the boundaries. Previous studies have investigated the sensitivity of the results to the frequency of boundary conditions, size and resolution of the domain (e.g., Noguer et al. 1998; Seth and Giorgi 1998), lateral conditions (Denis et al. 2003) and frequency of reinitialization (Lo et al. 2008). Those studies show that the internal variability of RCMs can strongly influence the results at regional scales and that the small-scale field inside the domain is not always consistent with the driving field (Laprise et al. 2008). To ensure the consistency between the small- and large-scale fields, the model can be driven using nudging techniques (e.g., Omrani et al. 2012a, b). The choice of the physical package that allows the model to simulate all the sub-grid scale processes using parameterizations is also very important and induces large discrepancies between model outputs (e.g., Flaounas et al. 2011). Despite the increase of computing power, running an RCM including all those different formulations still requires important computational resources. This often puts limits on the number, the resolution and the time period length of the RCMs runs.

The alternative approach to RCMs is based on Statistical Downscaling Models (SDMs) that rely on determining statistical relationships between large- and local-scale variables and do not try to solve the physical equations modelling the dynamic of the atmosphere. Due to their statistical formulation, they generally have a low computational cost and provide relatively fast simulations. SDMs are now considered as complementary to RCMs, for example in terms of applications for ensembles uncertainties studies (Sachindra et al. 2014). SDMs are based on a static relationship, i.e. the mathematical formulation of the relation between predictand (i.e., the local-scale variable to simulate) and predictors (i.e., the large-scale information or data used as inputs in the SDMs) is supposed to be valid for any time period: not only for the current climate on which the relationship is calibrated, but also, for example, for future climates. This does not mean that the statistical properties of the predictands are stationary (i.e., are the same in current and future climates): if the statistical properties of the large-scale predictors evolve in time, those of the local-scale predictands will evolve as well. Hence, if the relationship is said to be "static" (or "stationary"), the statistical distribution of the predictands is "non-stationary" and the SDMs can be said to be non-homogeneous (e.g., Vrac et al. 2007b). Most state-of-the-art SDMs can be divided into the four following families: transfer functions (TFs), stochastic weather generators (WGs), weather typing (WT) based methods and model output statistics (MOS).

The TFs approach regroups the deterministic functions which "transfer" the large-scale information to the local scale. Those mathematical functions characterize the nature of the dependencies between the predictors and the predictands. They could be linear [e.g. through a multi-linear regression (MLR), see Jeong et al. 2012] or non-linear [.g. through polynomial regression or artificial neural network (ANN) see Xiaoli et al. 2008] These methods are usually easy to implement and apply but tend to underestimate the variance (see the variance inflation procedure in Wilby et al. 2002). One solution is to use a stochastic modelling in order to adapt the statistical distribution instead of "inflating" the variance.

Stochastic WGs simulate daily weather scenarios thanks to probability distribution functions (pdfs) estimated from observations. A wide range of WGs has been developed to generate weather variables (e.g., see for a review Wilks 2012). Historically, WGs were used to reproduce the observed rain statistical properties (Wilks 2010). However, in a downscaling context where the statistical properties may evolve in time, WGs have to be based on pdfs that depend on atmospheric predictors. These conditional pdfs can evolve in time, i.e., their parameters can change with the predictors (e.g. Bardossy and Plate 1992). This

approach is particularly interesting to generate variability in data.

The WT approach defines large-scale patterns from circulation variables and rely on clustering techniques. The main assumption is that for a given large-scale pattern, the relationship between the large- and local-scale variables is always the same. One particular method is the "analog" method where each daily large-scale situation is considered as a pattern. For a day to be downscaled, the day in the past which has the closest large-scale situation (according to a similarity metric) is chosen (Zorita and Storch 1999). The local-scale observations of the selected day are then the downscaled values. This approach also provides methods easy to implement. However, in climate change context these methods can miss a possible climate change signal because of their inability to generate values beyond the range of past values.

All the previous approaches need daily synchronicity between large-scale and local-scale data to be calibrated. They are referred to as "Perfect-prog" downscaling (Klein et al. 1959). Model output statistics approach is quite different by essence because it generally works directly on model outputs, without calibration based on reanalysis data. MOS aims to link characteristics like the mean, the variance or the probability (or cumulative) distribution function (pdf or CDF). This approach presents many interesting applications in terms of downscaling and bias correction but the performance is deeply linked to the quality of the modelled large-scale variable (Coiffier 2011).

Many different intercomparison studies have been conducted lately. These studies have a wide range of purposes. They can be discriminated for instance by the type of models which were compared: RCMs only, SDMs only or both SDMs and RCMs.

Concerning RCMs, several coordinated projects have been developed involving collaborations between Regional Climate Modelling groups. There are several projects taking place around the world over different regions. Over Europe, the MERCURE project (1997–2000, Machenhauer et al. 1998), aimed at identifying the strengths and weaknesses of RCM simulations driven by atmospheric analyses. It led to the project PRUDENCE,[1] (2001–2004, Christensen et al. 2007) where one important goal was to analyse future projections according to four uncertainty aspects: natural variability, greenhouse gases emissions and concentrations scenarios, the choice of the driving GCM atmospheric and oceanic boundary conditions and finally the RCM formulation. This was followed by the project ENSEMBLES[2] (2004–2009, Hewitt 2004). It produced for the first time a probabilistic estimate of uncertainty of future climate at several timescales, using an ensemble validated against observational datasets for Europe. Note that similar projects exist over other regions like the Asian RMIP project (Fu et al. 2005) or the North American projects PIRCS (Takle et al. 1999) and NARCCAP (Mearns et al. 2013). Lately, the Coordinated Regional Climate Downscaling Experiment (CORDEX, Giorgi et al. 2009) initiative from the World Climate Research Program promotes running multiple RCM simulations at 50 km and higher resolution for multiple regions. This initiative is mainly aiming to assess RCMs quality and uncertainty for the recent past and for twenty first century projections, covering the majority of populated land regions on the globe. The uncertainties are associated with varying GCMs simulations, varying greenhouse gas concentration scenarios, natural climate variability and different downscaling methods. In contrast to the former intercomparisons, the CORDEX initiative impose several additional and mandatory constraints which make the runs comparable. The constraints include domains definition, time period, same spatial resolution and boundary forcing (ERAinterim Reanalysis, Dee et al. 2011) for the hindcast evaluation to provide a framework for model evaluation and assessment.

SDMs-focused intercomparisons are also more and more available now but are mostly done by modest research initiatives compared to CORDEX for instance. One of the first intercomparison studies was brought by Wilby and Wigley (1997), who aimed to make a review of the available SDMs at the time and to compare precipitation models in terms of present and future climate over north America. Six SDMs calibrated on NCEP reanalyses have been compared with one GCM. The main result pointed out intervariable inconsistencies in the GCM which made unreliable the precipitation changes generated by the GCM. Even if this study was quite exhaustive, the MOS approach was not represented in it.

Since then, many intercomparison studies have been conducted, often not taking into account one or several SDM approaches and with specific purposes. For instance, Schoof and Pryor (2001) aimed to compare two TF methods calibrated over circulation indices on midwestern USA. The evaluation performed on present climate pointed out that the models failed to capture the variability of precipitation as governed by the large-scale circulation and suggested that other variables were necessary to capture precipitation. Although this paper is an important contribution, only TF methods were discussed in this study. Focussing also mostly on TFs methods, Harpham and Wilby (2005) evaluated two ANN-based SDMs (i.e., TF) and one WG to downscale heavy precipitation and their multisite behaviour in a present climate context over United Kingdoms. A follow up study included three supplementary SDMs and two

[1] http://prudence.dmi.dk/.

[2] http://www.ensembles-eu.org.

RCMs in a future climate context (Haylock et al. 2006). The results underlined the need of an ensemble approach when considering future climate projections. However, the WT approach was missing in the first study and the MOS approach in both studies. Similar studies conducted over the Serpent River basin (Quebec, Canada) aimed to focus on a particular temporal neural network (TNN, i.e., TF) to downscale precipitation (Dibike and Coulibaly 2006; Khan et al. 2006). Results showed the high-performance of that particular TNN model but the study did not include WT and MOS models in both cases. Moreover, in a recent study, Gaitan et al. (2014) aimed to compare high-resolution precipitation models over Ontario and Quebec, Canada to reproduce climate change signal based on the RCM pseudo-observation approach developed in Vrac et al. (2007c). Six rain occurrence WT models and four rain intensity TF models have been designed to this end from the same set of predictors. Their ANN (i.e., TF named ANN-F in their study) was found to be the best model. Although an interesting intercomparison, the study did not investigate MOS and WG approaches.

Other recent studies compared methods from the four statistical families but with different predictors. For instance, six SDMs and three RCMs precipitation outputs were compared over the Alps by Schmidli et al. (2007). The SDMs were calibrated on several reanalysis databases for present climate and applied to GCMs for future climate. Results showed that the statistics of most statistically and dynamically downscaled precipitation were similar. In another study by Bürger et al. (2012), five SDMs with their own set of predictors have been evaluated in a present climate context over British Columbia, Canada, focusing on extremes aspects of the compared SDMs for temperature and precipitation. It turned out that the use of hybrid models (i.e., models with components built from several families) made difficult to identify the component of the models which explains the model efficiency. Even if all the SDMs families were studied in both papers, the models were calibrated on different sets of predictors. A common set of predictors would have allowed an easier comparison.

All these references are examples and this list is by no means exhaustive. They give a general idea of some major studies. More generally, all those studies did not include a cross-validation procedure in their model evaluation (except in Gaitan et al. 2014 see also Vrac et al. 2007c) which is an important step to validate SDMs in a present climate context (this notion is illustrated in Sect. 2.2). Even if they compared many models with different interesting features and results, they all presented some inconsistencies. Indeed, one important argument is that in most studies the predictors were different: for instance they were selected according to the observation station (Harpham and Wilby 2005; Haylock et al. 2006; Dibike and Coulibaly 2006; Khan et al. 2006) or were specific to the models features (Schmidli et al. 2007; Bürger et al. 2012). Sometimes the purpose of the study was not the intercomparison itself but rather to underline the developments of a new model (e.g., Dibike and Coulibaly 2006; Harpham and Wilby 2005). As said above, in most of the studies at least one type of model is missing and the SDMs were calibrated on more or less sparse observation network. In this paper, to perform a consistent intercomparison, we want to compare models outputs from all types of models (i.e., from the four approaches of SDMs and from RCMs, see Schmidli et al. 2007) and observational data with similar resolution over a common area. Another criterion is to calibrate all the models with a common set of predictors (as much as possible) with a cross-validation procedure (see Gaitan et al. 2014). Thus, the three main requirements of this intercomparison study are: (1) models must have common predictors, (2) RCMs and SDMs model outputs and observations have to share the same area and resolution, (3) all SDMs families models have to be represented and a representative number of dynamical models have to be included. Recently, two initiatives sharing similar objectives have to be mentioned: CORDEX-ESD (http://wcrp-cordex.ipsl.jussieu.fr/index.php/community/cordex-esd), and to the COST Action VALUE (http://www.value-cost.eu/, Maraun et al. 2015). These projects aim at coordinating SDMs intercomparison at a continental scale and make SDMs comparable to RCMs.

The present intercomparison takes place under the CORDEX initiave hindcast evaluation: all the models have to be forced by ERAinterm reanalyses and run over the 1989–2008 period at 0.44° resolution. For the present study, the variable of interest is the precipitation. This choice is motivated by its high spatial and temporal variability and the difficulties faced to model precipitation compared to other variables like temperature. Another argument is that rainfall is one of the most important variables for many impact studies (e.g. for floods prediction, Raje and Mujumdar (2010) or crop yields, Oettli et al. 2011).

Hence, in this paper, several downscaling models are compared through a common and well defined framework. The aim is to set a generic intercomparison framework. More precisely, our goal is not to select the best model or to develop a model with new features. The objective is to design an intercomparison experiment in which the models are easily confrontable. The performance criteria are expected to be wide enough to correctly inspect the main aspects of the models representing each statistical downscaling family. The chosen indicators are relevant for climatological studies. Indeed, these can be different when considering other application domains (e.g. hydrology), which can produces different performance evaluation results. The proposed framework would help to point out main models strengths

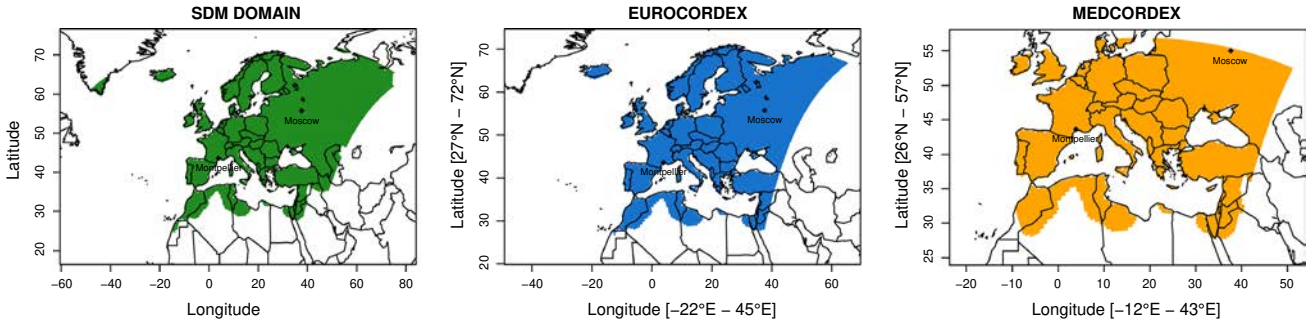

Fig. 1 The models are run and evaluated over the following geographical areas: the SDMs domain in *green* corresponding to the E-OBS domain, the intersection between EURO-CORDEX and E-OBS domain in *blue* which is the evaluation domain of EURO-CORDEX models and the intersection between MED-CORDEX and E-OBS domain in *orange* which is the evaluation domain of MED-CORDEX models

and weaknesses, identify the needed improvements and provide statistically simulated time series to be compared to RCMs over a common area and forced by a common set of predictors (ERAinterim). Differences between models with specific features both in conceptual terms, e.g., dynamical versus statistical or deterministic versus stochastic, and technical details are going to be described and evaluated. This intercomparison is also designed in a way that other models or indicators can be easily added.

This paper is organized as follows: the data and experimental set-up are presented in Sect. 2, while Sect. 3 describes the downscaling models used in this study. The results of the comparison are presented in Sect. 4. Finally, in Sect. 5, some conclusions, perspectives and discussions are proposed.

2 Data and experimental setup

SDMs seek to establish a link between large-scale and local-scale climate data. The experimental setup thus has to state which large-scale variables will act as predictors and which local-scale variables will be predicted. In addition, the validation procedure has to be defined. In order to design the experiment rigorously, it is essential to keep in mind assumptions under which the SDMs are performed (Hewitson and Crane 1996): (1) the relationship between local-scale data and large-scale predictors is fixed in time (even if the statistical properties of the downscaled simulations can evolve in time), (2) the predictors fully represent the climate signal, (3) the large-scale variables are well reproduced by climate models, including reanalysis.

2.1 Local-scale predictands and large-scale predictors

In order to limit any RCM data transformation from their initial spatial resolution, the common resolution of the RCMs at 0.44° or local-scale predictands has been chosen. Therefore, the comparison with the E-OBS V8 gridded dataset from the EU-FP6 project ENSEMBLES[3] and the data providers in the ECA&D project is straightforward[4] (Haylock et al. 2008) at 0.44° × 0.44°. In the experimental setup, the E-OBS precipitation data will serve as local-scale reference predictand for the calibration of the statistical models that will therefore downscale largee-scale information to 0.44° spatial resolution, directly comparable to RCMs outputs. Note that there are some quality inconsistencies in this version of E-OBS data (Hofstra et al. 2009). The reader has to keep in mind that this intercomparaison is done using E-OBS data as reference, which can potentially induce some inexact results over specific areas. This issue is discussed in Sect. 5.2.

As one of the goal of this study is to make intercomparisons between SDMs and RCMs involved in the EURO-CORDEX (Jacob et al. 2014) and MED-CORDEX/HYMEX (Drobinski et al. 2014, www.medcordex.eu/medcordex.php) initiatives, the atmospheric data chosen to drive the statistical models (i.e., the large-scale predictors) are the same as those used as forcing for the RCMs for the hindcast evaluation. Figure 1 represents the geographical areas over which the models are evaluated: in green the SDMs domain corresponding to the domain of E-OBS data, in blue the EURO-CORDEX evaluation domain which is the intersection between EURO-CORDEX and E-OBS domain and in orange the MED-CORDEX evaluation domain which is the intersection between MED-CORDEX and E-OBS domain. The atmospheric variables used as predictors are selected from the ERAinterim Reanalysis (ERAi, Dee et al. 2011) at 1.125° × 1.125° resolution, over the North-Atlantic region which includes the

[3] http://ensembles-eu.metoffice.com.

[4] http://www.ecad.eu.

Table 1 Selected predictors for each season and their correlation and rank into parenthesis given by SPARSE CCA algorithm

Predictors	D2	SLP	T2	R850	U850	V850	Z850	PR
Winter	0.635 (2)	0.679 (1)	0.572 (3)	–	0.298 (5)	0.355 (4)	–	Common predictor
Summer	0.508 (2)	0.342 (5)	0.515 (1)	–	0.444 (3)	–	0.403 (4)	Common predictor

These variables are: the dew point at 2 m (D2), the sea level pressure (SLP), the temperature at 2 m (T2), the relative humidity, the zonal and meridional wind components and the geopotential height at 850 hPa pressure level (R850, U850, V850 and Z850). Besides the large-scale precipitation (PR) is added as common predictor

EURO-CORDEX and MED-CORDEX domains. It corresponds to 5,452 grid-points over the geographical area $[-52.875°\text{E}\,;\,76.50°\text{E}] \times [20.25°\text{N};\,72.00°\text{N}]$. All fields are taken at the daily time scale obtained by averaging 6 h reanalyses outputs. These variables are selected according to many criteria. First, considering the objective of our study which is to intercompare models, a common set of predictors for all the statistical models is needed. Such a consideration makes the study as fair as possible in the way the models are considered. The choice is also motivated by the physical relation of the variables to the precipitation and their role in the precipitation processes. Another criterion is the availability of the common variables in GCMs and Reanalysis products and also the correct representation of the predictors over the domain (Hewitson and Crane 1996). Note that this is not a requirement for the intercomparison itself (one can imagine an intercomparison with badly simulated predictors) but only for our choice of predictors. Table 1 shows the chosen variables. Some of them have been widely used in statistical downscaling context with good results. For instance, surface variables such as the temperature at 2 m (T2), the sea level pressure (SLP) or atmospheric variable as the geopotential height, the zonal and meridional wind components and relative humidity at 850 hPa (Z850, U850, V850, R850) can be found in studies like Cavazos et al. (2005) or Crawford et al. (2007). The dew point at 2 m (D2) was also added. Physically, precipitation results from saturation of water vapour due to a vertical lift of the atmospheric cell, that is to say a combination between atmospheric instability and humidity convergence. As saturation is a non-linear function of temperature and moisture, it is important to include both temperature and moisture (relative, specific, or dew point temperature) as predictors. Moreover, SLP (or geopotential height at some tropospheric level) is a good large-scale predictor candidate, as it includes the direction of the advection (which implicitly interacts with orography) and the convergent motions (which produce also vertical lift). The U and V components of the wind bring also relevant information in terms of synoptic motions. Finally, using two levels (SLP and Z850) enables to take into account, to some extent, the vertical stability of the lower troposphere through the baroclinicity. T2 also accounts for the degree of atmospheric stability. A statistical analysis based on sparse canonical correlation analysis (SPARSE CCA, Witten et al. 2009) was conducted and corroborates our choice of predictors. Traditional CCA seeks the best projections of two sets of variables (in our case, the predictors and the predictand over the spatial domain) by iteratively maximizing the correlation of the projections. The sparse version of CCA adds sparsity constraints on the projections resulting in projection vectors with a number of zero coefficients which depends on the sparsity enforced. Each potential predictor variable was first spatially summarized by taking its first principal component (PC) computed from a principal component analysis (PCA, Barnston and Livezey 1987) applied—separately for each climate variable—to the 5452 grid-points over the North-Atlantic region. Then SPARSE CCA was carried out between a set containing the first PC of each of the seven potential predictors and a second set of variables comprising the precipitation on the EURO-CORDEX area. Only the first PC is considered to summarize spatially a climate variable. Two points motivated this choice. First, the physical/atmospheric variables that make sense as predictors for precipitation downscaling only at the first order have been determined. Hence, a natural choice was to retain only the first PC. Second, applying the SPARSE CCA algorithm over the whole EURO-CORDEX region based on the relatively high-resolution E-OBS dataset is computationally intensive, even for a single principal component. Therefore it has been decided to limit this first exploratory step of SPARSE CCA to only the first PC. The sparsity constraints are tuned so that only one predictor variable appears in each projection vectors (only one non-zero coefficient). Thus, each predictor variable is associated with a rank given by the correlation (see Table 1) with the projected predictand. The representation of some predictors have known issues in some GCMs, in particular R850 and D2 included in this study. The humidity has proven itself to improve the quality of the downscaled precipitation estimates (e.g., Vrac et al. 2007b). Therefore, although some GCMs may have some problems to represent this variable, it was decided to include it among the variables to be tested in the CCA analysis. The outcome of the SPARSE CCA excludes the relative humidity as a predictor. Instead, the dew point temperature

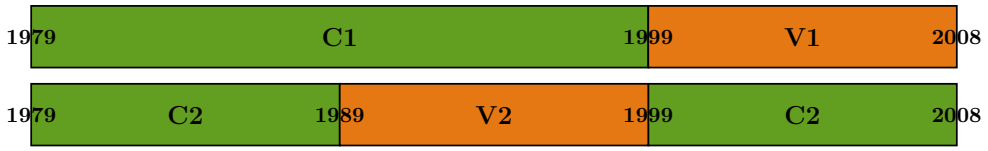

Fig. 2 Cross-validation scheme over the period 1979–2008 with two calibrations (C_1 = [1979–1998] and C_2 = [1979–1988] ∪ [1999–2008]) and two validations (V_1 = [1999–2008] and V_2 = [1989–1998]) periods. The intercomparison is done over the period $V_2 \cup V_1$ = [1989–2008] corresponding to the CORDEX RCM runs period

at 2 m (D2), an index of moisture saturation (Charles et al. 1999), is kept. Although D2 depends on humidity, it also integrates pressure and temperature in its computation, two physical variables that are expected to be relatively well represented by most GCMs. The dew point temperature is then expected to be relatively well represented. The MOS model uses only the large-scale variable as predictor (c.f. Sect. 3.1.5). As precipitation is usually not well represented by the GCMs, this variable is rarely employed as a predictor in Statistical Downscaling Models. Nevertheless, in the present intercomparison exercise since it is aimed at having predictors as common as possible, the large-scale precipitation has also been added in order to have at least one common predictor for all the SDMs. More precisely, to account for the non-Gaussian behaviour of the daily precipitation whose distribution is generally skewed, a transformation of the precipitation data has to be performed before applying PCA. Hence, as in Vrac and Friederichs (2014), the zero precipitation values have been set to a small value different from zero (0.00033) and the logarithm of all precipitation data (with 0's transformed to 0.00033) have been computed.

The SPARSE CCA was carried out over two 6-month periods: a 6-month "summer" (from April, 15th, to October, 14th) and 6-month "winter" (from October, 15th, to April, 14th). Table 1 shows the selected predictors for each season and their order according to the rank given by SPARSE CCA: the first five variables have been selected for each season. For the intercomparison, the first two PCs of each selected large-scale variable are kept as predictors. This choice is made to avoid the optimization of too many parameters since the SDMs calibrations/simulations are pointwise over 6043 E-OBS grid-points. This is a trade-off to keep a relatively low complexity (i.e., a relatively low number of parameters)—especially for the stochastic and TFs models—while including a significant number of physical variables as predictors. The variable selection preprocessing resulted in 12 predictors (2 first PCs for each of the 5 variables selected through the SPARSE CCA and precipitation). For example, this corresponds for the stochastic models, to 39 parameters to be estimated (13 for the occurrences, 26 for the intensity, see Sect. 3.1.3) for each of the 6043 E-OBS grid-points.

2.2 Cross-validation set up

In order to intercompare some SDMs and RCMs involved in the CORDEX exercise, all evaluations have to be made within the constraints of this program, i.e., over the 1989–2008 time period which is the hindcast evaluation time period. Figure 2 sketches the two calibrations (C_1 and C_2) and validations (V_1 and V_2) time periods used in this study for SDMs. The models are trained and validated sequentially, first over C_1 (i.e., [1979–1998]) and V_1 (i.e., [1999–2008]) respectively and then over C_2 (i.e., [1979–1988] ∪ [1999–2008]) and V_2 (i.e., [1989–1998]) respectively. The model evaluation is performed over $V_2 \cup V_1$ = 1989 to 2008, therefore with the outputs of two different calibrations per model. The rain occurrence threshold is set at 1 mm per day for the evaluation. In the literature, a wide panel of thresholds has been used: 0 mm in Semenov et al. (1998), 0.5 mm in Ambrosino et al. (2014) or 5mm in Bouvier et al. (2003). In this study, a middle ground is stroke and a threshold of 1 mm is selected since it is commonly used (e.g., Schmidli et al. 2007).

3 Statistical and dynamical downscaling models

3.1 Statistical downscaling models

One SDM per each of the four families of approaches (TF, WG, WT-based methods and MOS) has been selected—potentially with some variants—in order to evaluate the main philosophical and technical features between the different approaches, e.g., deterministic versus stochastic. Statistical modelling of precipitation is usually divided in two successive steps: first the occurrence and then the intensity. Section 3.1.1 describes rain occurrence modelling and Sects. 3.1.2 to 3.1.5 the different rain intensity models.

3.1.1 Rain occurrence

In this study, two ways to model rain occurrence are considered. In the first way, the model outputs are simply thresholded at a given level (1 mm in that case) from a model including zeros and making no difference between 0's and

positive values. If negative values are generated, they are set to 0. In the other way, rain occurrence at a given location is modelled as a binomial distribution $B(1,p)$ using a logistic regression (LR, see Buishand et al. 2004; Fealy and Sweeney 2007). Let p_i be the probability of rain for the day i conditionally to a N-length predictor (or covariate) vector $\mathbf{X_i}$. The conditional probability of occurrence p_i is formulated through an LR as:

$$\log\left(\frac{p_i}{1-p_i}\right) = \underbrace{P^0 + \sum_{j=1}^{N} P^j X_{i,j}}_{=\mathrm{S}} \tag{1}$$

$$\Leftrightarrow\ p_i = \frac{exp(S)}{1+exp(S)}, \tag{2}$$

where $(P^0, P^1, \ldots, P^N)$ is a vector of coefficients to be estimated. Based on the predictors for day i, Eqs. (1 and 2) provides the probability of rain from which it is easy to simulate a rainfall occurrence. Computational details to estimate p_i are available in "Appendix".

3.1.2 Transfer functions (TFs)

The models belonging to this family link directly the large-scale information to local-scale variables using deterministic functions. As stated in the introduction, those functions characterize the nature (linear or non-linear) of the predictors-predictand relationships. For this approach the Generalized Additive Models (GAM) framework (Hastie and Tibshirani 1990) has been chosen. It is a deterministic model which consists in modelling the expectation of Y (here, the precipitation) conditionally on the N large-scale predictors $(X_1 \ldots X_N)$ as a sum of spline functions $f_j(X_j)$:

$$E(Y|X_1 \ldots X_N) = \sum_{j=1}^{N} f_j(X_j) \tag{3}$$

where f_j are cubic regression spline functions. The cubic splines have a relatively low complexity while allowing a high non-linearity to model the link between X_j and Y, i.e., the large- and the local-scale data. This method has been applied for the present time period, for instance to downscale the near surface wind fields in Salameh et al. (2009), or for the Last Glacial Maximum time period (-21 ky), to retrieve monthly climatology for temperature and precipitation over Europe (Vrac et al. 2007a) or global permafrost (Levavasseur et al. 2011). GAM is a data-driven approach in the sense that it allows to model both piecewise linearities and non-linearities depending on the nature of the predictor-predictand dependence. Two variants have been defined in the present study: (1) GAM and (2) GAM-so. In the first one, GAM has been calibrated with all values (i.e., including 0's) and then rain intensity has been directly simulated and the rain occurrence is dealt by thresholding the outputs at 1 mm. In the second one (i.e., the GAM-so approach), the LR is first used to model the occurrence and then $E(\log(Y)|X_1 \ldots X_N)$ (instead of $E(Y|X_1 \ldots X_N)$) has been modelled for positive rain intensities. Computational details for GAM simulations are available in "Appendix".

3.1.3 Stochastic weather generator (WG)

WGs are models generating daily weather scenario thanks to pdfs estimated from observations. As previously stated they are mainly used to simulate data whose statistical properties are similar to those of observations. They present a large diversity in terms of techniques and complexity: starting from quite simple series generators (e.g., Semenov and Stratonovitch 2010), passing through Markov chain based models (e.g., Kilsby et al. 2007) to sophisticated approaches like the observed hierarchical organization of rainfall and rain-cell space and time-clustering processes (e.g., Onof et al. 2000). One way to build a stochastic SDM is based on generalized linear models (GLMs). GLMs have been first applied by Stern and Coe (1984) for the generation of precipitation. GLMs link the expected mean of a random variable to the N predictors as:

$$g(\mu) = \sum_{j=1}^{N} \theta^j \cdot X_j \tag{4}$$

where μ is the expected mean, θ^j are regression coefficients to be estimated and $g(\cdot)$ a monotonic link function. In this work, an extension of this formulation is used. Conditional pdfs are used to model the precipitation in a Vectorised Generalized Linear Models (VGLM) framework as in Chandler and Wheater (2002). It means that the distribution family is fixed and the distribution parameters are estimated by a GLM. Thus, the rain distribution parameters for each day are estimated from the selected predictors. This method allows also the simulation of spatio-temporal rainfall with an appropriate covariance function (Yang et al. 2005) or at subdaily temporal resolution (Mezghani and Hingray 2009). In all those works a two-step approach is implemented to model precipitation. It stands as follows:

1. Rain occurrence is modelled by an LR as given in Eqs. (1 and 2),
2. Rain intensity is supposed to follow a Gamma distribution $\Gamma_{\alpha,\beta}(\cdot)$ whose parameters shape α and rate β are functions of the large-scale predictors at day i:

$$\begin{cases} \log(\alpha_i) &= \alpha^0 + \sum_{j=1}^{N} \alpha^j X_{i,j} \\ \log(\beta_i) &= \beta^0 + \sum_{j=1}^{N} \beta^j X_{i,j} \end{cases} \tag{5}$$

Hence, for each day the parameters are calculated and a distribution is retrieved, which makes the model non-stationary and able to evolve with predictors. Then, simulations are performed based on the daily pdf. Note that the Gamma distribution parameters have been estimated from all values above 0 mm but only rain amounts above 1 mm are simulated. Indeed, estimating the Gamma distribution for values above 1 mm makes the hypothesis of a Gamma to simulate rain intensity no longer valid. Besides, calibrating the model over precipitation amounts above 1 mm causes an artificial increase of the variability of the generated time series: the variance is about twice the variance of the data generated from a calibration with all positive precipitation (not shown). Computational details to infer the Gamma distribution parameters are available in "Appendix".

In the following, two variants are applied and tested: (1) SWG, the non-stationary model described above in Eqs. (1 and 5), and (2) SWG-s, the stationary version of SWG, where the occurrence probability and the parameters α and β are constant and do not depend on any predictor. All the parameters of both variants are estimated by maximizing the likelihood function, except the constant occurrence probability (hereafter referred to as COP), which corresponds to the observed occurrence.

3.1.4 Weather typing (WT)

The WT family is based on large-scale circulation (LSC) patterns. It relies on the idea that the same LSC situation (i.e., predictors) produces the same local-scale effects (here E-OBS rain fields). WT consists in regrouping days with similar LSCs. This is classically done with statistical clustering methods: given a number K of clusters and a measure of similarity, data (here daily situations) are grouped in K clusters such that the situations into a given cluster are as similar as possible, while situations in different clusters have to be very different. The clustering methods are widely used to study weather regimes (e.g., Yiou 2004; Vrac et al. 2014, and the references therein). In terms of SDMs those methods are rather used to condition statistical models, for example a stochastic model as in Schnur et al. (1998), Bellone et al. (2000) or Vrac et al. (2007b).

In this study, the analog method is employed as representative of the WT family. This method considers each day as a cluster. A deterministic analog modelling as defined in Yiou et al. (2013) has been chosen here. It has been used in several previous studies (Zorita and Storch 1999; Yiou et al. 2007; Vautard and Yiou 2009; Chiriaco et al. 2014). It consists in determining for a given day to be downscaled in the validation period the day in the calibration period which has the closest atmospherical situation. It is determined by a similarity metric between the predictor set of the day to be downscaled (X_V) and the predictor set of the day in the calibration period (X_C). This approach is quite flexible to change the distance or the temporal window of the situations (Yiou 2014). Many family of metrics can be used (e.g., Grenier et al. 2013) and one of them is distance:

$$\mathrm{Day}_{\in C} = \mathrm{argmin}_{\mathrm{day}_{\in V}}(\mathrm{dist}(X_V, X_C)). \quad (6)$$

The Euclidian distance is chosen in this study. Only one experiment has been set and is called ANALOG. Note that one important difference with the other models is that this method is applied over the entire predictor dataset anomalies, not only over the first two PCs. Hence, much more information than for the other models has been provided to this model, this will be discussed in Sect. 5.2. A threshold at 1 mm has also been applied to the output values for rainfall occurrences.

3.1.5 Model output statistics (MOS)

This approach regroups all the "quantile-mapping" related methods, more precisely all the methods relating the large-scale CDFs to the local-scale CDFs. For instance quantile-quantile based methods have been widely used for downscaling (e.g., Vrac et al. 2012 and references therein) or to correct bias in model outputs thanks to observations CDFs (e.g., Gudmundsson et al. 2012, and references therein) and the correspondences between predictors and predictands quantiles. Those methods can be directly calibrated on models outputs (e.g., GCM or RCM). Those correspondences can be based on non-parametric (Déqué 2007) or parametric (Piani et al. 2010) models. Many methods have been implemented and compared in Gudmundsson et al. (2012). The MOS technique used here is the "Cumulative Distribution Function-transform" (CDF-t) initially developed in Michelangeli et al. (2009) to downscale wind and applied later to temperature and precipitation, for example in Lavaysse et al. (2012), Vrac et al. (2012) and Vigaud et al. (2013).

The CDF-t model consists in relating local-scale (i.e., here E-OBS precipitation) CDF to the large-scale (i.e., here ERAi reanalysis precipitation) CDF. The CDF-t and quantile-quantile methods are similar in philosophy, except that CDF-t takes into account the change in the large-scale CDF from the calibration to the projection (or validation) time period, while quantile-quantile projects the simulated large-scale values onto the historical CDF to compute and match quantiles. Let $F_{\mathrm{Rc}}(x)$ and $F_{\mathrm{Ec}}(x)$ define respectively the rain CDFs from the Reanalyses (subscript R) and from E-OBS (subscript E) over the calibration period (subscript c) and $F_{\mathrm{Rv}}(x)$ and $F_{\mathrm{Ev}}(x)$ the CDFs over the validation period (subscript v). An estimation of $F_{\mathrm{Ev}}(x)$ is assumed to be:

$$F_{\mathrm{Ev}}(x) = F_{\mathrm{Ec}}\left(F_{\mathrm{Rc}}^{-1}(F_{\mathrm{Rv}}(x))\right), \quad (7)$$

Table 2 Statistical Downscaling Models features concerning the occurrence model (LR: logistic regression, COP: constant occurrence probability, T: thresholded) and the predictors (Anom.: anomalies over all the variables, ERAi PR: ERAi reanalyses precipitation, 6 × 2 PCs: the first two PCs of the five selected predictors and precipitation) used in each case

Statistical models features											
Weather pattern based		MOS bias correction		Transfer function				Stochastic weather generator			
ANALOG		CDFt-so		GAM		GAM-so		SWG		SWG-s	
Occ.	Pred.	Occ.	Pred.	Occ.	Pred.	Occ.	Pred.	Occ.	Pred.	Occ.	Pred.
T	Anom.	LR	ERAi PR	T	6 × 2 PCs	LR	6 × 2 PCs	LR	6 × 2 PCs	COP	6 × 2 PCs

with x in the range of the physical variable of interest. Thus, the local-scale CDF over the validation period, F_{Ev} is obtained from the large-scale CDF F_{Rv} over the validation period, on which a transformation T defined from the CDFs over the calibration period, $T(u) = F_{\mathrm{Ec}}\left(F_{\mathrm{Rc}}^{-1}(u)\right)$ is applied. Then, a quantile mapping between $F_{\mathrm{Ev}}(x)$ and $F_{\mathrm{Rv}}(x)$ is performed to retrieve the precipitation values at local scale. More detailed information, descriptions and evaluations of CDF-t are available in Vrac et al. (2012). CDFt-so is the only experiment set for this approach. In the same way as GAM-so, rain amount is modelled by CDF-t and rain occurrence by the LR. Because the ERAi precipitation presented too few days with precipitation amounts above 1 mm, CDF-t has been calibrated over precipitation above 0 mm. Indeed, too few rainy days (rain above 1 mm in that case) at the large-scale will produce too few rainy days in the downscaled data (not shown). That is why the calibration has been made for days above 0 mm and then the outputs have been thresholded at 1 mm. This model is the one which has the lowest quantity of information in terms of predictors: the large-scale precipitation only. Indeed the other models have six variables with precipitation among them. Computational details on CDF-t are available in "Appendix".

3.2 Regional Climate Models

Concerning dynamical models, five runs have been selected: two from EURO-CORDEX and three from MED-CORDEX experiment. These simulations cover two different domains (Fig. 1) but use the same horizontal resolution (0.44°) and are all initialised and forced at their boundaries by ERAinterim data. None of the models uses nudging inside the domain except IPSL-WRF311. A relaxation region of different widths (a few hundreds of km, depending on the model) is used to account for boundary imbalance effects. The common period of simulation is 1989–2008 and each model uses its own set of parameterizations. Details on each run can be found in the following references: Flaounas et al. (2013) for IPSL-WRF311, Nabat et al. (2014) for CNRM-ALADIN52 (see also Colin et al. 2010; Herrmann et al. 2011), Domínguez et al. (2013) and Jiménez-Guerrero et al. (2013) for UCLM-PROMES and Table 1 of Vautard et al. (2013) for IPSL-INERIS44 and for CNRM-ARPEGE51. These models are hereafter referred to as MED-IPSL, MED-CNRM, MED-UCLM, EURO-IPSL and EURO-CNRM respectively. As indicated in Table 3, CNRM and UCLM models repeat the year 1989 two or three times to take into account the spin-up associated to the surface scheme initialization. This is widely sufficient to equilibrate moisture in the levels of the soil that interacts with the atmosphere through evapotranspiration. Repeating the year 1989 two or three times is considered as negligible in the final results. IPSL models do not repeat this year but this does not influence the results. Indeed, year 1989 has been tested and similar behaviour compared to other years has been observed and several tests have shown that simulations were converging after a few days. Moreover, the use of nudging for the MED-IPSL simulations reduces the spin-up period. Besides, this investigation is beyond the scope of this paper.

Tables 2 and 3 summarize all the models (SDMs and RCMs) and their features.

4 Intercomparison results

The quality of the simulations is assessed by comparison to the data product considered as pseudo local-scale observations (E-OBS) in terms of rain occurrence and intensity, as well as spatial and temporal properties through selected indicators. In the view of the relatively equivalent results over the two seasons, only the results over the "summer" season will be presented hereafter. Besides, even if impacts studies generally need annual precipitation data, impact studies focusing for example on agricultural impacts, heatwaves or droughts studies need accurate precipitation data during spring and summer. Intense precipitation events around the Mediterranean usually take place between mid August and mid November and cause floods. Precipitation during winter is easier to model by the RCMs because of the stratiform nature of precipitation, whereas summer

Table 3 Dynamical downscaling models features

Dynamical models features					
Region	EURO-CORDEX		MED-CORDEX		
Model	WRF-IPSL-INERIS44	CNRM-ARPEGE51	IPSL-WRF311	CNRM-ALADIN52	UCLM-PROMES
Institute	IPSL and INERIS	CNRM	IPSL	CNRM	UCLM
Number of levels	32	31	28	31	36
Radiation scheme	RRTMG, Iacono et al. (2008)	Morcrette (1990)	RRTMG, Iacono et al. (2008)	Morcrette (1990)	ECMWF (2004), with fractional cloud cover Chaboureau and Bechtold (2002, 2005)
Convection scheme	Grell and Dévényi (2002)	Bougeault (1985)	Kain (2004)	Bougeault (1985)	Kain and Fritsch (1993)
Microphysics scheme	Hong et al. (2004)	Ricard and Royer (1993)	WSM 5-class Hong and Lim (2006)	Ricard and Royer (1993)	Hong et al. (2004)
Land-surface scheme	NOAH, Ek et al. (2003)	Douville et al. (2000)	RUC Smirnova et al. (1997)	Douville et al. (2000)	Krinner et al. (2005)
Boundary layer scheme	YSU, Hong et al. (2006)	Ricard and Royer (1993)	YSU, Hong et al. (2006)	Ricard and Royer (1993)	Cuxart et al. (2000)
Soil initialization and spin-up	Soil initialized from ERAInterim. No spin up	Year 1989 is played twice	Soil initialized from ERAInterim. No spin up	Year 1989 is played twice	Soil initialized from ERAInterim. Year 1989 is played three times
Land use	USGS Land Use	ECOCLIMAP	USGS Land Use	ECOCLIMAP	Assigned by the land-surface scheme

rainfalls are driven by convective rain processes, more difficult to represent and resulting from a parametrization in the RCMs. All the indicators are computed over the 1989–2008 period. Results specific for the "winter" season will be described and the corresponding figures are available as auxiliary material. In the following section, most evaluations are presented in terms of bias of the indicators with respect to those of the pseudo-observations defined as "Indicator(simulation) minus Indicator(observation)". In terms of colours, blue means that the model underestimates and orange/red means that it overestimates the considered criterion with respect to the observations.

4.1 Occurrence indicators

The evaluation begins by exploring the ability of the models to reproduce the occurrence properties: Do the models respect the observed proportions of wet or dry days and the time they occur? In this part, only nine models are considered for occurrence evaluation (ANALOG, GAM, LR, COP, EURO-CNRM, EURO-IPSL, MED-CNRM, MED-IPSL and MED-UCLM) since SWG, CDFt-so and GAM-so share the same LR occurrence model presented in Eq. (1).

First, bias (in %) of wet days frequency have been investigated in Fig. 3. The LR, COP and ANALOG models perform well. They show biases close to zero with very small positive or negative values and distributed over all the area. All the other SDMs and RCMs are strongly biased. Most of them are mainly positively biased which is a well known problem for RCMs: the models produce little rainfall amount too often (see Sun et al. 2006; Stephens et al. 2010). The negative bias of MED-IPSL is due to land surface/atmosphere feedbacks that are not well reproduced generating dry soil too early in spring over most Western Europe then less clouds and precipitation and higher temperature in summer. Except ANALOG, LR and COP, all the models are globally producing rainfall too frequently. Both IPSL RCMs show patterns at the borders of the domain. This is a consequence of the relaxation zone at the domain boundaries. Similar patterns are observed on Figs. 6 and 7 for the same reason. Note the very poor performance of GAM which largely overestimates the percentage of rainy days.

For the winter season the results are more or less the same for all the models except for the EURO-CNRM model where the biases are smaller and distributed in terms of sign all over the domain. MED-IPSL presents also some interesting differences. The biases evolve from negative at the south–west to positive at the north–east of the domain (see auxiliary materials). This gradient is a consequence of a humidity bias observed in winter in the model (compared to GPS measures). Indeed, there is a light positive

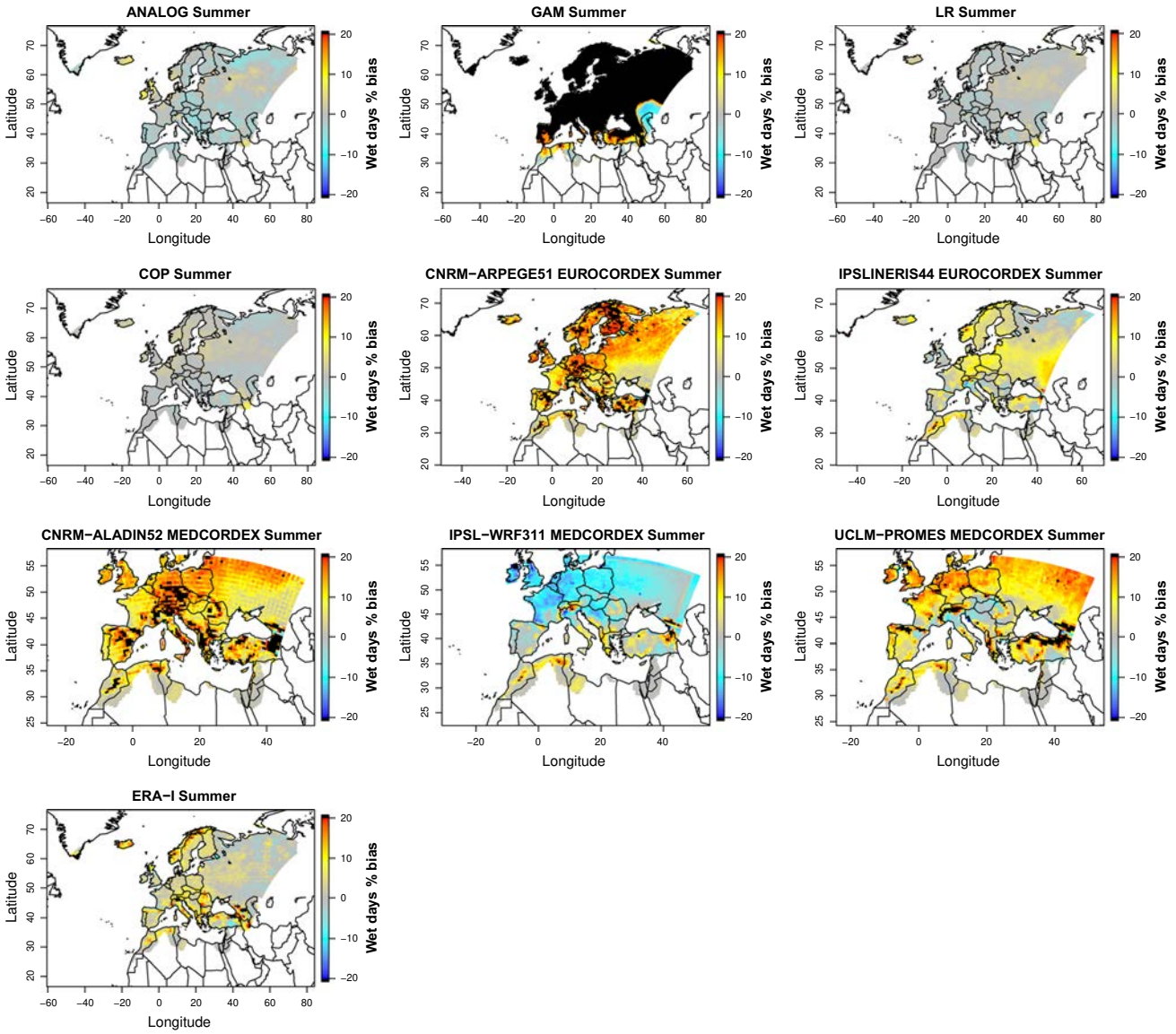

Fig. 3 Wet day frequency bias in percentage (%). Some bias values have been saturated to ease the reading and to improve the *colour contrast*: *black* means that the bias values exceed 20 % in absolute values. The real range of values for each map is given in auxiliary material table 3

humidity bias in Western Europe and it increases when going eastward. One explanation is that the microphysics scheme is not efficient enough for precipitation and can induce a lower precipitation amount for a given humidity rate. Besides, in winter the air mass flows from west to east which also increases the humidity and therefore the precipitation in the east.

Periods of consecutive wet or dry days (or spells) have been also considered, in particular the mean length of the wet spells and dry spells biases (expressed in days). In other words, the mean wet and dry persistence biases are investigated. They are pictured by boxplots respectively in Fig. 4a, b. In order to remain consistent with the domains presented in the maps of Fig. 1, they are computed over different domains. These boxplots are nevertheless relevant since, when the indicators are calculated only over the MED-CORDEX domain for all the models, the ranking of the models and the global aspects of the boxplots are similar (not shown). All the models except GAM show skills for reproducing the wet spells of E-OBS, especially the EURO-IPSL and ANALOG models. On the opposite, GAM is strongly biased. Interestingly, although not perfect, the LR occurrence provides better results than the constant occurrence probability (COP) approach. In other words, the non-stationarity brought by the logistic regression improves the wet occurrence modelling compared to the stationary COP model. Concerning the mean dry spells, the models uniformly underperform (i.e., they present larger

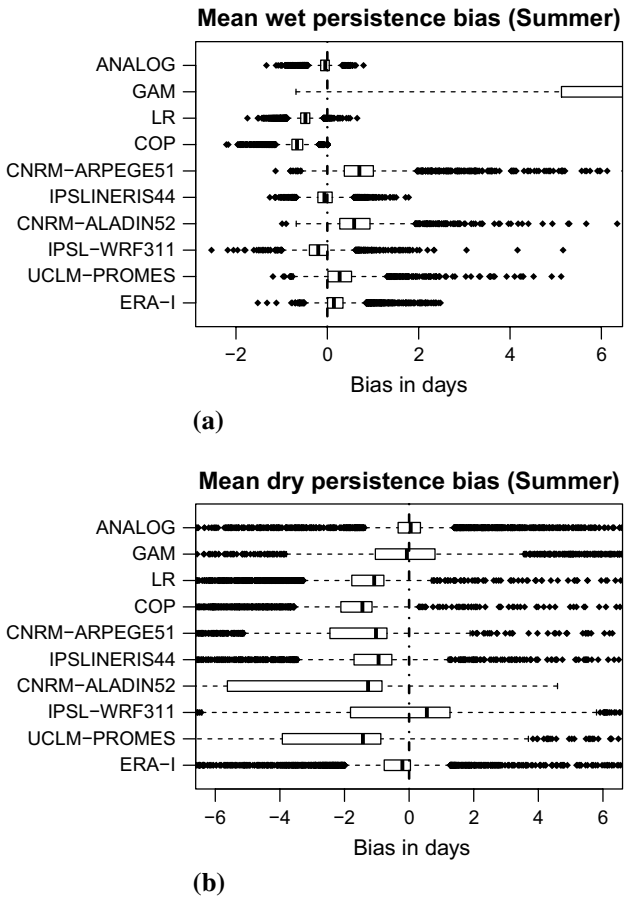

Fig. 4 Mean wet and dry spells length bias in days for the summer season. The *left* and *right fences* are situated at 1.5 times the interquartile range from the *box* and the *dots* are the values beyond these fences. Bias values above 6 days or below −6 days have been ignored in these plots for readability reasons (c.f. auxiliary material for the real range). **a** Wet spells bias. **b** Dry spells bias

biases) than for mean wet spells. They all have also difficulties to reproduce dry spells around the Mediterranean (not shown). They mainly underestimate the mean dry spell length except for MED-IPSL. The mean wet and dry spells biases do not cancel each other even if the MED-CORDEX models, GAM and EURO-CNRM show opposite bias signs between wet and dry spells. In other words, a deficit (or an excess) of the wet days persistence does not necessarily imply an excess (or a deficit) of the dry days persistence. In winter, the results are similar except that the mean wet spells biases absolute values are smaller for all the models (see auxiliary materials).

Until now, the rain occurrence has been tested only in terms of frequencies. In order to characterize the time synchronicity of the rainy events, the Brier score (hereafter referred to as BS, Brier 1950) is computed. The BS describes how close to the daily observed occurrences the daily estimated probabilities are:

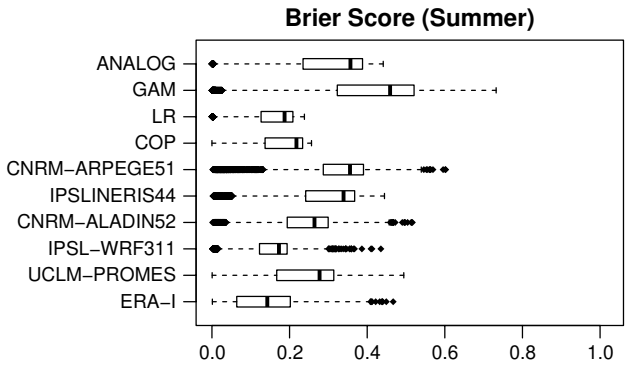

Fig. 5 Brier score: the scores *boxplots* of each model are represented between 0 and 1. Models with scores close to zero have good estimation of the daily rain occurrence probabilities

$$BS = \frac{1}{N}\sum_{t=1}^{N}(p_t - o_t)^2, \tag{8}$$

where p_t is the estimated probability at the time t from LR and 1 or 0 for deterministic models for rain or no rain respectively, o_t is the observed occurrence in observation at time t which takes the values 1 or 0 (meaning rain or no rain) and N is the number of days. Hence, the closer the score to 0, the more synchronized the model is. Figure 5 shows the scores computed for each model. LR and MED-IPSL have the smallest values, on average below 0.2. The other models, except GAM and ANALOG, have a BS on average below 0.4. Note that the Analog approach has better results in terms of rainy days proportion than for the timing of rainfall events. This means that the Analog model produces sequences of wet or dry days with correct proportions but not at the right moment. In winter (see auxiliary materials), the results are similar.

4.2 Intensity indicators

The statistical properties of the downscaled rain intensity at individual grid-points of the 11 models are now compared to those of the observations. Figure 6 shows mean daily precipitation biases (in mm) for the precipitation above 1 mm. The average rain amounts are well represented by SWG, SWG-s, ANALOG, CDF-so models. CDFt-so shows small positive biases over almost the whole domain while SWG, SWG-s and ANALOG models present small positive and negative biases distributed all over the domain. GAM, GAM-so and the dynamical models are more or less strongly positively and negatively biased. MED-IPSL is the best among them with positive and negative biases distributed all over the domain which is also the case for MED-UCLM. EURO-CNRM, MED-CNRM present mostly negative biases while EURO-IPSL has mostly positive ones.

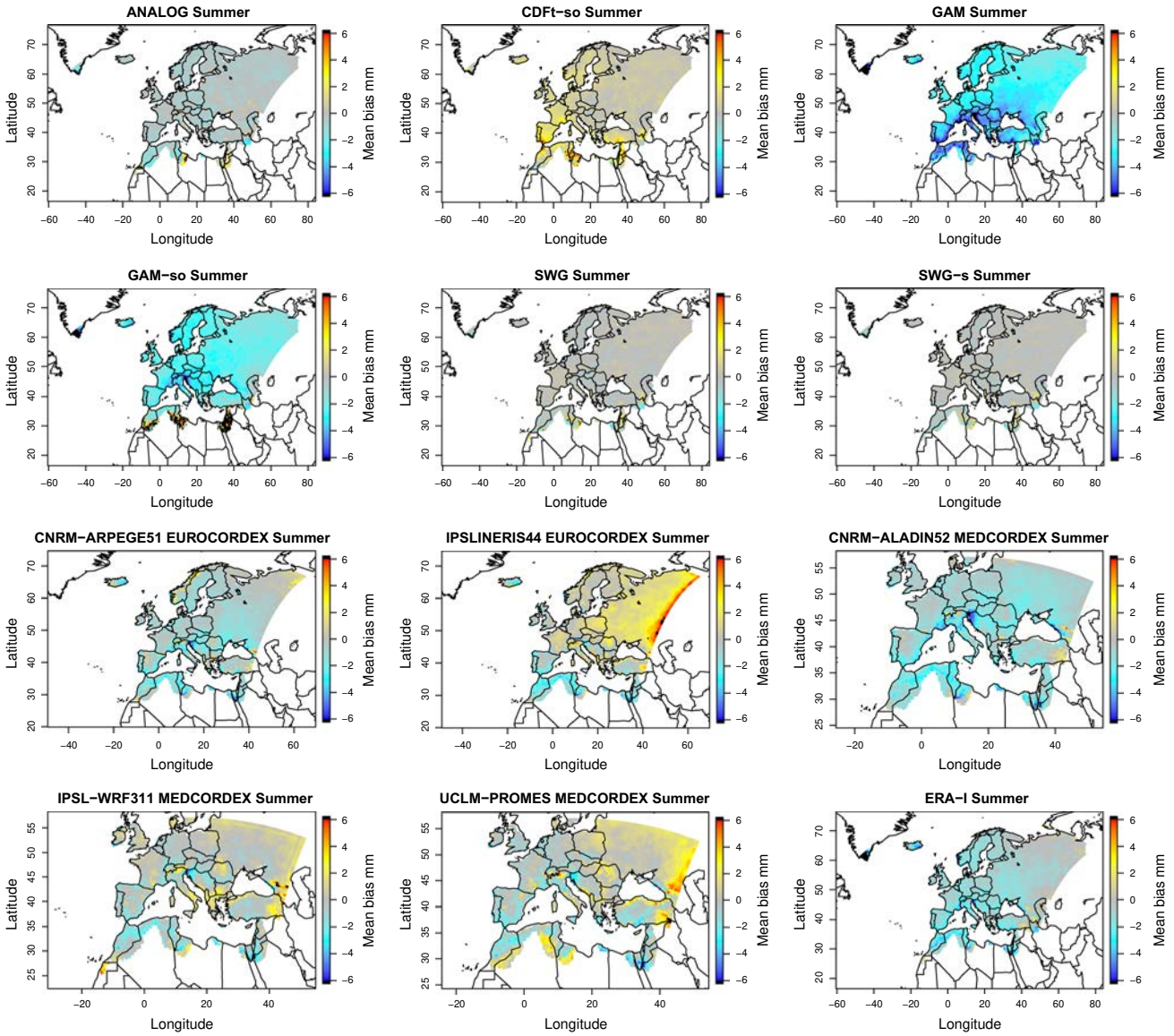

Fig. 6 Mean precipitation bias in mm $\cdot$ day^{-1}. Bias values above 6 mm or below −6 mm have been saturated in *black* in order to have a good *colour contrast* (c.f. auxiliary material for the real range)

Border patterns are visible for all the RCMs which are a consequence of the relaxation zone. Similar results are found for winter although with smaller biases for all the models except MED-IPSL (see auxiliary materials).

Figure 7 displays the variance ratio (in percentage). It is the ratio between the variance of the simulations and that of the observations:

$$\%rv = \frac{\sum_{i=1}^{n}(S_i - \overline{S})^2}{\sum_{i=1}^{n}(O_i - \overline{O})^2} \times 100, \tag{9}$$

with S_i is the simulated value for day i, O_i is the observed value at day i, $\overline{O}$ is the mean of the observations for the period, and $\overline{S}$ is the mean of the simulated data. While CDFt-so performs well with some variations and mostly overestimates the variance over the area, ANALOG, SWG and SWG-s tend to underestimate the variance. For the stochastic models, it is caused by the way rain amounts have been simulated. Indeed, the SWG and SWG-s models have been forced to simulate precipitation above 1 mm which can reduce the variability of the generated data. Once again GAM and GAM-so perform poorly. While the other SDMs reach an average ratio between 80 and 150 %, GAM and GAM-so barely reach 25 % and are the worst among all models. Concerning RCMs, CNRM models are the best among them although they mainly underestimate the variance around 80 %. Others are much more biased and mostly overestimate it. Their variance ratios are above

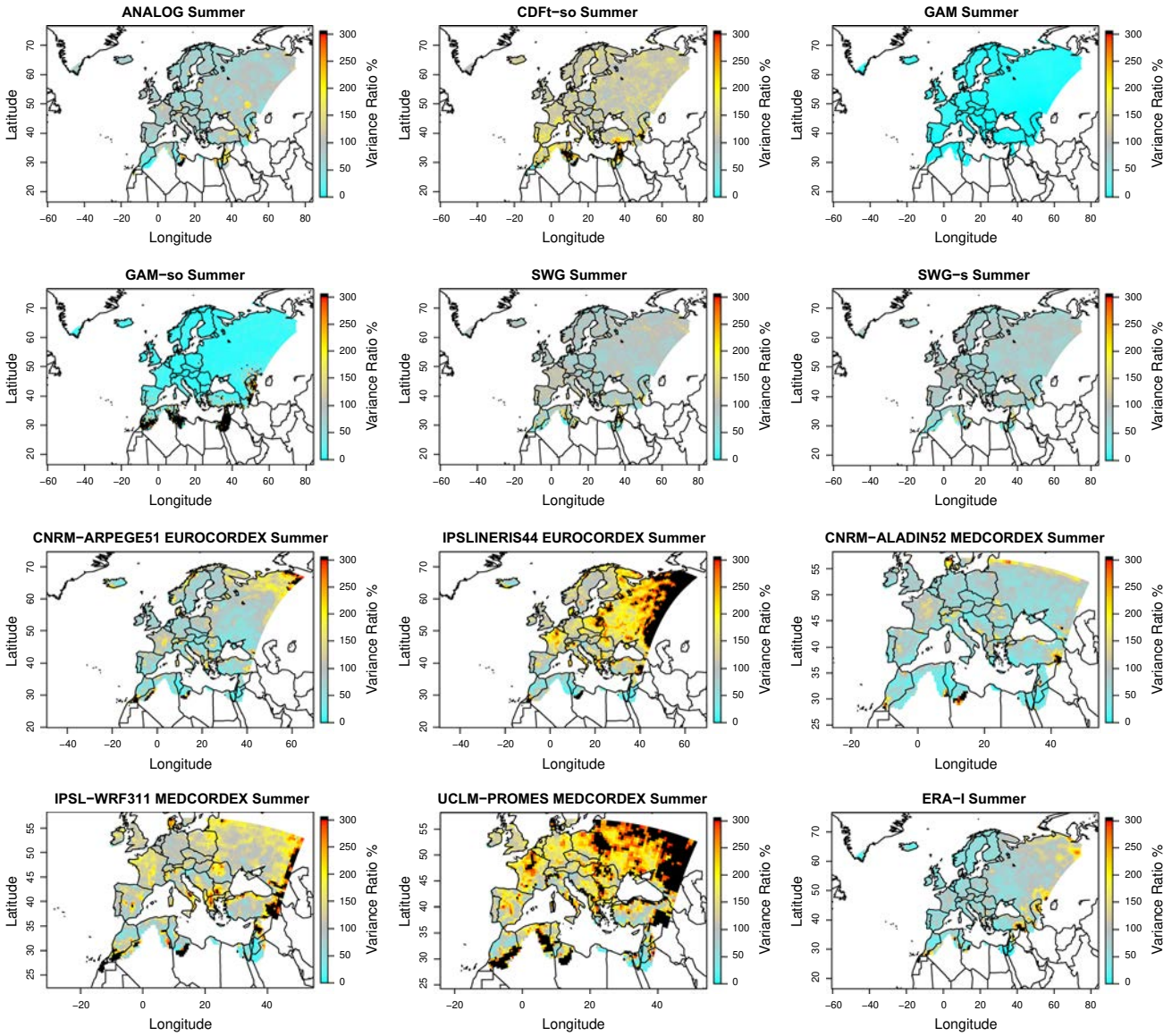

Fig. 7 Variance ratio in percentage (%). Values above 300 % have been saturated in *black* for the same reason as previously (c.f. auxiliary material for the real range)

150 %. Here, the patterns at the boundaries for RCMs are stronger than for the previous indicators. In winter, the SDMs have the same behaviour unlike RCMs. CNRM models and MED-IPSL present variance ratio larger than for summer: closer to 100 % for CNRM models and ratios above 150 % for MED-IPSL (see auxiliary materials).

As a last indicator of marginal intensity, the reproduction of extreme values is investigated. The 99th quantile bias (in mm) is considered and shown in Fig. 8. Overestimation and underestimation patterns are quite similar to those observed for the variance ratio (see Fig. 7) transposed to the 99th quantile bias i.e., biases are quite similarly distributed all over the area (not shown). Thus, similarly to the variance ratio ANALOG, CDFt-so, SWG and SWG-s are good to reproduce extremes. ANALOG, SWG and SWG-s slightly underestimate the 99th quantile, while CDFt-so overestimate it. Note that MOS models like CDF-t may be unstable to simulate extremes especially for future projections. In order to deal with this issue, the constant correction method defined in Déqué (2007) is used in CDF-t. The underestimation for SWG and SWG-s results from the marginal Gamma pdf used here which is not able to reproduce correctly the extremes. This is something known and investigated in literature (e.g., Vrac and Naveau 2007). GAM and GAM-so reach a median bias below -10 mm and therefore widely underestimate the 99th quantile. RCMs over- or underestimate depending on the model. CNRM RCMs present mostly negative biases and the others positive biases.

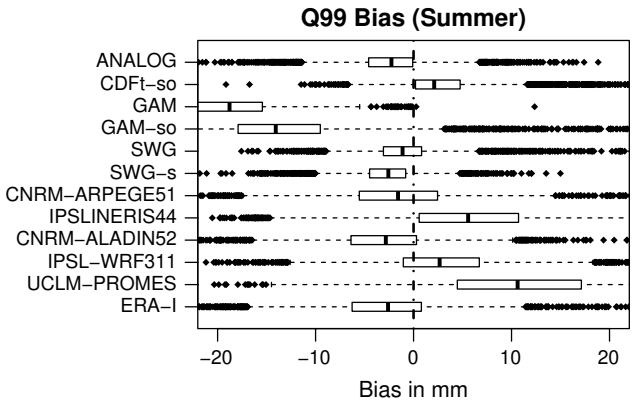

Fig. 8 99th quantile bias in mm. Bias values exceeding 20 mm in absolute values are not represented (c.f. auxiliary material for the real range)

In winter (see auxiliary materials), the results are similar except that biases are smaller in absolute values. The only remarkable difference is for CNRM models which present mostly positive 99th quantile bias.

4.3 Spatial indicators

The spatial properties of the downscaling models, more precisely the spatial variability are now evaluated. To this end, a PCA is performed on daily downscaled precipitation outputs for each of the 11 models and on E-OBS data. Figure 9 pictures the first summer EOF of E-OBS and of each model. Since the distribution of precipitation is skewed, and therefore non-Gaussian, a transformation of the precipitation data has to be performed before applying a PCA. Here, the approach suggested by Vrac and Friederichs (2014) is followed: the zero precipitation values have been set to a small value different from zero (0.00033) and we then computed the logarithm of all precipitation data, with 0's transformed to 0.00033. The PCA is actually performed on those transformed precipitation outputs. The variance explained by the first EOF is indicated for each model. Even if the values are generally low (mostly around 10 %), in the present case, it is a valuable tool to spatially compare modes of variability. The EOF coefficient characterizes the contribution of each grid-point to the variability explained by a PC. The aim is to see if the EOF values for each model have the same spatial distribution as for E-OBS. Similar patterns means that the models have a good ability to reproduce the spatial variability of the observations. The ANALOG model has almost the same spatial structure as the observations. This was expected since ANALOG is based on a resampling procedure and therefore keeps the spatial structure. The other statistical models have quite different spatial patterns even if CDFt-so, GAM and SWG are quite close. In some cases, they even present "flat" spatial patterns (i.e., EOF coefficients are almost equal). The "flat" spatial patterns come from models that are not able to reproduce any spatial variability in their simulations. That is the case for GAM-so and SWG-s for example, whose simulations are made pointwise without spatial constraints. EURO-CORDEX models well reproduce the observation pattern whereas MED-CORDEX models. In winter (see auxiliary materials), the spatial variability of all the models is better caught than in summer, except for GAM-so and SWG-s again. It is probably a consequence that the rain processes involved are different depending on the season. In winter the precipitation is stratiform or dynamic which is related to large-scale atmospheric system. In summer, the precipitation relies on convective processes (i.e. isolated storms for instance) which have a complex spatial structure.

The pattern correlation of the daily maps has also been investigated. It was computed between the previously transformed precipitation outputs used to compute the EOF and the transformed E-OBS. In Fig. 10, the boxplots of daily pattern correlation are given. RCMs—which are spatially constrained—are better than SDMs. Even ANALOG, which is considered as efficient for reproducing the spatial variability, fails in reproducing daily spatial pattern. It is consistent with the result given by the Brier score which indicates that ANALOG fails in terms of synchronicity of the events. The best model is the MED-IPSL model; this might be explained by the fact that it is nudged. Note that ERAi presents the best pattern correlation with E-OBS, with the exception of MED-IPSL. Even if MED-IPSL model is nudged with ERAi, it seems to improve the pattern correlation of MED-IPSL with E-OBS.

4.4 Temporal indicators

The temporal aspect is studied through two angles: by studying the interannual variability and studying the seasonality. Naturally these indicators are examined over the whole year.

In Fig. 11, the cumulated annual rain amount over two illustrative stations (see Fig. 1 for their location): Montpellier (Fig. 11a) and Moscow (Fig. 11b) is represented. The top panels display the E-OBS amounts, all the statistical models and ERAi, while the bottom panels show the results from the dynamical models and E-OBS. The reanalysis precipitation is plotted since it is the only predictor of CDFt-so. First, the case of Montpellier is considered: among statistical models all deterministic models (in purple) except GAM-so seem to be better than the stochastic models (in green) to reproduce the inter-annual variability. GAM-so annual amounts are low because of the combination of LR

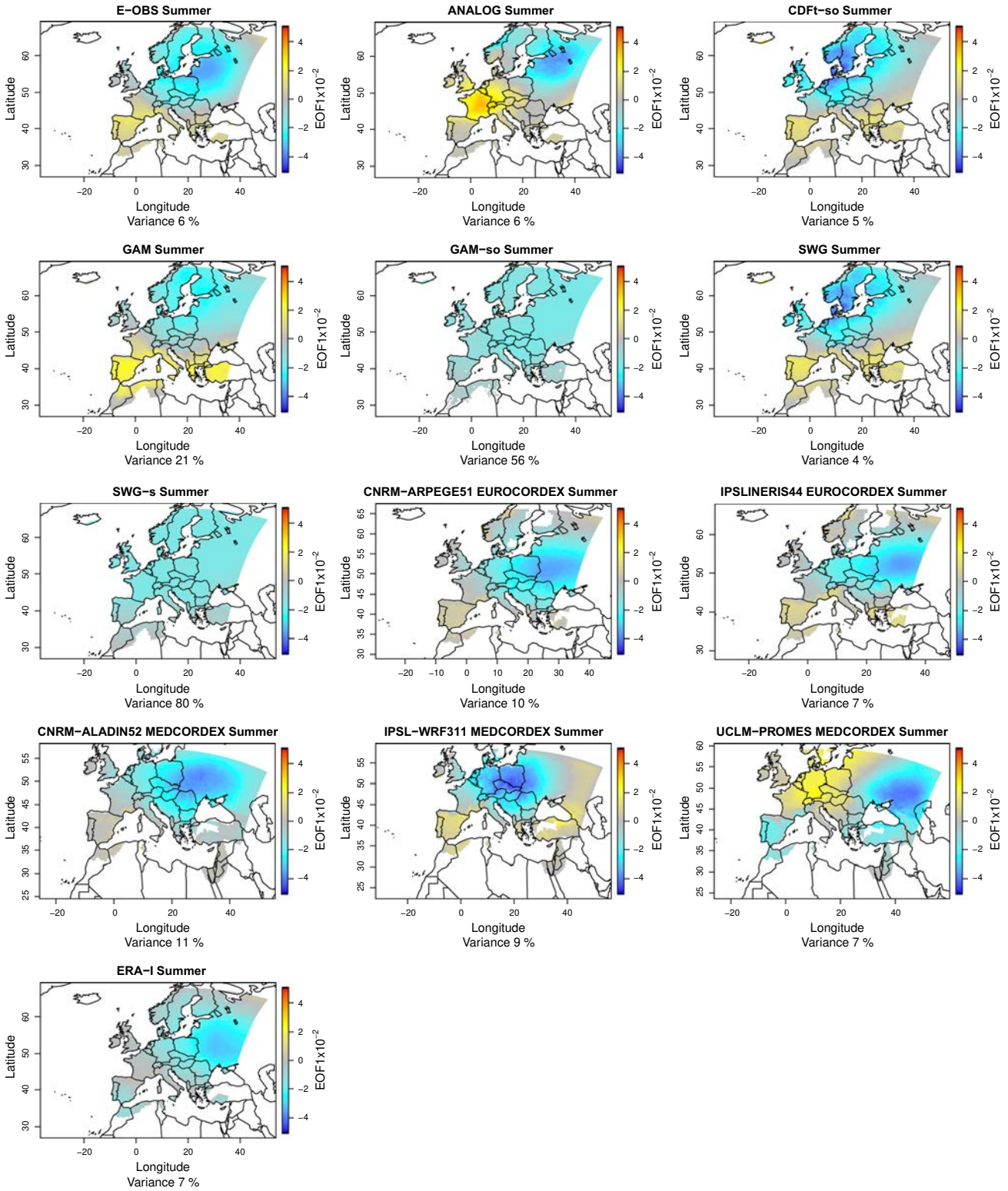

Fig. 9 Spatial distribution of the first EOF loadings of each model. The variance explained by each EOF is indicated for each model

to model rain occurrence and GAM to model rain intensity. The latter is designed to simulate the average rain amount but the random trial for the rain occurrence reduces the annual amount. The dynamical models are better than the statistical models for the inter-annual variability (except ANALOG and CDFt-so).

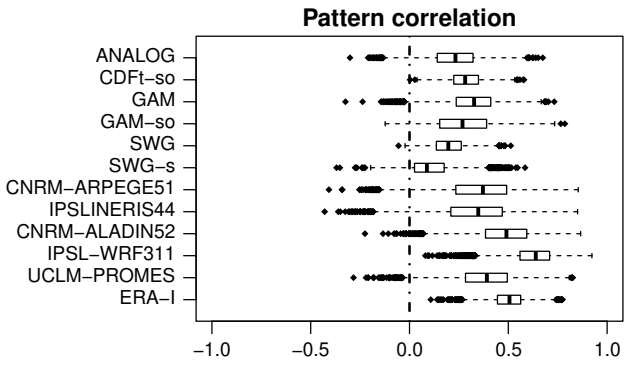

Fig. 10 Pattern correlation of daily maps between each model and E-OBS

For Moscow, the evaluation result is quite different. In this case, no SDMs seem to reproduce the inter-annual variability of the observations. As for Montpellier, low annual rainfall amounts are observed for GAM-so. Almost all dynamical models overestimate precipitation for this station except EURO-CNRM which is particularly close to E-OBS in this case. In order to have a more global overview over the domain, the correlation between cumulated annual rain amount time series of each model and that of E-OBS have been computed pointwise. The boxplots of the correlations are available in Fig. 12. Obviously the SDMs have difficulties to reproduce the inter-annual variability compared to the RCMs except the CDFt-so whose predictor is ERAi total precipitation (c.f. the boxplot of total precipitation above 1 mm of ERAi in Fig. 12). The performance of the other SDMs is poor (with correlation from 0.2 to 0.4) and the stochastic models and ANALOG have the worst performance while they were the best for occurrence and intensity marginal properties. RCMs are more satisfactory, especially the EURO-CNRM and MED-IPSL models. However, these results have to be considered carefully because they characterize the year-to-year synchronisation of the variability i.e., if the variations of the annual amount increase or decrease at the same moment. In terms of RMSE (given in auxiliary material) SDMs are better than RCMs (except GAM-so) as already suggested by the evaluation of the Brier score. This observation does not stand for EURO-CNRM, which is good in terms of RMSE and correlation but not for the Brier score.

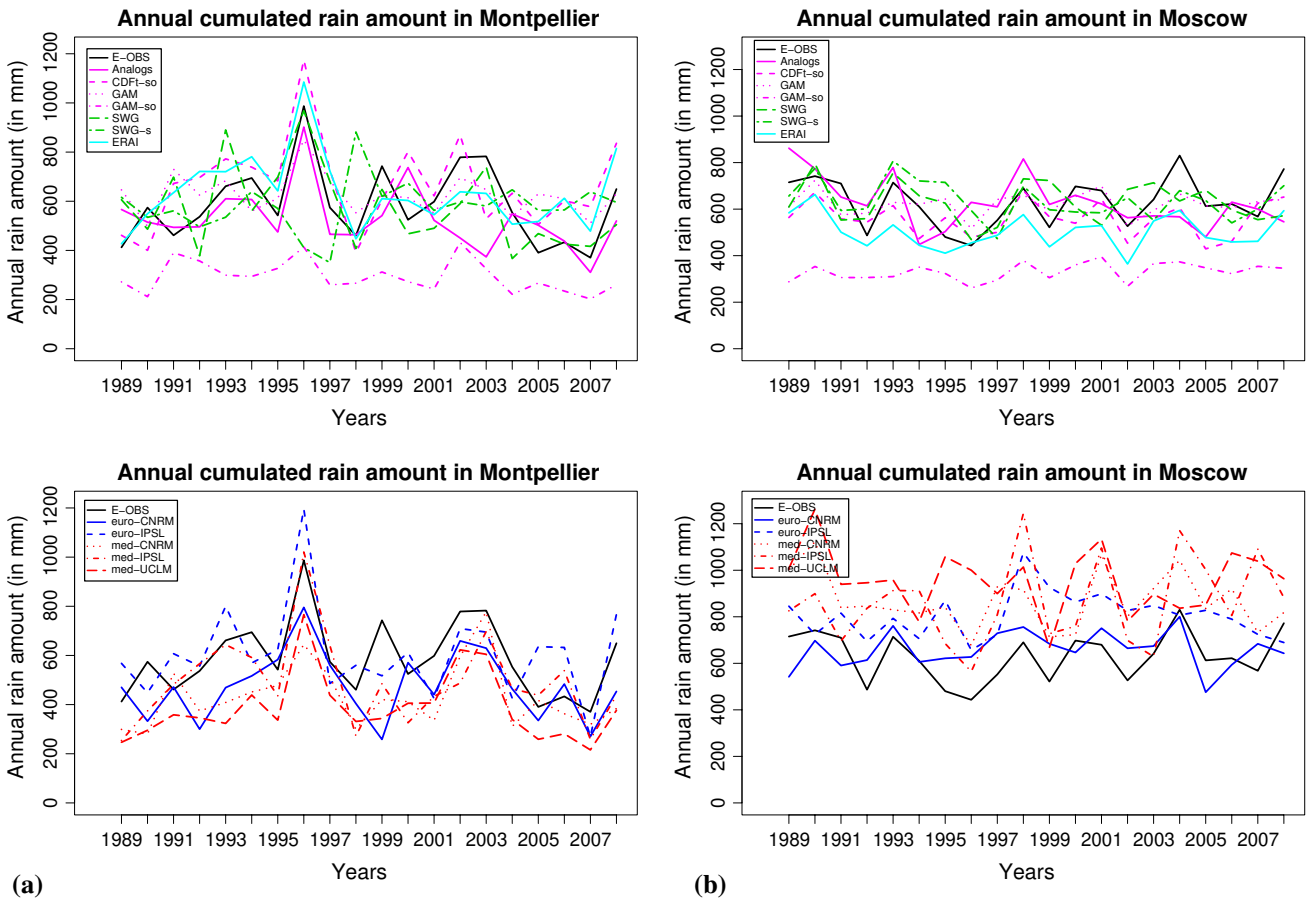

Fig. 11 Annual rain amount in mm · year^{-1} of the models, *left* for Montpellier and *right* Moscow. *Top panels* regroups all the SDMs, ERAi and E-OBS, *bottom panels* all the RCMs and E-OBS. **a** Montpellier, **b** Moscow

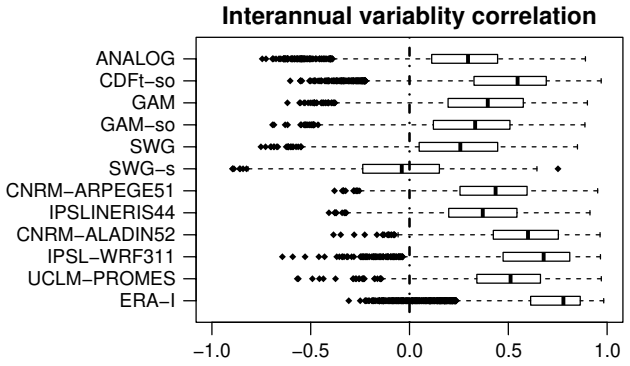

Fig. 12 *Boxplots* of the correlation coefficients between annual rain amount time series of each model and E-OBS

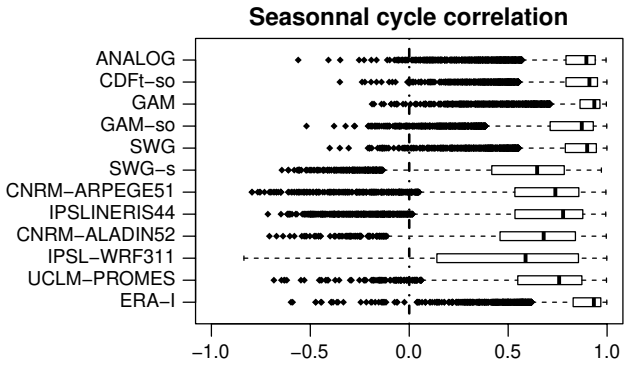

Fig. 13 Correlation coefficient between the seasonal cycle of each model and E-OBS

Now the seasonality is examined. To this end, the daily mean of each month (including zeros) over the 20 years is computed (i.e., 12 values, one for each month) for each model and E-OBS. Then the correlation between the seasonal cycle of each model and E-OBS is calculated. Figure 13 shows the corresponding boxplots. Here the results are opposite compared to the previous figure when comparing SDMs and RCMs. This time SDMs achieve higher correlation (except SWG-s), reaching correlations around 0.9 while RCMs have more troubles to reproduce the seasonal cycle, reaching correlations around 0.75 nevertheless. In the case of MED-IPSL, the bad seasonal cycle is partly a consequence of the land surface/atmosphere feedbacks described in Sect. 4.1.

A third index has been considered to evaluate the temporal properties. In Fig. 14, the first order summer autocorrelation coefficients (AR1) for each model are pictured. As for the first EOF, the aim is to see if the spatial distribution of the coefficients of each model is the same as for E-OBS. The AR1 coefficient is computed over the precipitation outputs gaussianized as in Sect. 4.3. GAM gives too high autocorrelation due to the fact that it generates a little amount of rain too frequently. Other SDMs have very low autocorrelation except the ANALOG which reproduces closely the autocorrelation of observations. Note that the CDFt-so model achieve very different AR1 coefficients than ERAi coefficients, it is a consequence of the LR used for the occurrence of this model. This widely modify the rain occurrence observed in ERAi and therefore influence the auto correlation. RCMs have autocorrelation values different from that of the observations but are very close to E-OBS in terms of range. The 2-day and 3-day-lag-autocorrelation values have also been computed (not shown), these coefficients decrease quite fast as expected for rain and the ranking of the models compared to E-OBS is the same as for AR1. In winter (see auxiliary materials) the results are similar for SDMs. RCMs are globally much better and their autocorrelation coefficients are really close to those computed for E-OBS.

5 Conclusions and discussion

5.1 Conclusions

In this study, an intercomparison of several precipitation downscaling models has been conducted. To this end, an intercomparison framework has been built following some essential requirements. First, all the models had to have common predictors (as much as possible) coming from the same database, here ERAi reanalyses. Second, observations and models outputs with the same spatial resolution and over a common area were considered. So, considering the available RCMs and observational data resolution (E-OBS), a resolution at 0.44° has been chosen. Third, the selected models had to represent all the downscaling approaches the authors have defined (TF, WG, WT, MOS statistical families and some dynamical models). So 11 models (six SDMs and five RCMs) have been selected and their outputs confronted according to criteria characterizing the four following aspects of the rain: occurrence, intensity, as well as spatial and temporal properties. This study is an opportunity to set-up and test the consistency of the intercomparison framework to compare outputs coming from SDMs as well as RCMs. Very different downscaling models, at least in terms of model philosophy, have been compared.

All the RCMs (except MED-IPSL), as well as GAM, seem to produce too many rainy days. For general consideration, modelling the rain occurrence by an LR (logistic regression) reveals itself to be a better approach than thresholding the outputs. Concerning the spells, all the models have better abilities to reproduce the wet spells than the dry ones and ANALOG is the best to reproduce them. However, even if ANALOG is good in terms of occurrence statistics, it fails in terms of time accuracy (Brier score).

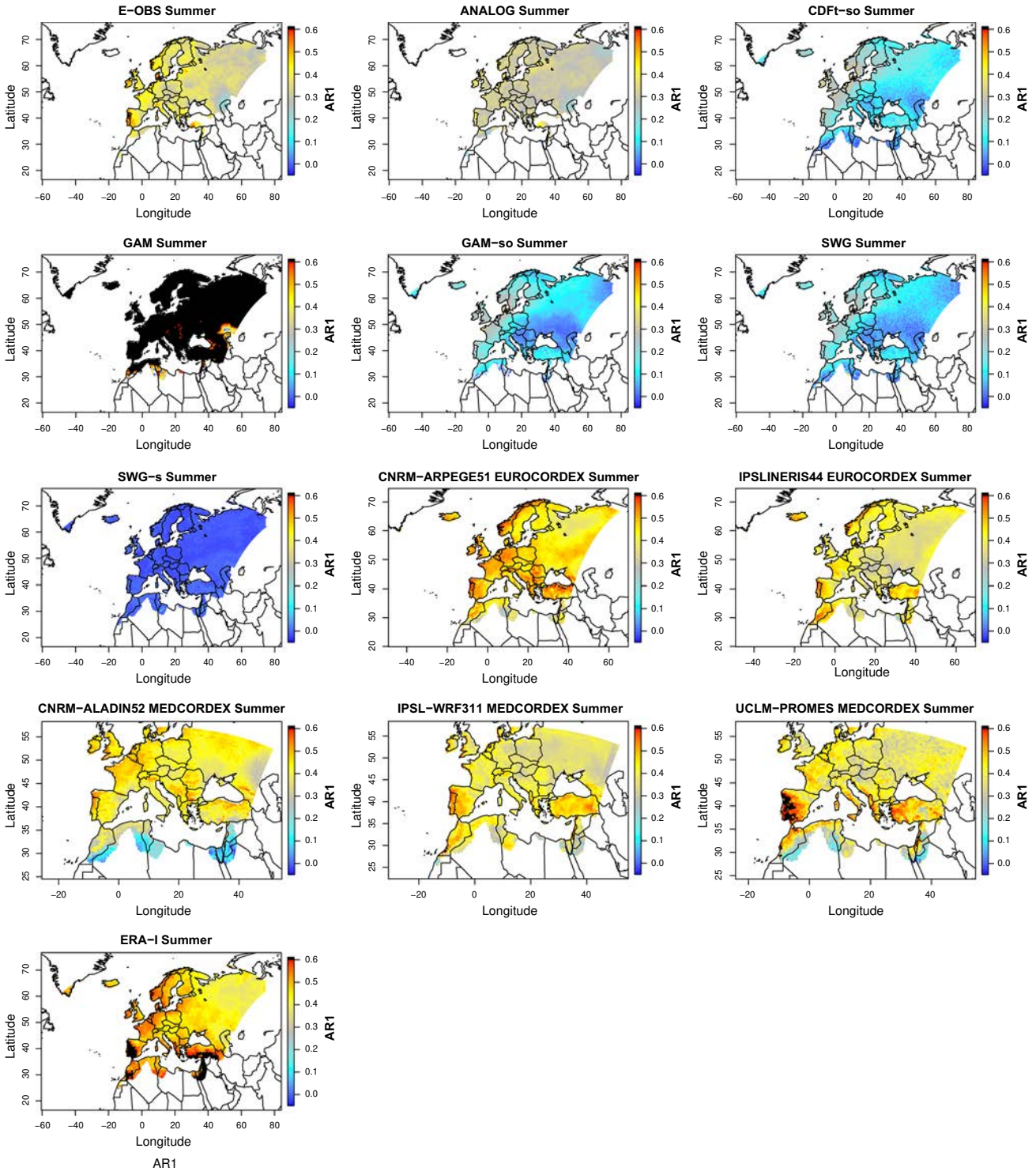

Fig. 14 First order autocorrelation coefficient for each model. Only values below 0.6 are represented, above they are saturated in *black*

The second examined aspect is the rain intensity. Here, the mean climatology is better reproduced by the stochastic models (SWG and SWG-s). While variability and extremes are better dealt by ANALOG and CDFt-so, SWG and SWG-s are close behind. All the other models present strong biases with variations over the domain. GAM-so and GAM completely fail in reproducing intensity properties. This is in agreement with Schoof and Pryor (2001) concerning TFs models performances. Concerning RCMs, the study corroborates the classical results found in the

literature, namely that they are producing too many rainfall events (occurrence) but with low intensity (Sun et al. 2006; Stephens et al. 2010) except for MED-IPSL.

Spatial pattern are studied through two specific angles: first, the spatial variability thanks to an EOF analysis; second the pattern correlation of the daily maps. Concerning the spatial variability the ANALOG and EURO-CORDEX, are better reproducing E-OBS spatial rain patterns while the others show quite different or no patterns at all. The models with good spatial pattern are the models which have spatial constraints: by construction for RCMs and by keeping the observations spatial structure for ANALOG model. This shows the importance of developing statistical spatial models in the future. In terms of daily pattern correlation, the only model which has been nudged, MED-IPSL, is the best to achieve the daily pattern of E-OBS (even if the nudging has been done with ERAi).

Finally the temporality was investigated. In this study, SDMs fail to retrieve E-OBS inter-annual variability especially SWG and SWG-s models. It is probably due to the random nature of the simulations which can generate too large or too little rain amounts and thus simulate very different annual rain amount compared to the observations. Another explanation could be the lack of information in terms of inter-annual variability provided by the predictors. RCMs on the opposite are in general better with a good performance of EURO-CNRM and MED-IPSL. For several aspects, MED-IPSL model achieve good performances. This can be partly explained by the nudging performed inside the domain with ERAi. In the mean time, the SDMs succeed in reproducing the seasonality. RCMs have more difficulties to achieve a good seasonal cycle. Finally, in terms of autocorrelation, ANALOG, followed by the RCMs are close to the E-OBS autocorrelation. Other SDMs, are quite far from E-OBS autocorrelation values.

In order to synthesize the results, the statistical models and the dynamical models are ranked according to each criterion. The models are scored according to the domain-wide averaged indicators in Tables 4 and 5 over MED-CORDEX and EURO-CORDEX domain respectively: the lower the score, the better the model. A global score can be obtained by simply adding each indicator rank over each of the considered aspects of the model evaluation (occurrence, intensity, spatial and temporal). Tables 4 and 5 can be used as a guideline for the users of the simulations. It allows to choose the model(s) to be used, depending on the needed statistical properties that the simulations must satisfy for some particular applications. Indeed, there is not one model in particular which really takes the advantage on the others considering the four aspects of the evaluation. Their performances really rely on the considered indicators and therefore on the use of the model simulations. Thus, the model quality depends on the end-users needs and the properties they expect the data to have to define their "best" model.

5.2 Perspectives and discussion

Many perspectives can be foreseen for this work. The choice between SDM and RCM methods can not be done solely on the reproduction of ERAi climate. A direct continuation can be the intercomparison in a future climate context. First, the couple "GCM/SDM" over the historical (or CTRL) period has to be evaluated. From the SDMs fitted over the historical period (e.g., 1979–2008) to the observations (E-OBS) and reanalyses (ERAi) (i.e., basically similar to which has been done in this study), new time series driven this time by GCMs as predictors will be generated and evaluated. A good agreement of those time series with observations would mean that GCMs provide good predictors to simulate local-scale variables. Thus, the ability of the SDMs to reproduce the climatological present characteristics of the precipitation when driven by historical GCM fields would be assessed. The evaluation would be performed only in terms of statistics. In other words, indicators needing day-to-day synchronicity (e.g., Brier score and daily maps correlation) would not be relevant in that case. The next step would be to assess the capability of the SDMs to capture changes in future spatial and/or temporal local-scale properties. The couple "GCM / SDM" would be evaluated in a climate change context with a RCM-based pseudo-observations approach, for example as developed in Vrac et al. (2007c) and applied in Gaitan et al. (2014). RCMs will be considered as proxies of future climate conditions and RCMs and SDMs have to be driven by the same GCM simulations. SDMs fitted to CTRL GCM simulations and pseudo-observations coming from RCM over the same time period will be driven with future GCM simulations (multiple emissions scenarios can be used) to generate new time series. Good agreement between those time series and the future RCM time series would mean that the SDM is able to capture a similar climate change signal as that simulated by the RCM.

A multi-model approach can also be an interesting follow-up study. It has been first tested in Sanders (1963) for subjective and Perrone and Miller (1985) for objective weather forecasting and has proven itself to be superior to the methodologies applied individually. There are many occasions when this result is verified. Even theoretical contributions are made to support these experimental facts (e.g. Hagedorn et al. 2005). However, it is not generalized until the 2000s (e.g. Palmer and Shukla 2000; Pavan and Doblas-Reyes 2000; Lambert and Boer 2001; Gillett et al. 2003; Jacob et al. 2007; Ruti et al. 2011; Solman et al. 2013; Gallardo et al. 2013) and is consolidated as the standard in studies of climate performed with dynamical models. Therefore, future studies should include the multi-model approach when MED-CORDEX and EURO-CORDEX databases are completed. This methodology could be thus extended, as noted by Haylock et al. (2006)

Table 4 Score and rank table for summer season computed over the MED-CORDEX domain

Model	ANALOG	CDFt-so	GAM	GAM-so	SWG	SWG-s	EURO-CNRM	EURO-IPSL	MED-CNRM	MED-IPSL	MED-UCLM	ERA-I
Wet days % bias	−1.04 (3)	−0.29 (2)	32.87 (9)	−0.29 (2)	−0.29 (2)	0.02 (1)	10.75 (7)	4.10 (5)	11.19 (8)	−2.80 (4)	8.33 (6)	3.66
Wet mean persistence	−0.09 (2)	−0.50 (5)	9.01 (9)	−0.50 (5)	−0.50 (5)	−0.66 (6)	0.72 (8)	−0.02 (1)	0.66 (6)	−0.17 (3)	0.34 (4)	0.24
Dry mean persistence	−28.13 (8)	−14.97 (2)	286.75 (9)	−14.97 (2)	−14.97 (2)	6.28 (1)	−25.98 (5)	−26.38 (6)	−27.37 (7)	−21.57 (4)	−16.70 (3)	−16.17
Brier Score	0.26 (6)	0.14 (1)	0.37 (9)	0.14 (1)	0.14 (1)	0.16 (3)	0.28 (8)	0.26 (6)	0.23 (4)	0.15 (2)	0.24 (5)	0.14
Total rank occurrence	19	**10**	36	**10**	**10**	11	28	18	25	13	18	–
Mean bias mm	−0.26 (6)	1.18 (9)	−3.74 (11)	−0.15 (3)	0.16 (4)	−0.01 (1)	−1.08 (8)	0.33 (7)	−1.26 (10)	−0.06 (2)	0.19 (5)	−1.47
Variance ratio %	88.45 (2)	146.71 (7)	2.65 (11)	42.09 (10)	98.93 (1)	82.79 (3)	80.56 (4)	166.53 (8)	75.52 (5)	140.27 (6)	224.06 (9)	76.97
Q99 bias mm	−2.57 (2)	4.00 (7)	−21.90 (11)	8.45 (9)	0.17 (1)	−3.01 (5)	−2.85 (3)	5.23 (8)	−3.18 (6)	2.85 (3)	10.99 (10)	−4.04
Total rank intensity	10	23	33	22	**6**	9	15	23	22	11	24	–
EOF1	1	2	4	11	2	11	6	6	6	5	10	–
Spatial pattern correlation	0.23 (8)	0.27 (8)	0.32 (6)	0.28 (7)	0.19 (10)	0.11 (11)	0.36 (4)	0.36 (4)	0.49 (2)	0.62 (1)	0.41 (3)	0.53
Total rank spatial	9	10	10	18	12	22	10	10	8	**6**	13	–
Cor. annual amount	0.26 (9)	0.46 (4)	0.34 (7)	0.27 (8)	0.20 (10)	−0.10 (11)	0.43 (5)	0.37 (6)	0.58 (2)	0.63 (1)	0.50 (3)	0.69
Cor. seas. cycle	0.80 (3)	0.81 (2)	0.87 (1)	0.75 (5)	0.80 (3)	0.58 (9)	0.58 (9)	0.61 (7)	0.61 (7)	0.49 (11)	0.68 (6)	0.85
AR1 (E-OBS:0.38)	0.31 (6)	0.12 (7)	0.68 (10)	0.11 (8)	0.10 (9)	0.00 (11)	0.46 (5)	0.41 (4)	0.37 (2)	0.39 (2)	0.38 (1)	0.52
Total rank temporal	18	13	18	21	22	31	19	17	11	14	**10**	–

The value corresponds to the mean over the domain and their rank is indicated into bracket

Bold values indicate the best models for each one of the four evaluation aspect

Table 5 Score and rank table for summer season computed over the EURO-CORDEX domain

Model	ANALOG	CDFt-so	GAM	GAM-so	SWG	SWG-s	EURO-CNRM	EURO-IPSL	ERA-I
Wet days % bias	−0.75 (3)	−0.10 (1)	35.43 (6)	−0.10 (1)	−0.10 (1)	0.14 (2)	11.90 (5)	3.98 (4)	3.48
Wet mean persistence	−0.08 (2)	−0.51 (3)	10.26 (6)	−0.51 (3)	−0.51 (3)	−0.70 (4)	0.78 (5)	−0.04 (1)	0.23
Dry mean persistence	−18.44 (3)	−9.70 (2)	193.63 (6)	−9.70 (2)	−9.70 (2)	5.06 (1)	−18.45 (4)	−18.60 (5)	−10.74
Brier Score	0.30 (4)	0.16 (1)	0.40 (6)	0.16 (1)	0.16 (1)	0.18 (2)	0.31 (5)	0.29 (3)	0.14
Total rank occurrence	12	**7**	24	**7**	**7**	9	19	13	–
Mean bias mm	−0.31 (3)	1.02 (7)	−3.37 (8)	−0.86 (5)	0.13 (2)	0.00 (1)	−0.86 (5)	0.54 (4)	−1.31
Variance ratio%	85.26 (3)	138.18 (5)	3.38 (8)	30.59 (6)	96.97 (1)	82.57 (4)	90.17 (2)	183.01 (7)	74.69
Q99 bias mm	−2.51 (4)	3.49 (6)	−20.39 (8)	0.80 (2)	−0.25 (1)	−2.81 (5)	−1.67 (3)	5.98 (7)	−4.00
Total rank intensity	10	18	24	13	**4**	10	10	18	–
EOF1	1	2	4	7	2	7	5	5	–
Spatial pattern correlation	0.24 (6)	0.29 (4)	0.32 (3)	0.28 (5)	0.20 (7)	0.10 (8)	0.37 (1)	0.37 (1)	0.53
Total rank spatial	7	**6**	7	12	9	15	**6**	**6**	–
Cor. annual amount	0.27 (6)	0.50 (1)	0.38 (3)	0.32 (5)	0.25 (7)	−0.05 (8)	0.42 (2)	0.38 (3)	0.72
Cor. seas. cycle	0.83 (4)	0.85 (2)	0.90 (1)	0.80 (5)	0.84 (3)	0.59 (8)	0.65 (7)	0.69 (6)	0.88
AR1 (E-OBS:0.38)	0.31 (3)	0.14 (5)	0.69 (4)	0.12 (6)	0.11 (7)	−0.01 (8)	0.44 (2)	0.38 (1)	0.50
Total rank temporal	13	**8**	**8**	16	17	24	11	10	–

The value corresponds to the mean over the domain and their rank is indicated into bracket

Bold values indicate the best models for each one of the four evaluation aspect

to a mix of dynamical and statistical models. Note that one major difference between ensemble methods in weather forecasting and in climate studies is that the first must deal effectively with uncertainty in initial conditions, while in climate studies this uncertainty is not as much relevant.

Moreover, a way to refine the results would be to study the impacts of E-OBS uncertainties on the downscaled data. Some studies pointed out some quality inconsistencies. For instance in Hofstra et al. (2009) problems such as data homogeneity over the E-OBS domain or oversmoothing in interpolation scheme causing difficulties to catch correctly the extremes or rain patterns over mountains have been pointed out. Therefore, the data uncertainty caused by the interpolation is ill-estimated. In this study, E-OBS V8 has been used. Potentially some improvements can be expected, if the last version E-OBS V11 is used instead, since the network density has been increased and an artefact of drizzle occurrence has been corrected. However, concerning the drizzle effect, it should not influence our results since the rain occurrence threshold is set at 1 mm. This occurrence threshold could also influence the results. In our case, simulations with a 0 mm threshold have also been tested for all the models (not shown). This changes the indicators values but does not influence the ranking of the models. The poor performance of GAM is not a consequence of the threshold since the same poor performances of GAM have also been observed for the 0 mm threshold. This mainly comes from the fact that the deterministic TF based models are not suited to simulate precipitation. Besides, concerning the drizzle effect of the RCMs, the results show that the tested RCMs produce too many rainy days even with this threshold except for MED-IPSL model (see Fig. 3).

Improvements can also be made on SDMs calibration, for instance by improving the predictors selection process or adding other predictors. It is worth noticing that the first exploratory step based on the SPARSE CCA algorithm (i.e., to determine the variables that make sense as predictors for precipitation downscaling) has been performed only on the first principal component of each variable. Although the SPARSE CCA method is computationally intensive, it would be interesting to have additional leading PCs in this exploratory analysis to bring more robustness to the choice of the predictors. Moreover, as the SPARSE CCA has not been applied in the cross-validation context, the performance of the SDMs as assessed via cross-validation could be overly optimistic (or at least biased). Although the differences could be minor, it would be interesting to perform the selection of the predictors within the cross-validation procedure. Note that the cross-validation scheme used in this study has a rather short calibration period (20 years), which may underestimate or even overestimate the skills for some methods. One solution could be to use a "29-leave-one-out" scheme, with calibrations of the models made on 29 years and evaluations on the left-out year. This 29-leave-one-out strategy, however, may not be an adapted strategy to evaluate the performances of the models in a changing climate context. Indeed, as the one-year left out would be either surrounded by the 29 calibration years, or appended (before or after) to the 29 calibration years, the basic statistical properties of the large-scale predictors and of the local-scale data should be the same in the 29 calibration years

and the evaluation year. Hence, this strategy could provide overly optimistic results compared to an evaluation performed on a whole decade (or more). Besides, the 20-leave-10-out method is closer to the framework in which the downscaling methods are applied (calibration on historical period and application on future period).That is why, despite the limited length (20 years) of the calibration period, the "turning"? 20 leave-10-out cross-validation procedure has been favoured in this study. Predictors relevant in terms of rain physical process such as the CAPE (convective available potential energy, Foufoula-Georgiou and Tsonis 1996), the vertical wind shear (Wingo and Cecil 2009) or moisture flux (Yang et al. 2010) characterizing the atmosphere instability can be also considered. Some temporal information could be added by including the previous day precipitation observation especially for the occurrence model (Kleiber et al. 2012). Weather regimes or seasonal cycle indicators could also bring interesting information leading to potential improvements. Globally, the intercomparison could be broadened by adding more statistical and dynamical models or adding new variables of interest such as temperature or wind. Thus, an inter-variable analysis could be carried out based on adapted indicators.

Besides, the SDMs' features can be improved. According to the results, it would be legitimate to focus on the ANALOG model. However, this model presents some limitations. Indeed, it is limited by its range over the calibration period: in case of future projections in context of climate change signal it is possible to miss that signal because ANALOG cannot go beyond the calibration climate range. Besides, this model has more large-scale information than the other models tested here. This could also explain its performance. One can object that the ANALOG model could have been run with the same set of predictors that have been used for the other SDMs (i.e., the 12 PCs). The authors are not aware of any application of the ANALOG model with PCs as predictors. The usual way to apply it is to work with fields of anomalies. However, the ANALOG model has also been run with PCs as predictors for comparison. This approach strongly degrades the results of the ANALOG model compared to using the anomalies as predictors. This model presents large biases and sometimes the results are even unrealistic (not shown). Some analog approaches combine multiple analogs (e.g., Radanovics et al. 2013; Chardon et al. 2014; Yiou 2014). In the way the analogs are computed in our study, the use of a combination (e.g., through a mean or weighted average) of multiple analogs would decrease the quality of the ANALOG simulation. Indeed, it would undermine the mean and the variance of the ANALOG model output and could also introduce a bias in the wet days frequency. An artificial variance-inflating procedure would then be necessary to maintain the main statistical properties.

On the opposite, a focus can be given to the SWG model. Indeed, in spite of its caveats in terms of spatiality and temporality, it seems to be very promising. There are many ways of improvement for instance by giving the model a spatial structure through a covariance function (e.g., Vischel et al. 2009) or by improving the Bernoulli/Gamma marginal probability distribution function used here. It would allow us to generate daily rain fields with a spatial coherence and one model for an entire region instead of a model per grid-point. Instead of two seasons, considering weather regimes could also lead to a potential improvement (Vrac et al. 2007b). Of course, the CORDEX regions are probably too large to define a simple but realistic dependence model. However, improving the SWG model seems a good compromise between the many leads of improvements and the model flexibility. Spatial coherence can also be ensured in other modelling framework: for instance the spatial MOS model, EC-BC, developed in Vrac and Friederichs (2014). Another path can be a combination of a stochastic model with an ANALOG model.

Finally, the present study has focused entirely on the intercomparison framework and the results that have come out of it. This work aspires to set an easily reproducible ground rules to conduct a RCM intercomparison which includes RCMs as well as SDMs and allows the SDMs to fit into the CORDEX initiative. Based on that, it is expected to perform consistent future intercomparison studies between SDMs as well as RCMs.

Acknowledgments The authors are thankful to all the RCM data providers, especially to R. Vautard (IPSL) and A. Colette (INERIS) for the WRF-IPSL-INERIS44 EURO-CORDEX run and Météo-France/CNRM (A. Alias, S. Somot) for the CNRM-ALADIN52 MED-CORDEX run. The MED-CORDEX simulations used in this work are downloaded from the MED-CORDEX data portal (www.medcordex.eu/medcordex.php). This work has been partially funded by the Spanish Ministry of Education and Science and the European Regional Development Fund, through Grant CGL2007-66440-C04-02. We also thank F. Blondot (HSM) who, in collaboration with Julie Carreau, helped us for the predictors selection. All the estimations and simulations for the stochastic and the TF models have been done with the R-package "VGAM" (Yee 2010). Special thanks are due to Thomas Yee, the "VGAM" package author for his help. The MOS model has been computed thanks to the R-package CDFt (Michelangeli et al. 2009). This work has been supported by the ANR StaRMIP project, the ANR REMEMBER project and the REMedHE GICC project. It is a contribution to the HyMeX program (HYdrological cycle in The Mediterranean EXperiment) through INSU-MISTRALS support and the MED-CORDEX program. It was supported by the IPSL group for regional climate and environmental studies, with granted access to the HPC resources of IDRIS (under allocation i2011010227). It is a contribution to the CORDEX-ESD initiative (http://wcrp-cordex.ipsl.jussieu.fr/index.php/community/cordex-esd) and to the COST Action VALUE (http://www.value-cost.eu/, Maraun et al. 2015).

Table 6 The main R features to reproduce the simulations are indicated in this table

Model	LR	GAM	SWG intensity	CDFt-so intensity
R-package	VGAM	VGAM	VGAM	CDFt
Function	vgam	vgam	vgam	CDFt
Parameters	binomialff (mv = TRUE)	gaussianff	gammaR (zero = NULL)	npas = 100, dev = 2

The distribution parameters are inferred by maximum-likelihood with the vgam function

Appendix: Technical features

See Table 6.

References

Ambrosino C, Chandler R, Todd M (2014) Rainfall-derived growing season characteristics for agricultural impact assessments in South Africa. Theor Appl Climatol 115(3–4):411–426. doi:10.1007/s00704-013-0896-y

Bardossy A, Plate EJ (1992) Space–time model for daily rainfall using atmospheric circulation patterns. Water Resour Res 28(5):1247–1259. doi:10.1029/91WR02589

Barnston AG, Livezey RE (1987) Classification, seasonality and persistence of low-frequency atmospheric circulation patterns. Mon Weather Rev 115(6):1083–1126. doi:10.1175/1520-0493(1987)115<1083:CSAPOL>2.0.CO;2

Bellone E, Hughes JP, Guttorp P (2000) A hidden Markov model for downscaling synoptic atmospheric patterns to precipitation amounts. J Hydrol 15(1):1–12. http://www.int-res.com/abstracts/cr/v15/n1/p1-12/

Bougeault P (1985) A simple parameterization of the large-scale effects of cumulus convection. Mon Weather Rev 113(12):2108–2121. doi:10.1175/1520-0493(1985)113<2108:ASPOTL>2.0.CO;2

Bouvier C, Cisneros L, Dominguez R, Laborde JP, Lebel T (2003) Generating rainfall fields using principal components (pc) decomposition of the covariance matrix: a case study in mexico city. J Hydrol 278(1–4):107–120. doi:10.1016/S0022-1694(03)00122-7. http://www.sciencedirect.com/science/article/pii/S0022169403001227

Brier GW (1950) Verification of forecasts expressed in terms of probability. Mon Weather Rev 78(1):1–3. doi:10.1175/1520-0493(1950)078

Buishand TA, Shabalova MV, Brandsma T (2004) On the choice of the temporal aggregation level for statistical downscaling of precipitation. J Clim 17(9):1816–1827. doi:10.1175/1520-0442(2004)017<1816:OTCOTT>2.0.CO;2

Bürger G, Murdock TQ, Werner AT, Sobie SR, Cannon AJ (2012) Downscaling extremes—an intercomparison of multiple statistical methods for present climate. J Clim 25(12). doi:10.1175/JCLI-D-11-00408.1

Cavazos T, Hewitson C Bruce (2005) Performance of NCEP-NCAR reanalysis variables in statistical downscaling of daily precipitation. Clim Res 28(2):95–107. doi:10.3354/cr028095. http://www.int-res.com/abstracts/cr/v28/n2/p95-107/

Chaboureau JP, Bechtold P (2002) A simple cloud parameterization derived from cloud resolving model data: diagnostic and prognostic applications. J Atmos Sci 59(15):2362–2372. doi:10.1175/1520-0469(2002)059<2362:ASCPDF>2.0.CO;2

Chaboureau JP, Bechtold P (2005) Statistical representation of clouds in a regional model and the impact on the diurnal cycle of convection during tropical convection, cirrus and nitrogen oxides (troccinox). J Geophys Res Atmos 110(D17). doi:10.1029/2004JD005645

Chandler RE, Wheater HS (2002) Analysis of rainfall variability using generalized linear models: a case study from the west of Ireland. Water Resour Res 38(10):1192. doi:10.1029/2001WR000906

Chardon J, Hingray B, Favre A, Autin P, Gailhard J, Zin I, Obled C (2014) Spatial similarity and transferability of analog dates for precipitation downscaling over france. J Clim 27(13):5056–5074. doi:10.1175/JCLI-D-13-00464.1

Charles SP, Bates BC, Whetton PH, Hughes JP (1999) Validation of downscaling models for changed climate conditions: case study of southwestern Australia. Clim Res 12(1):1–14. doi:10.3354/cr012001. http://www.int-res.com/abstracts/cr/v12/n1/p1-14/

Chiriaco M, Bastin S, Yiou P, Haeffelin M, Dupont JC, Stéfanon M (2014) European heatwave in July 2006: observations and modeling showing how local processes amplify conducive large-scale conditions. Geophys Res Lett 41(15):5644–5652. doi:10.1002/2014GL060205

Christensen J, Carter T, Rummukainen M, Amanatidis G (2007) Evaluating the performance and utility of regional climate models: the PRUDENCE project. Clim Change 81(1):1–6. doi:10.1007/s10584-006-9211-6

Coiffier J (2011) Fundamentals of numerical weather prediction. Cambridge University Press. doi:10.1017/CBO9780511734458 (Cambridge books online)

Colin J, Déqué M, Radu R, Somot S (2010) Sensitivity study of heavy precipitation in limited area model climate simulations: influence of the size of the domain and the use of the spectral nudging technique. Tellus A 62(5):591–604. doi:10.1111/j.1600-0870.2010.00467.x

Crawford T, Betts NL, Favis-Mortlock D (2007) GCM grid-box choice and predictor selection associated with statistical downscaling of daily precipitation over Northern Ireland. Clim Res 34(2):145–160. doi:10.3354/cr034145. http://www.int-res.com/abstracts/cr/v34/n2/p145-160/

Cuxart J, Bougeault P, Redelsperger JL (2000) A turbulence scheme allowing for mesoscale and large-eddy simulations. Q J R Meteorol Soc 126(562):1–30. doi:10.1002/qj.49712656202

Dee DP, Uppala SM, Simmons AJ, Berrisford P, Poli P, Kobayashi S, Andrae U, Balmaseda MA, Balsamo G, Bauer P, Bechtold P, Beljaars ACM, van de Berg L, Bidlot J, Bormann N, Delsol C, Dragani R, Fuentes M, Geer AJ, Haimberger L, Healy SB, Hersbach H, Hólm EV, Isaksen L, Kållberg P, Köhler M, Matricardi M, McNally AP, Monge-Sanz BM, Morcrette JJ, Park BK, Peubey C, de Rosnay P, Tavolato C, Thépaut JN, Vitart F (2011) The ERA-interim reanalysis: configuration and performance of the data assimilation system. Q J R Meteorol Soc 137(656):553–597. doi:10.1002/qj.828

Denis B, Laprise R, Caya D (2003) Sensitivity of a regional climate model to the resolution of the lateral boundary conditions. Clim Dyn 20(2–3):107–126. doi:10.1007/s00382-002-0264-6

Dibike YB, Coulibaly P (2006) Temporal neural networks for downscaling climate variability and extremes. Neural Netw 19(2):135–144. doi:10.1016/j.neunet.2006.01.003. http://www.

sciencedirect.com/science/article/pii/S0893608006000062 (Earth Sciences and Environmental Applications of Computational Intelligence)

Domínguez M, Romera R, Sánchez E, Fita L, Fernández J, Jiménez-Guerrero P, Montávez J, Cabos W, Liguori G, Gaertner M (2013) Present-climate precipitation and temperature extremes over spain from a set of high resolution rcms). Clim Res 58(2):149–164. doi:10.3354/cr01186. http://www.int-res.com/abstracts/cr/v58/n2/p149-164/

Douville H, Planton S, Royer J, Stephenson D, Tyteca S, Kergoat L, Lafont S, Betts R (2000) The importance of vegetation feedbacks in doubled-CO_2 time-slice experiments. Ann Geophys 11(12):1095–1115

Drobinski P, Ducrocq V, Alpert P, Anagnostou E, Béranger K, Borga M, Braud I, Chanzy A, Davolio S, Delrieu G, Estournel C, Boubrahmi NF, Font J, Grubišić V, Gualdi S, Homar V, Ivančan-Picek B, Kottmeier C, Kotroni V, Lagouvardos K, Lionello P, Llasat MC, Ludwig W, Lutoff C, Mariotti A, Richard E, Romero R, Rotunno R, Roussot O, Ruin I, Somot S, Taupier-Letage I, Tintore J, Uijlenhoet R, Wernli H (2014) Hymex: a 10-year multidisciplinary program on the mediterranean water cycle. Bull Am Meteorol Soc 95(7):1063–1082. doi:10.1175/BAMS-D-12-00242.1

Déqué M (2007) Frequency of precipitation and temperature extremes over France in an anthropogenic scenario: model results and statistical correction according to observed values. Glob Planet Change 57(1–2):16–26. doi:10.1016/j.gloplacha.2006.11.030. http://www.sciencedirect.com/science/article/pii/S0921818106002748

Déqué M, Piedelievre J (1995) High resolution climate simulation over Europe. Clim Dyn 11(6):321–339. doi:10.1007/BF00215735

ECMWF (2004) IFS documentation CY28r1. ECMWF, reading, pp 7–32. http://www.oldecmwfint/research/ifsdocs/CY28r1/pdf_files/Physics.pdf

Ek MB, Mitchell KE, Lin Y, Rogers E, Grunmann P, Koren V, Gayno G, Tarpley JD (2003) Implementation of Noah land surface model advances in the national centers for environmental prediction operational mesoscale eta model. J Geophys Res Atmos 108(D22). doi:10.1029/2002JD003296

Fealy R, Sweeney J (2007) Statistical downscaling of precipitation for a selection of sites in Ireland employing a generalised linear modelling approach. Int J Climatol 27(15):2083–2094. doi:10.1002/joc.1506

Flaounas E, Bastin S, Janicot S (2011) Regional climate modelling of the 2006 West African monsoon: sensitivity to convection and planetary boundary layer parameterisation using wrf. Clim Dyn 36(5–6):1083–1105. doi:10.1007/s00382-010-0785-3

Flaounas E, Drobinski P, Vrac M, Bastin S, Lebeaupin-Brossier C, Stéfanon M, Borga M, Calvet JC (2013) Precipitation and temperature space–time variability and extremes in the mediterranean region: evaluation of dynamical and statistical downscaling methods. Clim Dyn 40(11–12):2687–2705. doi:10.1007/s00382-012-1558-y

Foufoula-Georgiou E, Tsonis A (1996) Preface [to the special section on space–time variability and dynamics of rainfall]. J Geophys Res Atmos 101(D21):26,161–26,163. doi:10.1029/96JD03121

Fu C, Wang S, Xiong Z, Gutowski WJ, Lee DK, McGregor JL, Sato Y, Kato H, Kim JW, Suh MS (2005) Regional climate model intercomparison project for Asia. Bull Am Meteorol Soc 86(2):257–266. doi:10.1175/BAMS-86-2-257

Gaitan C, Hsieh W, Cannon A (2014) Comparison of statistically downscaled precipitation in terms of future climate indices and daily variability for southern Ontario and Quebec, Canada. Clim Dyn 1–17. doi:10.1007/s00382-014-2098-4

Gallardo C, Gil V, Hagel E, Tejeda C, de Castro M (2013) Assessment of climate change in Europe from an ensemble of regional climate models by the use of Köppen–Trewartha classification. Int J Climatol 33(9):2157–2166. doi:10.1002/joc.3580

Gillett NP, Zwiers FW, Weaver AJ, Stott PA (2003) Detection of human influence on sea-level pressure. Nature 422(6929):292–294. doi:10.1038/nature01487

Giorgi F, Jones C, Asrar GR (2009) Addressing climate information needs at the regional level: the CORDEX framework. Bull World Meteorol Organ 58(3):175–183

Grell GA, Dévényi D (2002) A generalized approach to parameterizing convection combining ensemble and data assimilation techniques. Geophys Res Lett 29(14):38-€œ1–38-€œ4. doi:10.1029/2002GL015311

Grenier P, Parent AC, Huard D, Anctil F, Chaumont D (2013) An assessment of six dissimilarity metrics for climate analogs. J Appl Meteorol Climatol 52(4):733–752. doi:10.1175/JAMC-D-12-0170.1

Gudmundsson L, Bremnes JB, Haugen JE, Engen-Skaugen T (2012) Technical note: downscaling RCM precipitation to the station scale using statistical transformations—a comparison of methods. Hydrol Earth Syst Sci 16(9):3383–3390. doi:10.5194/hess-16-3383-2012. http://www.hydrol-earth-syst-sci.net/16/3383/2012/

Hagedorn R, Doblas-Reyes FJ, Palmer TN (2005) The rationale behind the success of multi-model ensembles in seasonal forecasting—i. Basic concept. Tellus A 57(3):219–233. doi:10.1111/j.1600-0870.2005.00103.x

Harpham C, Wilby RL (2005) Multi-site downscaling of heavy daily precipitation occurrence and amounts. J Hydrol 312(1–4):235–255. doi:10.1016/j.jhydrol.2005.02.020. http://www.sciencedirect.com/science/article/pii/S0022169405000922

Hastie T, Tibshirani R (1990) Generalized additive models. Monographs on statistics and applied probability. Chapman and Hall. http://books.google.co.uk/books?id=qa29r1Ze1coC

Haylock MR, Cawley GC, Harpham C, Wilby RL, Goodess CM (2006) Downscaling heavy precipitation over the United Kingdom: a comparison of dynamical and statistical methods and their future scenarios. Int J Climatol 26(10):1397–1415. doi:10.1002/joc.1318

Haylock MR, Hofstra N, Klein Tank AMG, Klok EJ, Jones PD, New M (2008) A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. J Geophys Res Atmos 113(D20). doi:10.1029/2008JD010201

Herrmann M, Somot S, Calmanti S, Dubois C, Sevault F (2011) Representation of spatial and temporal variability of daily wind speed and of intense wind events over the mediterranean sea using dynamical downscaling: impact of the regional climate model configuration. Nat Hazards Earth Syst Sci 11(7):1983–2001. doi:10.5194/nhess-11-1983-2011. http://www.nat-hazards-earth-syst-sci.net/11/1983/2011/

Hewitson B, Crane R (1996) Climate downscaling: techniques and application. Clim Res 7(2):85–95. doi:10.3354/cr007085. http://www.int-res.com/abstracts/cr/v07/n2/p85-95/

Hewitt CD (2004) Ensembles-based predictions of climate changes and their impacts. Eos, Trans Am Geophys Union 85(52):566. doi:10.1029/2004EO520005

Hofstra N, Haylock M, New M, Jones PD (2009) Testing E-OBS European high-resolution gridded data set of daily precipitation and surface temperature. J Geophys Res Atmos 114(D21). doi:10.1029/2009JD011799

Hong SY, Lim JOJ (2006) The wrf single-moment 6-class microphysics scheme (wsm6). J Korean Meteorol Soc 42(2):129–151

Hong SY, Dudhia J, Chen SH (2004) A revised approach to ice microphysical processes for the bulk parameterization of

clouds and precipitation. Mon Weather Rev 132(1):103–120. doi:10.1175/1520-0493(2004)132<0103:ARATIM>2.0.CO;2

Hong SY, Noh Y, Dudhia J (2006) A new vertical diffusion package with an explicit treatment of entrainment processes. Mon Weather Rev 134(9):2318–2341. doi:10.1175/MWR3199.1

Hourdin F, Musat I, Bony S, Braconnot P, Codron F, Dufresne JL, Fairhead L, Filiberti MA, Friedlingstein P, Grandpeix JY, Krinner G, LeVan P, Li ZX, Lott F (2006) The LMDZ4 general circulation model: climate performance and sensitivity to parametrized physics with emphasis on tropical convection. Clim Dyn 27(7–8):787–813. doi:10.1007/s00382-006-0158-0

Iacono MJ, Delamere JS, Mlawer EJ, Shephard MW, Clough SA, Collins WD (2008) Radiative forcing by long-lived greenhouse gases: calculations with the aer radiative transfer models. J Geophys Res Atmos 113(D13). doi:10.1029/2008JD009944

Jacob D, Bärring L, Christensen O, Christensen J, de Castro M, Déqué M, Giorgi F, Hagemann S, Hirschi M, Jones R, Kjellström E, Lenderink G, Rockel B, Sánchez E, Schär C, Seneviratne S, Somot S, van Ulden A, van den Hurk B (2007) An inter-comparison of regional climate models for europe: model performance in present-day climate. Clim Change 81(1):31–52. doi:10.1007/s10584-006-9213-4

Jacob D, Petersen J, Eggert B, Alias A, Christensen O, Bouwer L, Braun A, Colette A, Déqué M, Georgievski G, Georgopoulou E, Gobiet A, Menut L, Nikulin G, Haensler A, Hempelmann N, Jones C, Keuler K, Kovats S, Kröner N, Kotlarski S, Kriegsmann A, Martin E, van Meijgaard E, Moseley C, Pfeifer S, Preuschmann S, Radermacher C, Radtke K, Rechid D, Rounsevell M, Samuelsson P, Somot S, Soussana JF, Teichmann C, Valentini R, Vautard R, Weber B, Yiou P (2014) EURP-CORDEX: new high-resolution climate change projections for European impact research. Reg Environ Change 14(2):563–578. doi:10.1007/s10113-013-0499-2

Jeong D, St-Hilaire A, Ouarda T, Gachon P (2012) Multisite statistical downscaling model for daily precipitation combined by multivariate multiple linear regression and stochastic weather generator. Clim Change 114(3–4):567–591. doi:10.1007/s10584-012-0451-3

Jiménez-Guerrero P, Montávez J, Domínguez M, Romera R, Fita L, Fernández J, Cabos W, Liguori G, Gaertner M (2013) Mean fields and interannual variability in RCM simulations over Spain: the ESCENA project. Clim Res 57(3):201–220. doi:10.3354/cr01165. http://www.int-res.com/abstracts/cr/v57/n3/p201-220/

Kain J, Fritsch J (1993) Convective parameterization for mesoscale models: the Kain–Fritsch scheme. The representation of cumulus convection in numerical models. No. 46 in Meteorological Monographs, American Meteorological Society

Kain JS (2004) The Kain–Fritsch convective parameterization: an update. J Appl Meteorol 43(1):170–181. doi:10.1175/1520-0450(2004)043<0170:TKCPAU>2.0.CO;2

Khan MS, Coulibaly P, Dibike Y (2006) Uncertainty analysis of statistical downscaling methods. J Hydrol 319(1–4):357–382. doi:10.1016/j.jhydrol.2005.06.035. http://www.sciencedirect.com/science/article/pii/S0022169405003719

Kilsby C, Jones P, Burton A, Ford A, Fowler H, Harpham C, James P, Smith A, Wilby R (2007) A daily weather generator for use in climate change studies. Environ Model Softw 22(12):1705–1719. doi:10.1016/j.envsoft.2007.02.005. http://www.sciencedirect.com/science/article/pii/S136481520700031X

Kleiber W, Katz RW, Rajagopalan B (2012) Daily spatiotemporal precipitation simulation using latent and transformed Gaussian processes. Water Resour Res 48(1). doi:10.1029/2011WR011105

Klein WH, Lewis BM, Enger I (1959) Objective prediction of five-day mean temperatures during winter. J Meteorol 16(9):972–682. doi:10.1175/1520-0469(1959)016<0672:OPOFDM>2.0.CO;2

Krinner G, Viovy N, de Noblet-Ducoudré N, Ogée J, Polcher J, Friedlingstein P, Ciais P, Sitch S, Prentice IC (2005) A dynamic global vegetation model for studies of the coupled atmosphere–biosphere system. Glob Biogeochem Cycles 19(1). doi:10.1029/2003GB002199

Lambert SJ, Boer GJ (2001) Cmip1 evaluation and intercomparison of coupled climate models. Clim Dyn 17(2–3):83–106. doi:10.1007/PL00013736

Laprise R, de Elía R, Caya D, Biner S, Lucas-Picher P, Diaconescu E, Leduc M, Alexandru A, Separovic L (2008) Challenging some tenets of regional climate modelling. Meteorol Atmos Phys 100(1–4):3–22. doi:10.1007/s00703-008-0292-9

Lavaysse C, Vrac M, Drobinski P, Lengaigne M, Vischel T (2012) Statistical downscaling of the French Mediterranean climate: assessment for present and projection in an anthropogenic scenario. Nat Hazards Earth Syst Sci 12(3):651–670. doi:10.5194/nhess-12-651-2012. http://www.nat-hazards-earth-syst-sci.net/12/651/2012/

Levavasseur G, Vrac M, Roche DM, Paillard D, Martin A, Vandenberghe J (2011) Present and LGM permafrost from climate simulations: contribution of statistical downscaling. Clim Past 7(4):1225–1246. doi:10.5194/cp-7-1225-2011. http://www.clim-past.net/7/1225/2011/

Lo JCF, Yang ZL, Pielke RA (2008) Assessment of three dynamical climate downscaling methods using the Weather Research and Forecasting (WRF) model. J Geophys Res Atmos 113(D9). doi:10.1029/2007JD009216

Machenhauer B, Windelband M, Botzet M, Hesselbjerg J, Déqué M, Jones G, Ruti P, Visconti G (1998) Validation and analysis of regional present-day climate and climate change simulations over europe. Max-Planck Institute of Meteorology Report No 275, pp 87

Maraun D, Widmann M, Gutiérrez JM, Kotlarski S, Chandler RE, Hertig E, Wibig J, Huth R, Wilcke RA (2015) Value: a framework to validate downscaling approaches for climate change studies. Earth's Future 3(1):1–14. doi:10.1002/2014EF000259

Mearns L, Sain S, Leung L, Bukovsky M, McGinnis S, Biner S, Caya D, Arritt R, Gutowski W, Takle E, Snyder M, Jones R, Nunes A, Tucker S, Herzmann D, McDaniel L, Sloan L (2013) Climate change projections of the North American Regional Climate Change Assessment Program (NARCCAP). Clim Change 120(4):965–975. doi:10.1007/s10584-013-0831-3

Mezghani A, Hingray B (2009) A combined downscaling-disaggregation weather generator for stochastic generation of multisite hourly weather variables over complex terrain: development and multi-scale validation for the Upper Rhone River basin. J Hydrol 377(3–4):245–260. doi:10.1016/j.jhydrol.2009.08.033. http://www.sciencedirect.com/science/article/pii/S0022169409005149

Michelangeli PA, Vrac M, Loukos H (2009) Probabilistic downscaling approaches: application to wind cumulative distribution functions. Geophys Res Lett 36(11). doi:10.1029/2009GL038401

Morcrette JJ (1990) Impact of changes to the radiation transfer parameterizations plus cloud optical. Properties in the ECMWF model. Mon Weather Rev 118(4):847–873. doi:10.1175/1520-0493(1990)118<0847:IOCTTR>2.0.CO;2

Nabat P, Somot S, Mallet M, Sevault F, Chiacchio M, Wild M (2014) Direct and semi-direct aerosol radiative effect on the Mediterranean climate variability using a coupled regional climate system model. Clim Dyn 1–29. doi:10.1007/s00382-014-2205-6

Noguer M, Jones R, Murphy J (1998) Sources of systematic errors in the climatology of a regional climate model over Europe. Clim Dyn 14(10):691–712. doi:10.1007/s003820050249

Oettli P, Sultan B, Baron C, Vrac M (2011) Are regional climate models relevant for crop yield prediction in West Africa? Environ

Res Lett 6(1):014008. http://stacks.iop.org/1748-9326/6/i=1/a=014008

Omrani H, Drobinski P, Dubos T (2012a) Investigation of indiscriminate nudging and predictability in a nested quasi-geostrophic model. Q J R Meteorol Soc 138(662):158–169. doi:10.1002/qj.907

Omrani H, Drobinski P, Dubos T (2012b) Spectral nudging in regional climate modelling: how strongly should we nudge? Q J R Meteorol Soc 138(668):1808–1813. doi:10.1002/qj.1894

Onof C, Chandler RE, Kakou A, Northrop P, Wheater HS, Isham V (2000) Rainfall modelling using poisson-cluster processes: a review of developments. Stoch Environ Res Risk Assess 14(6):384–411. doi:10.1007/s004770000043

Palmer TN, Shukla J (2000) Editorial. Q J R Meteorol Soc 126(567):1989–1990. doi:10.1002/qj.49712656701

Pavan V, Doblas-Reyes FJ (2000) Multi-model seasonal hindcasts over the Euro-Atlantic: skill scores and dynamic features. Clim Dyn 16(8):611–625. doi:10.1007/s003820000063

Perrone TJ, Miller RG (1985) Generalized exponential markov and model output statistics: a comparative verification. Mon Weather Rev 113(9):1524–1541. doi:10.1175/1520-0493(1985)113<1524:GEMAMO>2.0.CO;2

Piani C, Weedon G, Best M, Gomes S, Viterbo P, Hagemann S, Haerter J (2010) Statistical bias correction of global simulated daily precipitation and temperature for the application of hydrological models. J Hydrol 395(3–4):199–215. doi:10.1016/j.jhydrol.2010.10.024. http://www.sciencedirect.com/science/article/pii/S0022169410006475

Radanovics S, Vidal JP, Sauquet E, Ben Daoud A, Bontron G (2013) Optimising predictor domains for spatially coherent precipitation downscaling. Hydrol Earth Syst Sci 17(10):4189–4208. doi:10.5194/hess-17-4189-2013. http://www.hydrol-earth-syst-sci.net/17/4189/2013/

Raje D, Mujumdar P (2010) Reservoir performance under uncertainty in hydrologic impacts of climate change. Adv Water Resour 33(3):312–326. doi:10.1016/j.advwatres.2009.12.008. http://www.sciencedirect.com/science/article/pii/S0309170810000047

Ricard J, Royer J (1993) A statistical cloud scheme for use in an AGCM. Ann Geophys 11(12):1095–1115

Ruti PM, Williams JE, Hourdin F, Guichard F, Boone A, Van Velthoven P, Favot F, Musat I, Rummukainen M, Domínguez M, Gaertner MA, Lafore JP, Losada T, Rodriguez de Fonseca MB, Polcher J, Giorgi F, Xue Y, Bouarar I, Law K, Josse B, Barret B, Yang X, Mari C, Traore AK (2011) The west african climate system: a review of the amma model inter-comparison initiatives. Atmos Sci Lett 12(1):116–122. doi:10.1002/asl.305

Sachindra DA, Huang F, Barton AF, Perera BJC (2014) Multi-model ensemble approach for statistically downscaling general circulation model outputs to precipitation. Q J R Meteorol Soc 140(681):1161–1178. doi:10.1002/qj.2205

Salameh T, Drobinski P, Vrac M, Naveau P (2009) Statistical downscaling of near-surface wind over complex terrain in southern France. Meteorol Atmos Phys 103(1–4):253–265. doi:10.1007/s00703-008-0330-7

Sanders F (1963) On subjective probability forecasting. J Appl Meteorol 2(2):191–201. doi:10.1175/1520-0450(1963)002<0191:OSPF>2.0.CO;2

Schmidli J, Goodess CM, Frei C, Haylock MR, Hundecha Y, Ribalaygua J, Schmith T (2007) Statistical and dynamical downscaling of precipitation: an evaluation and comparison of scenarios for the European Alps. J Geophys Res Atmos 112(D4). doi:10.1029/2005JD007026

Schnur R, Lettenmaier DP (1998) A case study of statistical downscaling in Australia using weather classification by recursive partitioning. J Hydrol 212–213(0):362–379. doi:10.1016/S0022-1694(98)00217-0. http://www.sciencedirect.com/science/article/pii/S0022169498002170

Schoof J, Pryor S (2001) Downscaling temperature and precipitation: a comparison of regression-based methods and artificial neural networks. Int J Climatol 21(7):773–790. doi:10.1002/joc.655

Semenov MA, Stratonovitch P (2010) Use of multi-model ensembles from global climate models for assessment of climate change impacts. Clim Res 41(1):1–14. doi:10.3354/cr00836. http://www.int-res.com/abstracts/cr/v41/n1/p1-14/

Semenov MA, Brooks RJ, Barrow EM, Richardson CW (1998) Comparison of the WGEN and LARS-WG stochastic weather generators for diverse climates. Clim Res 10(2):95–107. doi:10.3354/cr010095. http://www.int-res.com/abstracts/cr/v10/n2/p95-107/

Seth A, Giorgi F (1998) The effects of domain choice on summer precipitation simulation and sensitivity in a regional climate model. J Clim 11(10):2698–2712. doi:10.1175/1520-0442(1998)011<2698:TEODCO>2.0.CO;2

Skamarock W, Klemp J, Dudhia J, Gill D, Barker D, Duda M, Huang X, Wang W, Powers J (2008) A description of the advanced research wrf version 3. Technical Report, NCAR

Smirnova TG, Brown JM, Benjamin SG (1997) Performance of different soil model configurations in simulating ground surface temperature and surface fluxes. Mon Weather Rev 125(8):1870–1884. doi:10.1175/1520-0493(1997)125<1870:PODSMC>2.0.CO;2

Solman S, Sanchez E, Samuelsson P, da Rocha R, Li L, Marengo J, Pessacg N, Remedio A, Chou S, Berbery H, Le Treut H, de Castro M, Jacob D (2013) Evaluation of an ensemble of regional climate model simulations over South America driven by the era-interim reanalysis: model performance and uncertainties. Clim Dyn 41(5–6):1139–1157. doi:10.1007/s00382-013-1667-2

Stephens GL, L'Ecuyer T, Forbes R, Gettlemen A, Golaz JC, Bodas-Salcedo A, Suzuki K, Gabriel P, Haynes J (2010) Dreary state of precipitation in global models. J Geophys Res Atmos 115(D24). doi:10.1029/2010JD014532

Stern RD, Coe R (1984) A model fitting analysis of daily rainfall data. J R Stat Soc Ser A (Stat Soc) 147(1):1–34

Sun Y, Solomon S, Dai A, Portmann RW (2006) How often does it rain? J Clim 19(6):916–934. doi:10.1175/JCLI3672.1

Takle ES, Gutowski WJ, Arritt RW, Pan Z, Anderson CJ, da Silva RR, Caya D, Chen SC, Giorgi F, Christensen JH, Hong SY, Juang HMH, Katzfey J, Lapenta WM, Laprise R, Liston GE, Lopez P, McGregor J, Pielke RA, Roads JO (1999) Project to intercompare regional climate simulations (PIRCS): description and initial results. J Geophys Res Atmos 104(D16):19443–19461. doi:10.1029/1999JD900352

Vautard R, Yiou P (2009) Control of recent European surface climate change by atmospheric flow. Geophys Res Lett 36(22). doi:10.1029/2009GL040480

Vautard R, Gobiet A, Jacob D, Belda M, Colette A, Déqué M, Fernández J, García-Díez M, Goergen K, Güttler I, Halenka T, Karacostas T, Katragkou E, Keuler K, Kotlarski S, Mayer S, Meijgaard E, Nikulin G, Patarčić M, Scinocca J, Sobolowski S, Suklitsch M, Teichmann C, Warrach-Sagi K, Wulfmeyer V, Yiou P (2013) The simulation of European heat waves from an ensemble of regional climate models within the EURO-CORDEX project. Clim Dyn 41(9–10):2555–2575. doi:10.1007/s00382-013-1714-z

Vigaud N, Vrac M, Caballero Y (2013) Probabilistic downscaling of GCM scenarios over southern India. Int J Climatol 33(5):1248–1263. doi:10.1002/joc.3509

Vischel T, Lebel T, Massuel S, Cappelaere B (2009) Conditional simulation schemes of rain fields and their application to rainfall-runoff modeling studies in the Sahel. J Hydrol 375(1–2):273–286. doi:10.1016/j.jhydrol.2009.02.028. http://www.sciencedirect.com/science/article/pii/S0022169409000900

(Surface processes and water cycle in West Africa, studied from the AMMA-CATCH observing system)

Vrac M, Friederichs P (2014) Multivariate–intervariable, spatial, and temporal–bias correction. J Clim 28(1):218–237. doi:10.1175/JCLI-D-14-00059.1

Vrac M, Naveau P (2007) Stochastic downscaling of precipitation: from dry events to heavy rainfalls. Water Resour Res 43(7). doi:10.1029/2006WR005308

Vrac M, Marbaix P, Paillard D, Naveau P (2007a) Non-linear statistical downscaling of present and LGM precipitation and temperatures over Europe. Clim Past 3(4):669–682. doi:10.5194/cp-3-669-2007. http://www.clim-past.net/3/669/2007/

Vrac M, Stein ML, Hayhoe K (2007b) Statistical downscaling of precipitation through nonhomogeneous stochastic weather typing. Clim Res 34(3):169–184. doi:10.3354/cr00696. http://www.int-res.com/abstracts/cr/v34/n3/p169-184/

Vrac M, Stein ML, Hayhoe K, Liang XZ (2007c) A general method for validating statistical downscaling methods under future climate change. Geophys Res Lett 34(18). doi:10.1029/2007GL030295

Vrac M, Drobinski P, Merlo A, Herrmann M, Lavaysse C, Li L, Somot S (2012) Dynamical and statistical downscaling of the french mediterranean climate: uncertainty assessment. Nat Hazards Earth Syst Sci 12(9):2769–2784. doi:10.5194/nhess-12-2769-2012. http://www.nat-hazards-earth-syst-sci.net/12/2769/2012/

Vrac M, Vaittinada Ayar P, Yiou P (2014) Trends and variability of seasonal weather regimes. Int J Climatol 34(2):472–480. doi:10.1002/joc.3700

van Vuuren D, Edmonds J, Kainuma M, Riahi K, Thomson A, Hibbard K, Hurtt G, Kram T, Krey V, Lamarque JF, Masui T, Meinshausen M, Nakicenovic N, Smith S, Rose S (2011) The representative concentration pathways: an overview. Clim Change 109(1–2):5–31. doi:10.1007/s10584-011-0148-z

Wilby R, Wigley T (1997) Downscaling general circulation model output: a review of methods and limitations. Prog Phys Geogr 21(4):530–548. doi:10.1177/030913339702100403. http://ppg.sagepub.com/content/21/4/530.abstract. http://ppg.sagepub.com/content/21/4/530.full+html

Wilby RL, Dawson CW, Barrow EM (2002) SDSM—a decision support tool for the assessment of regional climate change impacts. Environ Model Softw 17(2):145–157

Wilks DS (2010) Use of stochastic weathergenerators for precipitation downscaling. Wiley Interdiscip Rev Clim Change 1(6):898–907. doi:10.1002/wcc.85

Wilks DS (2012) Stochastic weather generators for climate-change downscaling, part ii: multivariable and spatially coherent multisite downscaling. Wiley Interdiscip Rev Clim Change 3(3):267–278. doi:10.1002/wcc.167

Wingo MT, Cecil DJ (2009) Effects of vertical wind shear on tropical cyclone precipitation. Mon Weather Rev 138(3):645–662. doi:10.1175/2009MWR2921.1

Witten DM, Tibshirani R, Hastie T (2009) A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. Biostatistics 10(3):515–534. doi:10.1093/biostatistics/kxp008. http://biostatistics.oxfordjournals.org/content/10/3/515.abstract. http://biostatistics.oxfordjournals.org/content/10/3/515.full+html

Xiaoli L, Coulibaly P, Evora N (2008) Comparison of data-driven methods for downscaling ensemble weather forecasts. Hydrol Earth Syst Sci 12(2):615–624. doi:10.5194/hess-12-615-2008. http://www.hydrol-earth-syst-sci.net/12/615/2008/

Yang C, Chandler RE, Isham VS, Wheater HS (2005) Spatial–temporal rainfall simulation using generalized linear models. Water Resour Res 41(11):W11415. doi:10.1029/2004WR003739

Yang W, Bárdossy A, Caspary HJ (2010) Downscaling daily precipitation time series using a combined circulation- and regression-based approach. Theor Appl Climatol 102(3–4):439–454. doi:10.1007/s00704-010-0272-0

Yee TW (2010) The VGAM package for categorical data analysis. J Stat Softw 32(10):1–34. http://www.jstatsoft.org/v32/i10

Yiou P (2014) AnaWEGE: a weather generator based on analogues of atmospheric circulation. Geosci Model Dev 7(2):531–543. doi:10.5194/gmd-7-531-2014. http://www.geosci-model-dev.net/7/531/2014/

Yiou P, Nogaj M (2004) Extreme climatic events and weather regimes over the North Atlantic: when and where? Geophys Res Lett 31(7). doi:10.1029/2003GL019119

Yiou P, Vautard R, Naveau P, Cassou C (2007) Inconsistency between atmospheric dynamics and temperatures during the exceptional 2006/2007 fall/winter and recent warming in Europe. Geophys Res Lett 34(21). doi:10.1029/2007GL031981

Yiou P, Salameh T, Drobinski P, Menut L, Vautard R, Vrac M (2013) Ensemble reconstruction of the atmospheric column from surface pressure using analogues. Clim Dyn 41(5–6):1333–1344. doi:10.1007/s00382-012-1626-3

Zorita E, von Storch H (1999) The analog method as a simple statistical downscaling technique: comparison with more complicated methods. J Clim 12(8):2474–2489. doi:10.1175/1520-0442(1999)012