Testing MOS-like Analog Precipitation Downscaling for ENSEMBLES Regional Climate Models over Spain

M. Turco[1], P. Quintana-Seguí[2], M.C. Llasat[1], S. Herrera[3], J. M. Gutiérrez[3]

M. Turco, GAMA (Meteorological Hazards Analysis Team), Dpt. of Astronomy & Meteorology, Faculty of Physics, University of Barcelona, Av. Diagonal 647, Barcelona, E-08028, Spain.(mturco@am.ub.es)

M. C. Llasat, GAMA (Meteorological Hazards Analysis Team), Dpt. of Astronomy & Meteorology, Faculty of Physics, University of Barcelona, Av. Diagonal 647, Barcelona, E-08028, Spain.(carmell@am.ub.es)

P. Quintana-Seguí, Observatori de l'Ebre (Universitat Ramon Llull - CSIC), Horta Alta 38, Roquetes, 43520, Spain. (pquintana@obsebre.es)

S. Herrera, Grupo de Meteorologia, Dpt. of Applied Mathematics and Computer Science, Universidad de Cantabria (UC), Av. de los Castros s/n, Santander, 39005, Spain. (herreras@unican.es)

J. M. Gutiérrez, Grupo de Meteorologia, Instituto de Física de Cantabria (IFCA), CSIC-UC, Av. de los Castros s/n, Santander, 39005, Spain. (gutierjm@unican.es)

[1]Universidad de Barcelona, Barcelona,

Abstract. Model Output Statistics (MOS) has been recently proposed as a convenient statistical downscaling alternative —as opposed to the standard perfect prognosis approach— for Regional Climate Model (RCM) outputs. In this case, the model output for the variable of interest (e.g. precipitation) is directly downscaled/calibrated using observations. In this paper we test the performance of a MOS implementation of the popular analog methodology applied to calibrate precipitation outputs over Spain. To this aim, we consider the state-of-the-art ERA40-driven RCMs provided by the EU-funded ENSEMBLES project and the *Spain02* gridded observations dataset, over the common period 1961-2000. The *MOS analog* method improves the representation of the mean regimes, the annual cycle, the frequency and the extremes of precipitation for all RCMs, regardless of the region and the model reliability (including relatively low-performing models), while preserving the daily accuracy. The good performance in different climates of Spain suggest the potential transferability of the method to other regions. Furthermore, in order to test the robustness of the method in changing climate conditions,

Spain

[2]Observatori de l'Ebre (URL-CSIC),

Roquetes, Spain

[3]Instituto de Física de Cantabria

CSIC-UC, Santander, Spain

a cross-validation in driest or wettest years was performed. The method improves the RCM results in both cases, especially in the former.

1. Introduction

Global Climate Models (GCM) are basic tools to study and simulate the climate, and to obtain future climate projections under different anthropogenic forcing scenarios [*Solomon et al.*, 2007]. However, due to their coarse resolution —generally few hundred kilometers— and their large biases, they are not suitable for regional studies [*Cohen*, 1990]. This is especially true for Spain, a geographically complex and heterogeneous region characterized by a great variability of precipitation regimes [*Serrano et al.*, 1999; *Trigo and Palutikof*, 2001]. Consequently, developing regional climate scenarios is a key problem for climate change impact/adaptation studies and has become a strategic topic in national and international climate programs [see, e.g. the WCRP CORDEX initiative, *Giorgi et al.*, 2009].

Usually, two different methodologies have been developed for downscaling GCM simulations over a region of interest (e.g. Europe). Firstly, *dynamical downscaling* is based on high resolution (e.g. 25 km) limited area models —also called Regional Climate Models (RCMs)— which are coupled at the boundaries to the GCM outputs [*Giorgi and Mearns*, 1991]. Secondly, *statistical downscaling* techniques [*Wilby et al.*, 2004; *Benestad et al.*, 2008] are based on statistical models, fitted to historical data to capture the empirical relationship between large-scale GCM variables (the predictors, e.g. 500mb geopotential) and local variables (the predictands, e.g. precipitation at a given location); these models are first trained using reanalysis data —following the Perfect Prognosis (PP) approach— and later applied to downscale GCM scenario outputs.

Traditionally, statistical downscaling has been used as a substitute for dynamical downscaling, or vice-versa. However, due to the increasing availability of reanalysis-driven RCM simulations, produced in projects like ENSEMBLES [*van der Linden and Mitchell*, 2009], some authors have recently suggested the possibility of combining the advantages of the two downscaling methodologies. The idea is applying the statistical downscaling directly to the RCM outputs following the Model Output Statistics (MOS) —or bias correction— approach [see *Maraun et al.*, 2010, and references therein]. In this case, as opposite to PP, the predictor is directly the RCM output variable (i.e. the RCM precipitation) which is calibrated to match the observed variable (local precipitation at a station or interpolated grid point). This alternative approach constitutes an advanced calibration method for end-users, allowing the calibration of RCM outputs for climate change impact studies [the importance of calibration in this context is discussed in several studies *Fowler et al.*, 2007; *Christensen et al.*, 2008; *Herrera et al.*, 2010a], and an increase of resolution (when high-resolution observations are available in the area of interest). Note that, although RCMs provide regional climate details compared to the GCMs, their resolution is still too coarse (typically tens of kilometers) for many climate change impact studies [see, e.g. *Quintana-Seguí et al.*, 2010].

In this study we analyze the state-of-the-art ensemble of ten RCMs produced in the ENSEMBLES project using a 25km grid, considering the ERA40-driven simulation in the control period 1961-2000, and focusing on precipitation over Spain [see *Kjellstrm et al.*, 2010, and other papers in the same special issue]. In a recent paper, *Herrera et al.* [2010a] show that some of these models have strong biases and exhibit a poor performance when reproducing the mean precipitation regime and annual cycle in this region. In

addition, they overestimate the frequency of rainfall and they deficiently represent the extreme events. In this paper, these models are statistically post-processed applying a downscaling technique based on analogs [*Lorenz*, 1969] to their precipitation fields, following the MOS approach. This study has two main objectives: (1) testing the skill of a MOS-like downscaling method for mean and extreme precipitation in a complex area (with both Atlantic and Mediterranean climates) at daily scale and (2) evaluating the possibility to obtain more homogeneous and calibrated ensembles by improving the reliability of the worst-performing RCMs (those with higher biases and larger improving potential).

The study is organized as follows. In Section 2, a short description of the precipitation characteristics in Spain is given and the RCM and observational datasets used in the paper are described; Section 3 presents the downscaling method used, and Section 4 analyzes the validation results. Finally, Section 5 synthesizes the main results and conclusions of this study.

2. Region of Study and Data

The Iberian peninsula —located on the south-west edge Europe, between 36° and 44°N and 10°W and 3°E— is an important region for precipitation studies for two main reasons. Firstly, precipitation plays a major role on water resources and natural hazards in this area [*Garrote et al.*, 2007; *Llasat*, 2009], thus leading to one of the most vulnerable countries to water scarcity, droughts and floods in Europe [*Kristensen*, 2010]. Secondly, its complex orography and particular location —at the transition area between extra-tropical and sub-tropical influence [*Jansá*, 1997; *Giorgi and Lionello*, 2008]— determines a great variety of climates with both Atlantic and Mediterranean influences. Thus, precipitation is charac-

terized by a complex spatial pattern [*Serrano et al.*, 1999], with a strong seasonal cycle and large interannual [*Trigo and Palutikof*, 2001] and spatial [*Rodriguez-Puebla et al.*, 1998; *Romero et al.*, 1998; *Martin-Vide*, 2004; *Rodrigo and Trigo*, 2007]) variability. The annual precipitation decreases from north-west (with a typical Atlantic precipitation regime) to south-east (with a Mediterranean precipitation regime). The north has the largest accumulated values (1000-2500 mm/year) with a maximum in winter and rainfall spread out over the year. The majority of the central part of the peninsula shaves less than 500 mm/year. The south-east is characterized by a semiarid climate with very dry condition, with areas with less than 100 mm/year. Finally, the Mediterranean coast and part of the Ebro basin exhibit bimodal Autumn-Spring maxima with an accumulated annual values of less than 700 mm/year, where frequent drought periods alternate to heavy rainfall events [*Llasat*, 2009]. Due to this strong variability, Spain represents a challenge area for downscaling studies [*Herrera et al.*, 2010a].

In the following we describe the ensemble of ten RCM simulations from the ENSEMBLES project, and the *Spain02* gridded dataset of interpolated observations.

2.1. Interpolated Observations: *Spain02*

The observed data of daily precipitation used in this study is provided by the high-resolution (0.2° x 0.2°, approximately 20km x 20 km) gridded dataset *Spain02* [*Herrera et al.*, 2010b], which is publicly available for research activities. This dataset was produced using data from 2756 quality-controlled stations from the Spanish Meteorological Agency (AEMET), covering the Iberian Peninsula and the Balearic Islands over the period 1950-2008 (Figure 1). This dataset has been used by *Herrera et al.* [2010a] to evaluate the

RCMs described in Sec. 2.2, assessing their performance to reproduce both the mean and extreme precipitation regimes.

2.2. ENSEMBLES RCM dataset

The EU-funded project ENSEMBLES produced an ensemble of regional simulations at a 25 km resolution using state-of-the-art RCMs [*van der Linden and Mitchell*, 2009] driven by both ERA40 reanalysis data [*Uppala et al.*, 2005] in a control period, and future A1B scenario simulations of different GCMs. In this paper we consider the ERA40-driven runs from ten models for the common period 1961-2000 (see Table1). The main advantage of these runs is the daily accuracy (or day-to-day correspondence with observations) of the reanalysis and, to some extent, of the corresponding RCM simulations (forced at the boundaries by the reanalysis). As we shall see in Sec. 4.3, this property is the key reason for the successful application of the MOS approach in this context. Moreover, the resulting validations is a characteristic of each particular RCM, since the reanalysis is a pseudo-observation of the atmosphere (at least in this extratropical region), therefore excluding the systematic biases of the global climate models.

Herrera et al. [2010a] evaluated the mean and extreme precipitation regimes from these RCMs over Spain (with the exception of the ICTP model, not available at the time of their work) and reported a subset of five models best performing over this region (indicated with an asterisk in Table 1). The resulting 5-model ensemble performed better than the individual models and than the total ensemble.

For practical reason, the daily outputs of the RCMs were bilinearly interpolated from their original resolution (25km) to the grid defined by *Spain02* ($0.2° \times 0.2°$, 20 km approximately). This manipulation might decrease the quality of the simulated data; however,

in this case the data will be used as predictor for the downscaling method and thus the interpolation does not influence the final results. Of course, if the method were to be used to assess the impact of climate change, this interpolation should be avoided.

3. Methodology: MOS based on Analogs

On the one hand, perfect prognosis (PP) is the most popular and widely used statistical downscaling methodology at seasonal and climate change scales [see *Benestad et al.*, 2008, and references therein]. In this case, a statistical link is derived from the reanalysis — quasi-observations— large scale predictors (e.g. sea level pressure) and the observed local predictands (e.g. precipitation) and, then, applied to future simulations from GCMs under different forcing scenarios. On the other hand, it has been recently suggested that the MOS methodology could be directly applied to the RCM outputs, using a reanalysis-driven control period to model, calibrate and validate the methods. In this case, the predictor is directly the RCM output variable (i.e. the predicted precipitation) which is corrected to match the observed variable. Thus, this alternative methodology may overcome some of the known drawbacks of the PP methods: like the underestimation of high intensity events [*Wilks and Wilby*, 1999] and local spatial and temporal variability [*Maraun et al.*, 2010]. Besides the PP methods do not necessarily provide consistency between different downscaled variables [*Wilby and Wigley*, 1997].

Some MOS methods have been recently proposed in the literature to correct RCM simulations by an additive term for temperature [*Déqué*, 2007] or by a scaling factor for precipitation [*Widmann et al.*, 2003]. Quantile mapping attempts to correct the whole distribution [*Déqué*, 2007; *Piani et al.*, 2010]. MOS methods are still in a rather premature state of development and substantial improvements are currently under development.

In this paper, we present a MOS method based on the adaptation of the analog methodology (hereafter referred to as "*MOS analog*"). This method was first developed for weather forecasting [*Lorenz*, 1969; *Obled et al.*, 2002; *Gibergans-Baguena and Llasat*, 2007] and later applied to climate scales [*Zorita et al.*, 1995]. After that, several climate studies were performed using the analog method [see, e.g. *Cubasch et al.*, 1996; *Zorita and von Storch*, 1999; *Timbal et al.*, 2003; *Benestad et al.*, 2008], so it is nowadays a popular and widely used technique. The analog method is based on the hypothesis that "analogue" weather patterns (predictors) should cause "analogue" local effects (predictands). This leads to a simple algorithm to infer the local occurrence associated with a given predictor (atmospheric pattern) based on the historical occurrences of a set of analog days (with closest predictors). This is simply done by considering the historical local occurrences corresponding to the atmospheric patterns closest/analog to the target predictor. The main advantages of this method are that (1) it is able to reproduce nonlinear relationship between predictors and predictands, (2) it is easy to implement with low computational cost, and (3) it is able to reproduce spatially coherent and realistic precipitation patterns.

The main drawback of the method is that it cannot simulate unobserved weather patterns, although it can produce accumulated values or frequencies over several days larger (or smaller) than the historical values. This limitation is related to the assumption of "stationarity" [*Wilby et al.*, 2004], a common weakness of all the downscaling methods —the parametrizations of the dynamical models and the statistical relationship between predictors and predictands must hold in the projected climate, which cannot be taken for granted [*Trenberth et al.*, 2003].— This limitation should be cautiously taken into account for climate change studies, although this problem can be mitigated using a long database

of observations with a great variety of situations [*Zorita and von Storch*, 1999] and using robust statistical relationships based on a small number of parameters and on a physical predictor/predictand relationship [*Benestad et al.*, 2008; *Maraun et al.*, 2010]. This is the case of the *MOS analog* method, where a unique predictor (model precipitation) is used.

Given an historical training period (with known predictors and predictands) and a projection period (with known predictors), the *MOS analog* downscaling consists of three main steps to estimate the corresponding projected predictands:

1. Selection of an appropriate subgrid within the RCM domain over the area of study, capturing the physical scales relevant for the predictand of interest (observed precipitation in this case). In our case we consider the 0.2° subgrid covering the Iberian peninsula described in Sec. 2.2 and consider the predictor pattern defined by the RCM precipitation on this grid.

2. For each predictor pattern from the projection period, the closest historical pattern (analog) within the training period is computed considering the Euclidean distance (according to *Matulla et al.* [2007] this is a reasonable first choice among the standard measures of similarity). A larger number of analogs were also considered, but a single analog exhibited the best performance according to the validation metrics considered in the study.

3. Then, the local precipitation projected for the predictand (*Spain02*) is simply obtained as the historical occurrence of the predictand on the analog day.

In order to select the predictor domain, different experiments were performed. We focused on the Ebro basin (see Figure 1), which is a demanding test for a downscaling method due to the great variability of the precipitation of this basin (due to Atlantic

and Mediterranean influences). Three predictor domains were considered: the *Spain02* domain (i.e. the Spanish Iberian Peninsula and the Balearic Islands), the Mediterranean coast (i.e. the union of the Mediterranean river basins show in Figure 1), and, finally, the Ebro basin itself. Several validation experiments were performed using these domains with different RCMs and different train/test periods. It came out that the skill was more influenced by the different test periods or different RCMs than by the different domains. For this reason, the predictor domain for our experiments over the Iberian Peninsula was *Spain02*.

4. Validation and Results

The skill of the MOS analog method has been evaluated using a cross-validation approach, considering reanalysis data and observations within the period 1961-2000. The data was split into two subsets, 30 years for model training/calibration and 10 years for validation. To test the robustness of the statistical relationship in a changing climate, two different test periods were used: The ten wettest years and the ten driest years, respectively. Since the annual precipitation in Spain does not exhibit any general trend [*Río et al.*, 2010], the wettest (driest) years have been identified in the following way: The annual total precipitation for each point has been standardized, spatially averaged and finally sorted. The resulting wettest (driest) years are given in Table 4. Note that cross validation requires that the test and training samples are randomly drawn from the population; thus, although the ten wettest (driest) years do not conform a proper test sample, our objective is measuring the performance of the method in changing climate conditions and, hence the present cross-validation procedure provides a more informative

assessment of the downscaling methodology regarding its suitability for future scenario simulations.

Two main approaches have been applied to evaluate the skill of the downscaling [see *Murphy*, 1993, for a description of forecast validation]. Firstly, we compare the simulated (both RCM outputs and MOS downscaled ones) and observed climatologies (spatial patterns) considering standard reliability measures (Sec. 4.1) and the annual cycle (Sec. 4.2). Secondly, since the reanalysis-driven RCM simulations acquire certain day-to-day correspondence with observations, the simulated and observed time series are also compared at a grid-point basis using standard accuracy measures (Section 4.3).

4.1. Reliability of the mean and extreme climates

The ability of RCMs and MOS analog to reproduce the annual climatology (spatial pattern) for the precipitation indices shown in Table 3 has been tested. These indices were computed working with daily data and are a subset of the standard ETCCDI indices characterizing total precipitation, dry and wet spells and extremes [*WMO*, 2009].

Simple performance scores (bias, mean absolute error and correlation) were computed for the spatial pattern of the annual indices and averaged over the ten year wet (dry) validation periods, respectively:

- *ME*: Normalized spatial mean error (or bias)

$$ME = \frac{1}{n \cdot \overline{O}} \sum_{i=1}^{n} (Y_i - O_i) \tag{1}$$

- *MAE*: Normalized spatial mean absolute error

$$MAE = \frac{1}{n \cdot \overline{O}} \sum_{i=1}^{n} |Y_i - O_i| \tag{2}$$

where Y_i and O_i are the simulated and observed indices, respectively, for the *i-th* grid-point ($n = 1445$), averaged over the ten year period of validation. Note that these values are normalized to the spatial mean of the observations $\overline{O}$ [*Bachner et al.*, 2008].

- *CORR*: Spatial correlation calculated by the Spearman rank correlation coefficient.

$$CORR = 1 - \frac{6 \cdot \sum_{i=1}^{n} D_i^2}{n \cdot (n^2 - 1)}, \tag{3}$$

Where D_i is the difference in ranks of the *i-th* data pair (Y_i, O_i). Note that the Spearman correlation is more robust to outliers and linearity than the classical Pearson correlation.

These scores were calculated both for the original RCM simulations and for the *MOS analog* downscaled values (ME_1 and ME_2, respectively, for the first score) and the resulting differences were statistically tested for significance (the null hypothesis is $ME_1 - ME_2 = 0$) applying bootstrap resampling with 1000 realizations, obtaining the 95% confidence intervals [*Efron and Tibshirani*, 1993]. *Bachner et al.* [2008] applied a similar test to evaluate the skill differences among RCMs.

As an illustrative example, and for the sake of simplicity, in Figure 2 we show the comparison maps for the KNMI model and the corresponding *MOS analog* values for the wet test period; note that this RCM has been chosen since it is one of the most skillful for precipitation in this region [see *Herrera et al.*, 2010a, and Figure 3]. The panels in this figure show the annual values of the indices (averaged in the validation period of ten years) for the observed grid *Spain02* (first column) the *MOS analog* downscaled values (second column) and the regional KNMI simulations (third column); the numbers below the figures indicate the correlation ($CORR$), bias (ME) and mean absolute (MAE) values for the MOS and RCM values with regards to the observed ones (an asterisk indicate those values where the MOS/RCM has a significantly better performance than the RCM/MOS,

respectively, at a 95% level). This figure shows that the MOS downscaled values clearly outperform the uncalibrated RCM outputs, with significant differences in most of the cases.

Figure 3 summarizes the verification results for all the models and the scores considered. The RCMs have been ranked from 1 to 10 according to the correlation value of total precipitation ($PRCPTOT$) for the wet period (i.e. according to the first score, in the upper left panel of the figure) [this ranking agrees with *Herrera et al.*, 2010a]. The 95% confidence interval for each individual score is also shown, as a vertical line displayed over the MOS downscaled values (filled circles). Thus, when the RCM values (circles) are outside this interval, the differences are statistically significant at a 95% level. The values above (or below) the upper (or lower) axis bounds are displayed as grey shaded circles; for instance, correlations smaller than 0.5 are not shown in the figure and, hence, cases with smaller values are just marked with a grey shaded circle. This figure shows that, overall, the same correlation and error patterns are obtained for wet (upper panels) and dry (lower) test periods, with slightly better results in the later case. The *MOS analog* downscaling method dramatically improves the RCM results for $PRCPTOT$, $SDII$, CWD, $R10$ and $R20$, with correlation values larger than 0.9 in all cases and with smaller $MAEs$ and biases. The improvement is also evident for the extremes $RX1DAY$ and $RX5DAY$, with correlations larger than 0.8; however, in this case the error and bias improvement is smaller than for the previous scores.

Regarding the ME (bias), even thought the *MOS analog* tends to underestimate the indices studied, it is able to reduce the ME of the RCMs in the dry period —with the exception of $RX5DAY$ index,— while it leads to similar or worse ME in the majority of

cases for the wet period. The *MOS analog* downscaling method is also able to improve the error (MAE score) for all the RCMs and for all the indices considered, with few exceptions for some RCMs in the case of $RX5DAY$ and CWD, considering the wet test period. Generally, the MAE of the *MOS analog* is slightly bigger when it is tested in the wettest period. This may be a result of the relatively short training period the *MOS analog* was based on, with a relatively low sampling of the heavy precipitation amount.

Note that although $PRCPTOT$ is overestimated for most of the RCMs, the precipitation on wet days ($SDII$) is underestimated. This problem is due to the overestimation of rainfall frequency by RCMs, as they tend to drizzle [see, e.g. *Gutowski et al.*, 2003]. However, as it is shown in this figure, the *MOS analog* solves this problem, leading to unbiased estimates of both indices (except in the wet test period, where the downscaled total precipitation is slightly underestimated).

The index CDD is the one with worst performance for the *MOS analog* method, providing only a slight improvement over the values of the RCM. In fact, our algorithm is able to improve the ME and MAE of the CDD, but not always its correlation (in particular in the case of the ITCP model, probably due to the great overestimation of the rainy days by this model). As we show in Sec. 4.3, the CDD index is highly sensitive to the autocorrelation of the time series and, consequently, to the accuracy of the RCM. This should be further investigated by adding temporal constrains in the choice of the analog.

Overall, the *MOS analog* method is able to improve the above considered reliability scores for all RCMs, thus attaining an appropriate calibration in all cases, regardless of their respective skills. This is the main advantage of the *MOS analog* methodology, based on a resampling of the observed space driven by the historical analogs of RCM fields. As

we show in Sec. 4.3 this calibration is done preserving the daily accuracy of the RCM and, thus, the downscaled output can be considered a calibrated local version of the RCM values. This is an important result since it permits to enlarge the ensemble of RCMs avoiding discarding those with bad reliability, since they can have a similar accuracy and could be calibrated as shown in this work.

Finally, the results reported in this section show that, although the *MOS analog* downscaling improved the RCM results in wet and dry periods, the added value in the former period is less evident (since it cannot simulate unobserved weather patterns) and, consequently, it should be cautiously considered in the projection of future climate scenarios.

4.2. Validation over the annual cycle

As already mentioned, the precipitation in Spain is characterized by a large variability in space and time. In particular the Iberian rainfall has a strong seasonal cycle that differs considerably among the river basins shown in Fig. 1a. In the previous section we evaluated the performance of the *MOS analog* method to represent the annual climatologies of different indices. In this section, in order to assess the correspondence of the simulated and observed annual cycles, we analyze the performance of the methods in the different basins at a monthly basis. A recent study has shown the capability of RCMs to simulate the annual precipitation cycle in the different basins, specially using a five-member ensemble formed by the best performing RCMs [*Herrera et al.*, 2010a]. In this section we also consider this five-model ensemble (indicated by stars in Table 1), but we focus on extremes. Thus, we consider the annual cycle of the $RX1DAY$, i.e. the monthly maximum value, averaging the grid point indices at a basin level, thus providing useful information for hydrological studies. It is important to underline that the spatial averages smooths

the peaks, since the distribution is not uniform over the area; to analyse this effect the calculations were repeated considering the standardized (to zero mean and unit variance) individual point series, obtaining similar results in qualitatively terms (e.g. the shape of the different precipitation regimes). For this reason we show the spatial averaged series calculated with the original series, since they provide useful quantitative information.

Figure 4 considers the wet test period, showing the observed $RX1DAY$ values (black line) and the simulated values for the ensemble of RCMs (light shade) and *MOS analog* values (dark shade). For a better comparison, all the plots have the same scale, ranging from 0 to 60 mm. The annual cycle is reproduced quite properly by both the RCMs and *MOS analog* downscalings, with a reduced spread (smaller uncertainty) in the later case. Similar results have been obtained in the dry test period, or considering the full ensemble (not shown). In the Mediterranean basin (Segura, Levante, Ebro, Catalana and Baleares), the $RX1DAY$ cycle presents two maximum periods, the major one in autumn (in the range of around 20-35 mm) and the secondary in spring (around 15-25 mm), although the amounts differ among the basins. This characteristic is also present considering the total precipitation instead of the maximum value and it is a representative aspect of the Western Mediterranean climatology [*Romero et al.*, 1998]. Here, the $RX1DAY$ values are usually due to convective events [*Llasat*, 2001], leading also to a higher RCM spread —note that the convective parametrization schemes are an important source of error in RCM simulated precipitation [*Hohenegger et al.*, 2008].— The remaining basins have a maximum in winter, with values in the range 20-35 mm, and a minimum in summer, with values ranging from 5 mm (Baleares, Guadalquivir and Sur) to 20 mm (North basin).

346 The performance of the *MOS analog* method to reproduce the observed seasonal cycles
347 in the different basins is quite remarkable, with the only exception of the autumn months
348 (mainly September) in the Segura and Levante Mediterranean basins, where the series
349 show the maximum values (also the maximum spread), which is underestimated by the
350 *MOS analog*. In order to better investigate this aspect, the $RX1DAY$ differences among
351 the different RCMs (and the corresponding *MOS analog* values) and the observed series for
352 September have been reported in Fig. 5, for each Mediterranean basin and for the two test
353 periods (wet and dry, in rows). The different colors in the figure show relative errors, i.e.,
354 the absolute difference of simulated and observed value divided by the observed value. The
355 biggest errors appear for the Segura, Levante and Baleares river basins. It is remarkable
356 how the *MOS analog* downscaling method reduces the error of the corresponding RCM
357 during the test dry period; however, it has similar or worst performance during the wet
358 periods.

359 Regarding the ten models analyzed in Fig. 5 in the five Mediterranean basins, ETHZ
360 and SMHI are the only ones with errors lower than 50% in all cases. Moreover, DMI
361 and KNMI have a single case (both in Segura basin) with errors larger than 50%. There
362 is no best RCM for all basins and situations. This variability of the performance of the
363 RCMs supports the use of an ensemble of RCM simulations in impact studies, both for
364 improving the performance and for estimating the uncertainty. Another consequence is
365 that disregarding one RCM because performs poorly in one fixed period in a certain area
366 could lead to a loss of valuable information for other periods/areas.

4.3. Accuracy of the Daily Series

In this section we test the daily accuracy of RCM simulations and the corresponding *MOS analog* values. To this aim, at each grid box, we computed the relative mean absolute error (as in Eq. 1) and the Spearman correlation between the simulated series and the observations. Table 4.3 summarizes the results for all models and test periods considered. The different performance metrics are provided in columns and the RCMs and test periods in rows. This table shows that the *MOS analog* technique greatly improves the correlation and preserves (or slightly improve) the RCM error (MAEr), with smaller spatial variability of the results (see, e.g. the $MAEr$ quantities in Table 4.3).

In order to illustrate the spatial distribution of the scores, for the sake of simplicity we only consider here the ETHZ model (see Fig. 6); the numbers above the figure indicate the spatial median and the interquartile range (IQR) of the scores considered, as in Table 4.3. The best correlation scores for the ETHZ model are obtained in central-south Spain, while lower correlation values are along the mountains on the north and the Mediterranean; this pattern is common to all the RCMs analyzed, with a west/east decreasing skill. However, the MOS analog technique provides a more uniform correlation pattern with low values restricted to the North. Considering the two test periods, it can be seen that the *MOS analog* shows lower correlation during the wet period than in the dry one, whereas the RCMs do not have this correlation dependence on the test period. Nevertheless, although the MOS correlation decreases in the wet period, it is still better than the RCM. This improvement is mainly related to ability of the *MOS analog* to reduce the drizzle days of the RCMs. Indeed, considering only the rain days ($> 1mm$), the correlation pattern is similar among the *MOS analog* and the respective RCMs, with a west to east gradient, with values of the same order, around 0.3. Note that aggregating the daily time series

at 5 days the correlations (similar values between the MOS and the respective RCM) are around 0.65 while considering 10 days they are around 0.70.

These results are also valid when considering the seasonal series instead of the annual ones. Indeed, the measures of accuracy have values of the same order of magnitude, with correlations around to 0.65 (0.25/0.30 if only the rain days are considered, with lower values in summer).

Finally, it has been tested how the daily accuracy of the RCMs influence the reliability of the *MOS analog* downscaling method. This was achieved by comparing the original MOS downscaled series with several surrogated series obtained applying the MOS method to the same test period, but considering different surrogate training periods in which the years have been gradually rearranged. The surrogates have been done iteratively by randomly swapping an increasing number of years, from 0 (original series) to 30 years, thus progressively destroying the accuracy of the RCM in the training period, while keeping the seasonal structure. As an illustrative example, and for sake of brevity, in Fig.7 we show the results for the ETHZ model, considering the wet test period. The accuracy of the RCM is measured as the daily temporal correlation between the RCM training surrogate data and the observations; the reliability (spatial MAE) of the RCM in the test period is shown by a dashed line in this figure whereas the reliability of the MOS downscaled values (for the different surrogate training periods) are marked with circles. Fig.7a shows the results for total precipitation ($PRCPTOT$), where the reliability of the RCM keeps a constant value of 0.23, whereas the *MOS analog* constantly improves this value up to 0.12 (for the original data). Therefore, as the accuracy of the RCM improves, the *MOS analog* allows improving the reliability of the downscaled series. Fig.7b shows the case for consecutive

dry days (CDD); in this case, since the variable is strongly related to the autocorrelation of the series, it is much more sensitive to the accuracy and an improvement of reliability is only obtained for high RCM accuracy values. This analysis gives valuable information regarding the minimum RCM accuracy needed for the *MOS analog* downscaling method in order to perform a proper calibration of the RCM, improving the reliability.

5. Summary and conclusions

In this study we introduced a new Model Output Statistics (MOS) downscaling technique based on analogs (*MOS analog*), and applied it to downscale precipitation in Spain. Our main goals were: (1) to test the skill of a MOS-like methodology for downscaling RCM simulated precipitation over a complex area and (2) to evaluate the possibility to calibrate relatively low performing RCMs using this methodology. To achieve these objectives we used the state-of-the-art ensemble of ERA40-driven RCM simulations provided by the EU-funded ENSEMBLES, as well as a gridded precipitation database developed from thousands of quality controlled stations (*Spain02*) for Spain, a region with high spatial and temporal precipitation variability. We considered ten RCMs over the common period 1961-2000 with ERA40-driven boundary conditions.

The *MOS analog* method was applied considering the RCM precipitation as the single predictor; this variable has been reported in different studies as the most informative for precipitation downscaling purposes, but it is avoided in perfect prognosis downscaling studies since it is very model dependent (e.g. different parameterizations in different models) and, thus, there may be significative differences between the reanalysis and the GCMs. This problem does not exist in the MOS setting (the RCM precipitation is used both for training and test) allowing us to define a very simple and parsimonious method.

One important limitation of the analog method is that it is not able to produce events outside those which are present in the historical archive. To test how this limitation affects our implementation, all the evaluations were carried out considering two test periods: a wet and a dry one.

One of the main advantages of the method is that it allows improving the reliability of the RCMs while preserving (or even improving, e.g. for correlation) their accuracy, regardless their own reliability. This is true for both dry and wet periods, although the performance of MOS decreases slightly in the latter case. The improvements are very good for the mean precipitation indices (e.g. the total precipitation and the intensity of precipitation) and also for the frequency (e.g. the consecutive dry days) and the extreme indices (e.g. $RX1DAY$, the maximum precipitation in one day). The ability of the method to reproduce the annual cycle of $RX1DAY$ was also tested. It has been found that this index is reproduced quite well at the basin scale by the RCMs and that the MOS improves the results of the RCMs, reducing the spread of the ensemble. In this regard, the method has more difficulties in the Mediterranean basins in autumn, which was expected, due to the importance of convective events.

Finally, the conditions under which the *MOS analog* improves the reliability of the RCM were tested by resampling the training years of the RCMs, i.e. varying the accuracy of the model. Generally, as the accuracy of the RCM improves, the *MOS analog* improves the reliability while keeping (errors, e.g. MAE) or improving (correlation) the original accuracy. Besides, being able to calibrate the RCMs, the *MOS analog* has other advantages: it maintains the spatial coherence of the precipitation fields (which is very important for hydrology); it is simple and parsimonious, so it is more robust than other complex

methods used in perfect prognosis; and it performs well in the different climates of Spain, which gives confidence in the transferability of the method to other regions.

We conclude that the two objectives of this study were fulfilled. The *MOS analog* was shown to work well in a difficult area as Spain, including in the Mediterranean side and it can be used to calibrate relatively low performing RCMs (in terms of reliability). The method performed better in the dry period, since, by construction, the method cannot reproduce unobserved weather patterns. This limitation must be seriously considered when working with future scenarios of precipitation. This limitation is common to all statistical downscaling methods [*Wilby et al.*, 2004].

In the future, we plan to test this method under sub-optimal conditions (using RCMs driven by GCM) and to apply this method to future RCM scenarios. In this case, it will be necessary to do further analysis of the applicability of our method, e.g. testing the validity of the statistical relationship in a surrogate climate as in *Frias et al.* [2006]. Finally, we intend to use the future scenarios in hydrological applications.

Acknowledgments. This work was supported by esTcena project (Exp. 200800050084078), a strategic action from Plan Nacional de I+D+i 2008-2011 funded by Spanish Ministry of Medio Ambiente y Medio Rural y Marino. For the RCM data used in this study, we acknowledge the ENSEMBLES project, funded by the European Commission's 6th Framework Programme through contract GOCE-CT-2003-505539. The authors thank AEMET and UC for the data provided for this work (*Spain02* gridded precipitation data set). Special thanks to the authors of the MeteoLab-Toolbox (www.meteo.unican.es/software/meteolab) which help us to post-process the data and to validate the method.

References

Bachner, S., A. Kapala, and C. Simmer (2008), Evaluation of daily precipitation characteristics in the CLM and their sensitivity to parameterizations, *Mteorologische Zeitschrift*, *17*(4, Sp. Iss. SI), 407–419, doi:{10.1127/0941-2948/2008/0300}.

Benestad, R., I. Hanssen-Bauer, and D. Chen (2008), *Empirical-Statistical Downscaling*, World Scientific Publishers.

Christensen, J. H., F. Boberg, O. B. Christensen, and P. Lucas-Picher (2008), On the need for bias correction of regional climate change projections of temperature and precipitation, *Geophysical Research Letters*, *35*(20), doi:{10.1029/2008GL035694}.

Christensen, O., M. Drews, J. Christensen, K. Dethloff, K. Ketelsen, I. Hebestadt, and R. A. (2008), The hirham regional climate model version 5, *Tech. Rep. Technical Report 06-17*, DMI, http://www.dmi.dk/dmi/en/print/tr06-17.pdf.

Cohen, S. (1990), Bringing the Global Warning Issue Closer to Home - The Challenge of Regional Impact Studies, *Bulletin of the American Meteorological Society*, *71*(4), 520–526.

Collins, M., B. B. B. Booth, G. R. Harris, J. M. Murphy, D. M. H. Sexton, and M. J. Webb (2006), Towards quantifying uncertainty in transient climate change, *Climate Dynamics*, *27*(2-3), 127–147, doi:{10.1007/s00382-006-0121-0}.

Cubasch, U., H. vonStorch, J. Waszkewitz, and E. Zorita (1996), Estimates of climate change in Southern Europe derived from dynamical climate model output, *Climate Research*, *7*(2), 129–149.

Déqué, M. (2007), Frequency of precipitation and temperature extremes over France in an anthropogenic scenario: Model results and statistical correction according to observed

values, *Global and Planetary Change*, *57*(1-2), 16–26, doi:10.1016/j.gloplacha.2006.11.030.

Efron, B., and R. Tibshirani (1993), *An Introduction to the Bootstrap*, Chapman & Hall.

Fowler, H. J., S. Blenkinsop, and C. Tebaldi (2007), Linking climate change modelling to impacts studies: Recent advances in downscaling techniques for hydrological modelling, *International Journal of Climatology*, *27*(12), 1547–1578, doi:{10.1002/joc.1556}, General Assembly of the European-Geosciences-Union, Vienna, AUSTRIA, APR, 2006.

Frias, M. D., E. Zorita, J. Fernandez, and C. Rodriguez-Puebla (2006), Testing statistical downscaling methods in simulated climates, *Geophysical Research Letters*, *33*(19), doi:{10.1029/2006GL027453}.

Garrote, L., F. Martin-Carrasco, F. Flores-Montoya, and A. Iglesias (2007), Linking drought indicators to policy actions in the Tagus basin drought management plan, *Water Resources Management*, *21*(5), 873–882, doi:{10.1007/s11269-006-9086-3}.

Gibergans-Baguena, J., and M. C. Llasat (2007), Improvement of the analog forecasting method by using local thermodynamic data. Application to autumn precipitation in Catalonia, *Atmospheric Research*, *86*(3-4), 173–193, doi:{10.1016/j.atmosres.2007.04.002}.

Giorgi, F., and P. Lionello (2008), Climate change projections for the Mediterranean region, *GLOBAL AND PLANETARY CHANGE*, *63*(2-3, Sp. Iss. SI), 90–104, doi:{10.1016/j.gloplacha.2007.09.005}.

Giorgi, F., and L. Mearns (1991), Approaches to the Simulation of Regional Climate Change - A review, *Reviews of Geophysics*, *29*(2), 191–216.

Giorgi, F., C. Jones, and G. Asrar (2009), Addressing climate information needs at the regional level: the CORDEX framework, *WMO Bulletin*, *58*(3), 175–183.

Gutowski, W. J., S. G. Decker, R. A. Donavon, Z. Pan, R. W. Arritt, and E. S. Takle (2003), Temporalspatial scales of observed and simulated precipitation in central u.s. climate, *Journal of Climate*, *16*(22), 3841–3847, doi:10.1175/1520-0442(2003)016⟨3841:TSOOAS⟩2.0.CO;2.

Haugen, J., and H. Haakensatd (2005), Validation of hirham version 2 with 50km and 25km resolution, *Tech. Rep. General Technical report 9*, RegClim, http://regclim.met.no/results/gtr9.pdf.

Herrera, S., L. Fita, J. Fernandez, and J. M. Gutierrez (2010a), Evaluation of the mean and extreme precipitation regimes from the ENSEMBLES regional climate multimodel simulations over Spain, *Journal of Geophysical Research-Atmospheres*, *115*, doi:{10.1029/2010JD013936}.

Herrera, S., J. M. Gutirrez, R. Ancell, M. R. Pons, and J. Frías, M. D.and Fernández (2010b), Development and analysis of a 50-year high-resolution daily gridded precipitation dataset over Spain (Spain02), *International Journal of Climatology*, doi:{10.1002/joc.2256}.

Hohenegger, C., P. Brockhaus, and C. Schaer (2008), Towards climate simulations at cloud-resolving scales, *Meteorologische Zeitschrift*, *17*(4, Sp. Iss. SI), 383–394, doi:{10.1127/0941-2948/2008/0303}.

Jacob, D. (2001), A note to the simulation of the annual and inter-annual variability of the water budget over the Baltic Sea drainage basin, *Meteorology and Atmospheric Physics*, *77*(1-4), 61–73.

548 Jaeger, E. B., I. Anders, D. Luethi, B. Rockel, C. Schaer, and S. I. Seneviratne (2008),
549 Analysis of ERA40-driven CLM simulations for Europe, *Meteorologische Zeitschrift*,
550 *17*(4, Sp. Iss. SI), 349–367, doi:{10.1127/0941-2948/2008/0301}.

551 Jansá, A. (1997), A general view about Mediterranean meteorology: cyclones and haz-
552 ardous weather, in *INM/WMO International Symposium on Cyclones and Hazardous*
553 *Weather in the Mediterranean (openning lecture)*, pp. 33–42, MMA-INM/UIB, Mallorca
554 (Spain).

555 Kjellstrm, E., F. Boberg, M. Castro, H. Christensen, G. Nikulin, and E. Sánchez (2010),
556 Daily and monthly temperature and precipitation statistics as performance indicators
557 for regional climate models, *Climate Research*, *44*, 135–150, doi:{10.3354/cr00932}.

558 Kjellstrm, E., L. Brring, and S. Gollvik (2005), A 140-year simulation of euro-
559 pean climate with the new version of the rossby centre regional atmospheric cli-
560 mate model (rca3), *Tech. Rep. Reports Meteorology and Climatology 108*, SMHI,
561 http://www.knmi.nl/knmi-library/knmipubTR/TR302.pdf.

562 Kristensen, P. (2010), *Water Resources: Quantity and Flows*, Publications Of-
563 fice of the European Union, http://www.eea.europa.eu/soer/europe/water-resources-
564 quantity-and-flows.

565 Llasat, M.-C. (2001), An objective classification of rainfall events on the basis of their con-
566 vective features: application to rainfall intensity in the northeast of spain, *International*
567 *Journal of Climatology*, *21*(11), 1385–1400, doi:10.1002/joc.692.

568 Llasat, M. C. (2009), High magnitude storms and floods, in *The Physical Geography of*
569 *the Mediterranean*, edited by J. Woodward, Oxford University Press, 513-540.

570 Lorenz, E. (1969), Atmospheric Predictability as Revealed by Naturally Occurring Ana-
571 logues, *Journal of the Atmospheric Sciences*, *26*(4), 636–&.
572 Maraun, D., F. Wetterhall, A. M. Ireson, R. E. Chandler, E. J. Kendon, M. Widmann,
573 S. Brienen, H. W. Rust, T. Sauter, M. Themessl, V. K. C. Venema, K. P. Chun,
574 C. M. Goodess, R. G. Jones, C. Onof, M. Vrac, and I. Thiele-Eich (2010), Precipitation
575 Downscaling under Climate Change: Recent Developments to Bridge the Gap Between
576 Dynamical Downscaling Models and the End User, *Reviews of Geophysics*, *48*, doi:
577 {10.1029/2009RG000314}.
578 Martin-Vide, J. (2004), Spatial distribution of a daily precipitation concentration index
579 in peninsular Spain, *INTERNATIONAL JOURNAL OF CLIMATOLOGY*, *24*(8), 959–
580 971, doi:{10.1002/joc.1030}.
581 Matulla, C., S. Schmutz, A. Melcher, T. Gerersdorfer, and P. Haas (2007), Assessing the
582 impact of a downscaled climate change simulation on the fish fauna in an Inner-Alpine
583 River, *INTERNATIONAL JOURNAL OF BIOMETEOROLOGY*, *52*(2), 127–137, doi:
584 {10.1007/s00484-007-0107-6}.
585 Murphy, A. (1993), What is a good forecast? An essay on the nature of goodness in
586 weather forecasting, *Weather Forecasting*, *8*, 281–293.
587 Obled, C., G. Bontron, and R. Garcon (2002), Quantitative precipitation forecasts: a
588 statistical adaptation of model outputs through an analogues sorting approach, *Atmo-*
589 *spheric Research*, *63*(3-4), 303–324.
590 Pal, J. S., F. Giorgi, X. Bi, N. Elguindi, F. Solmon, X. Gao, S. A. Rauscher, R. Francisco,
591 A. Zakey, J. Winter, M. Ashfaq, F. S. Syed, J. L. Bell, N. S. Diffenbaugh, J. Kar-
592 macharya, A. Konare, D. Martinez, R. P. da Rocha, L. C. Sloan, and A. L. Steiner

(2007), Regional climate modeling for the developing world - The ICTP RegCM3 and RegCNET, *Bulletin of the American Meteorological Society*, *88*(9), 1395+, doi:{10.1175/BAMS-88-9-1395}.

Piani, C., J. Haerter, and E. Coppola (2010), Statistical bias correction for daily precipitation in regional climate models over Europe, *Theoretical and Applied Climatology*, *99*(1-2), 187–192, doi:{10.1007/s00704-009-0134-9}.

Quintana-Seguí, P., A. Ribes, E. Martin, F. Habets, and J. Boe (2010), Comparison of three downscaling methods in simulating the impact of climate change on the hydrology of Mediterranean basins, *Journal of Hydrology*, *383*(1-2), 111–124, doi:{10.1016/j.jhydrol.2009.09.050}.

Radu, R., M. Deque, and S. Somot (2008), Spectral nudging in a spectral regional climate model, *Tellus Series A - Dynamic Meteorology and Oceanography*, *60*(5), 898–910, doi:{10.1111/j.1600-0870.2008.00341.x}.

Río, S., L. Herrero, R. Fraile, and A. Penas (2010), Spatial distribution of recent rainfall trends in Spain (1961-2006), *International Journal of Climatology*, doi:{10.1002/joc.2111}.

Rodrigo, F. S., and R. M. Trigo (2007), Trends in daily rainfall in the Iberian Peninsula from 1951 to 2002, *INTERNATIONAL JOURNAL OF CLIMATOLOGY*, *27*(4), 513–529, doi:{10.1002/joc.1409}.

Rodriguez-Puebla, C., A. Encinas, S. Nieto, and J. Garmendia (1998), Spatial and temporal patterns of annual precipitation variability over the Iberian Peninsula, *INTERNATIONAL JOURNAL OF CLIMATOLOGY*, *18*(3), 299–316.

615 Romero, R., J. Guijarro, C. Ramis, and S. Alonso (1998), A 30-year (1964-1993) daily
616 rainfall data base for the Spanish Mediterranean regions: First exploratory study, *In-*
617 *ternational Journal of Climatology*, *18*(5), 541–560.

618 Sanchez, E., C. Gallardo, M. Gaertner, A. Arribas, and M. Castro (2004), Future climate
619 extreme events in the Mediterranean simulated by a regional climate model: a first
620 approach, *Global and Planetary Change*, *44*(1-4), 163–180, doi:{10.1016/j.gloplacha.
621 2004.06.010}.

622 Serrano, A., J. Garcia, V. Mateos, M. Cancillo, and J. Garrido (1999), Monthly modes of
623 variation of precipitation over the Iberian peninsula, *Journal of Climate*, *12*(9), 2894–
624 2919.

625 Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K. Averyt, M. Tignor, and
626 H. Miller (2007), *Climate Change 2007: The Physical Science Basis. Contribution of*
627 *Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on*
628 *Climate Change*, Cambridge University Press.

629 Timbal, B., A. Dufour, and B. McAvaney (2003), An estimate of future climate change for
630 western France using a statistical downscaling technique, *Climate Dynamics*, *20*(7-8),
631 807–823, doi:{10.1007/s00382-002-0298-9}.

632 Trenberth, K., A. Dai, R. Rasmussen, and D. Parsons (2003), The changing character
633 of precipitation, *Bulletin of the American Meteorological Society*, *84*(9), 1205+, doi:
634 {10.1175/BAMS-84-9-1205}.

635 Trigo, R., and J. Palutikof (2001), Precipitation scenarios over Iberia: A comparison
636 between direct GCM output and different downscaling techniques, *Journal of Climate*,
637 *14*(23), 4422–4446.

Uppala, S., P. Kallberg, A. Simmons, U. Andrae, V. Bechtold, M. Fiorino, J. Gibson, J. Haseler, A. Hernandez, G. Kelly, X. Li, K. Onogi, S. Saarinen, N. Sokka, R. Allan, E. Andersson, K. Arpe, M. Balmaseda, A. Beljaars, L. Van De Berg, J. Bidlot, N. Bormann, S. Caires, F. Chevallier, A. Dethof, M. Dragosavac, M. Fisher, M. Fuentes, S. Hagemann, E. Holm, B. Hoskins, L. Isaksen, P. Janssen, R. Jenne, A. McNally, J. Mahfouf, J. Morcrette, N. Rayner, R. Saunders, P. Simon, A. Sterl, K. Trenberth, A. Untch, D. Vasiljevic, P. Viterbo, and J. Woollen (2005), The ERA-40 re-analysis, *Quarterly Journal of the Royal Meteorological Society*, *131*(612, Part B), 2961–3012, doi:{10.1256/qj.04.176}.

van der Linden, P., and J. Mitchell (Eds.) (2009), *ENSEMBLES: Climate Change and its Impacts: Summary of research and results from the ENSEMBLES project*, 160pp pp., Met Office Hadley Centre, FitzRoy Road, Exeter EX1 3PB, UK.

Van Meijgaard, E., L. van Ulft, W. van de Berg, B. Bosveld, B. van der Hurk, G. Lenderik, and A. Siebesma (2008), The knmi regional atmospheric climate model racmo version 2.1., *Tech. Rep. 302*, KNMI, http://www.knmi.nl/knmi-library/knmipubTR/TR302.pdf.

Widmann, M., C. Bretherton, and E. Salathe (2003), Statistical precipitation downscaling over the Northwestern United States using numerically simulated precipitation as a predictor, *Journal of Climate*, *16*(5), 799–816.

Wilby, R., S. Charles, E. Zorita, and B. Timbal (2004), Guidelines for use of climate scenarios developed from statistical downscaling methods, *Tech. rep.*, IPCC.

Wilby, R. L., and T. M. L. Wigley (1997), Downscaling general circulation model output: a review of methods and limitations, *Progress in Physical Geography*, *21(4)*, 530–548.

Wilks, D. S., and R. L. Wilby (1999), The weather generator game: A review of stochastic weather models, *Progress in Physical Geography*, *23*, 329–358.

WMO (2009), Guidelines on analysis of extremes in a changing climate in support of informed decisions for adaptation, *Tech. Rep. WCDMP No. 72 WMO/TD-No. 1500*, WMO.

Zorita, E., and H. von Storch (1999), The analog method as a simple statistical downscaling technique: Comparison with more complicated methods, *Journal of Climate*, *12*(8, Part 2), 2474–2489.

Zorita, E., J. Hughes, D. Lettemaier, and H. von Storch (1995), Stochastic Characterization of Regional Circulation Patterns for Climate Model Diagnosis and Estimation of Local Precipitation, *Journal of Climate*, *8*(5, Part 1), 1023–1042.

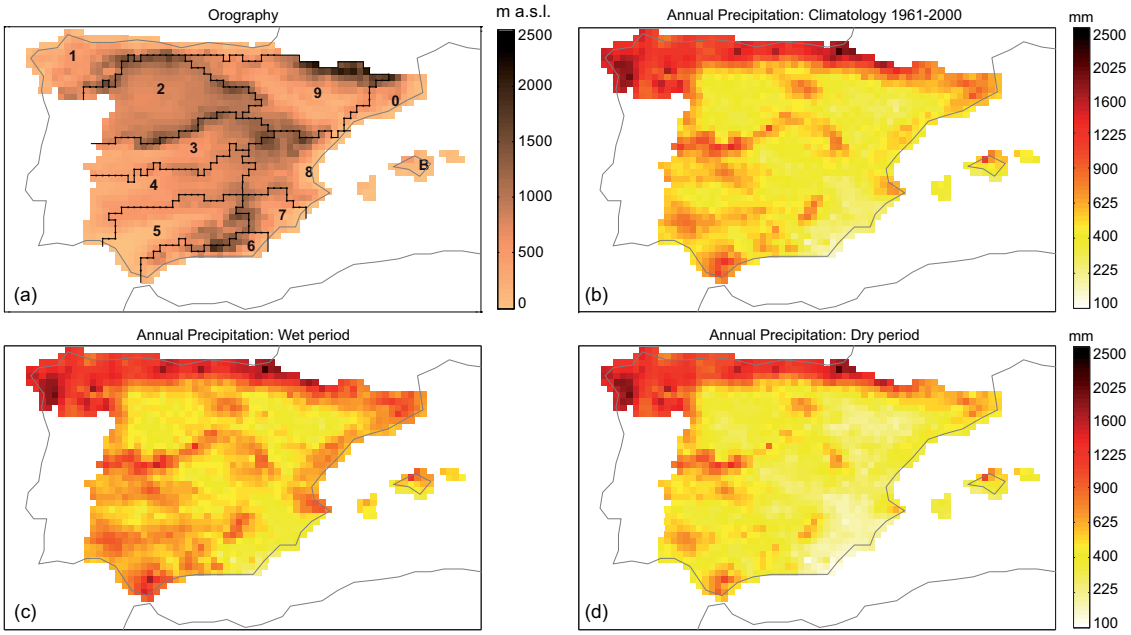

Figure 1. (a) Topography of Spanish Iberian Peninsula and the Balearic Islands as represented by *Spain02* at 0.2° x 0.2°, showing the main river basins: 0. Catalana, 1. North, 2. Duero, 3. Tajo, 4. Guadiana, 5. Guadalquivir, 6. South, 7. Segura, 8. Levante, 9. Ebro, B. Baleares. (b) Annual precipitation from *Spain02* (mm) in the period 1961-2000. (c) Annual precipitation of *Spain02* (mm) in the wettest years and (d) in the driest years; see Sec. 4 for the definition of the wettest and driest years.

Table 1. Summary of the RCM simulations nested in ERA40 data produced for the ENSEMBLES project. The columns are the acronym used in the paper, the institution running the simulation, the model used and a reference publication. The asterisks indicate the best performing models in this region according to *Herrera et al.* [2010a].

Acronym	**Institution**	**Model**	**Reference**
CNRM	Centre National de Recherches Meteorologiques	ALADIN-Climat	*Radu et al.* [2008]
DMI	Danish Meteorological Institute	HIRHAM	*Christensen et al.* [2008]
ETHZ(*)	Swiss Institute of Technology	CLM	*Jaeger et al.* [2008]
KNMI(*)	Koninklijk Nederlands Meteorologisch Instituut	RACMO	*Van Meijgaard et al.* [2008]
HC(*)	Hadley Center/UK MetOffice	HadRM3 Q0	*Collins et al.* [2006]
ICTP	Abdus Salam International Centre for Theoretical Physics	RegCM3	*Pal et al.* [2007]
METNO	The Norwegian Meteorological Institute	HIRHAM	*Haugen and Haakensatd* [2005]
MPI(*)	Max Planck Institute for Meteorology	M-REMO	*Jacob* [2001]
SMHI	Swedish Meteorological and Hydrological Institute	RCA	*Kjellstrm et al.* [2005]
UCLM(*)	Universidad de Castilla la Mancha	PROMES	*Sanchez et al.* [2004]

Period	Years
Wettest	1996, 1969, 1997, 1979, 1963, 1972, 1977, 1989, 1971 and 1987
Driest	1964, 1998, 1994, 1990, 1970, 1967, 1983, 1973, 1980 and 1981

Table 2. The ten wettest and the ten driest years in Spain within the period 1961-2000. The years have been obtained by ranking the grid-point standardized spatially averaged precipitation.

Table 3. Climatic mean and extreme indices for precipitation used in this work (see also ETCCDI http://cccma.seos.uvic.ca/ETCCDI).

label	description	units
PRCPTOT	total precipitation	mm
SDII	Mean precipitation amount on a wet day	mm
R10	number of days with precipitation over 10 mm/day	day
R20	number of days with precipitation over 20 mm/day	day
RX1DAY	maximum precipitation in 1 day	mm
RX5DAY	maximum precipitation in 5 days	mm
CDD	consecutive dry days ($<$ $1mm$)	day
CWD	consecutive wet days ($>$ $1mm$)	day

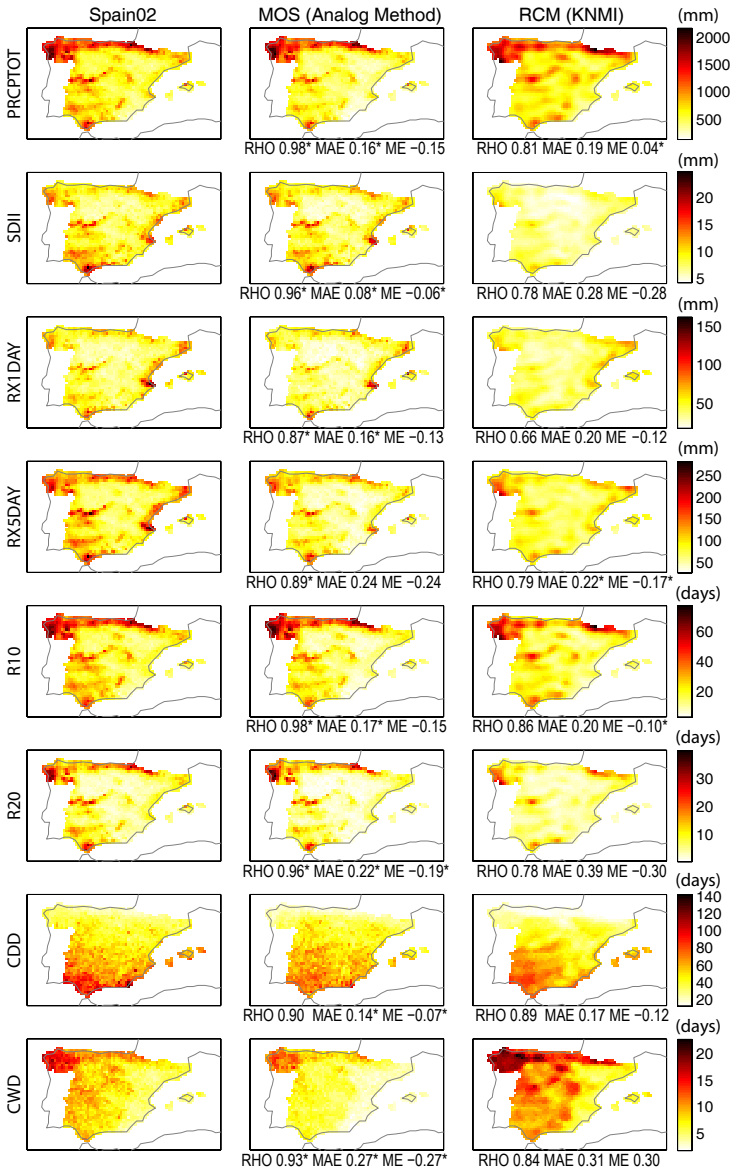

Figure 2. Spatial distribution of the observed (left), downscaled (central) and RCM (right) mean values (averaged over the wet validation period) for the precipitation reliability indices shown in Table 3. The spatial validation scores (correlation and errors) for the *MOS analog* and RCM simulated values, w.r.t. the observed values, are given below the corresponding panels. The asterisks next to the MOS (or RCM) scores indicate those situations where the score is significantly better (larger for correlation and smaller for errors) than the one corresponding to the RCM (or MOS), respectively.

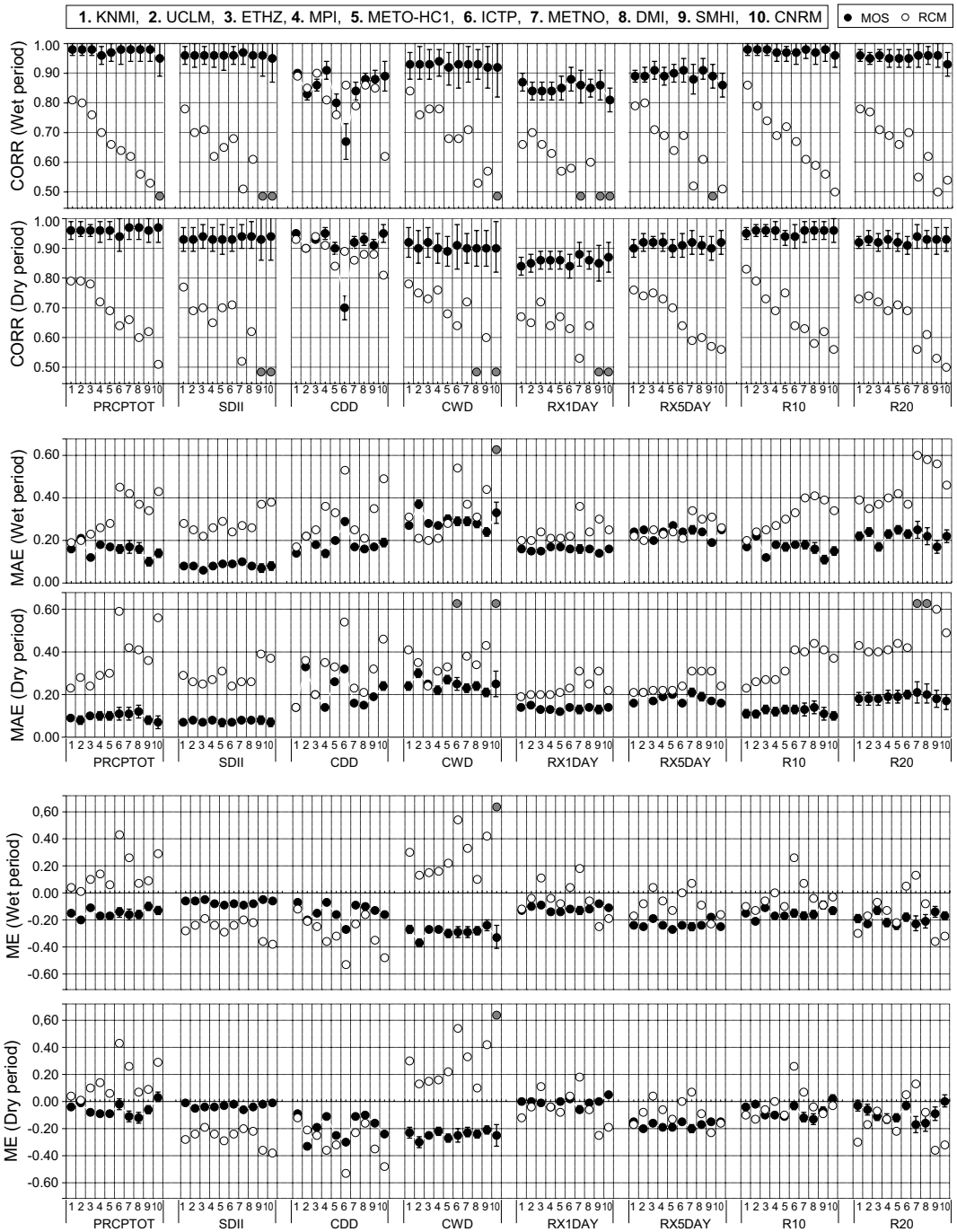

Figure 3. Summary of validation results for correlation ($CORR$), mean error (ME) and mean absolute error (MAE) for the different indices and validation (wet and dry) periods. Open circles represent the RCM values and the filled circles downscaled values. See the running text for more details.

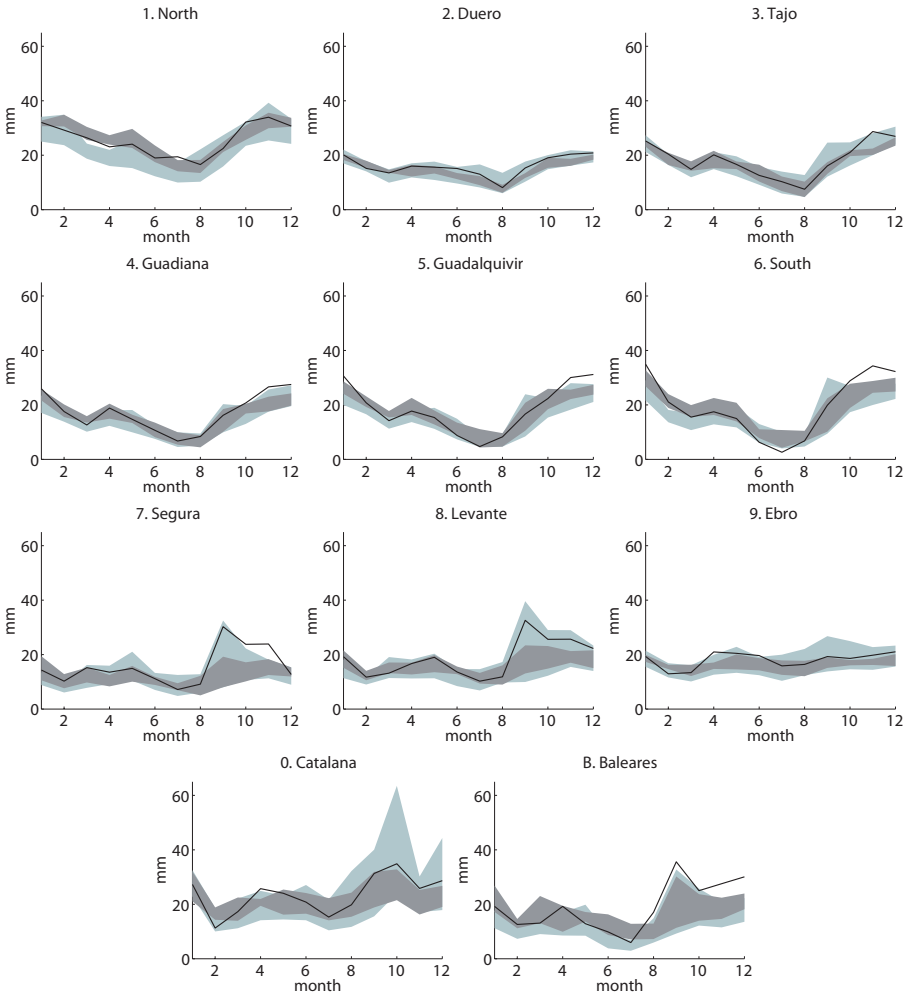

Figure 4. Seasonal cycle (yearly averaged monthly values) of the spatially averaged $RX1DAY$ index (in mm) for each river basin (according to Figure 1a). The black line represents the observed (*Spain02*) climatology. The light shaded band spans the values for the RCMs while the dark one spans the respective MOS downscaled values. The results correspond to the wet test period.

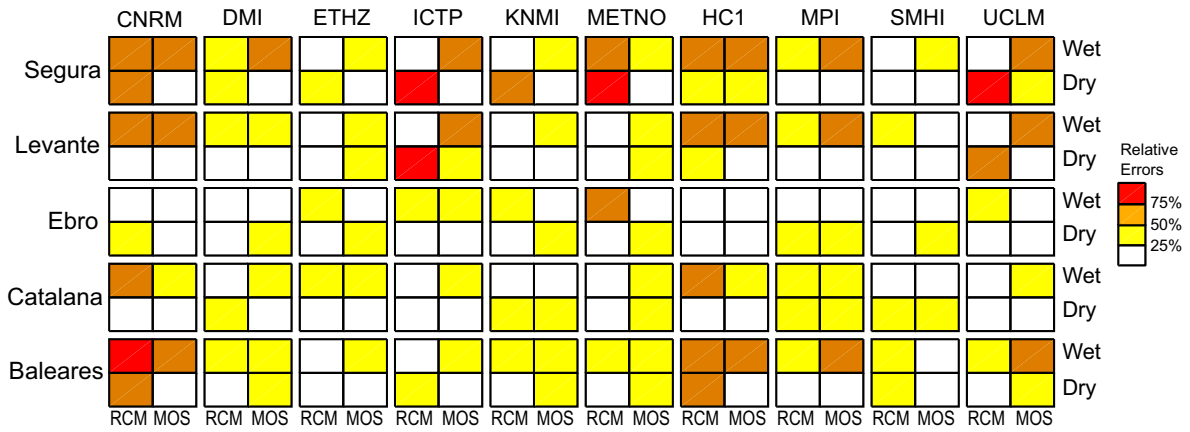

Figure 5. Differences among simulated (MOS as well the respective RCM) and observed $RX1DAY$ value for the month of September, relative to the observed values. The results are given for each Mediterranean river basin (according to Figure 1a) and test period (wet and dry). Red colors represent errors greater than 75%, orange between 50% and 75%, yellow between 25% and 50%, and white errors less than 25%.

		CORR		MAEr	
		MOS	RCM	MOS	RCM
CNRM	DRY	0.70 (0.61/0.75)	0.48 (0.45/0.52)	1.49 (1.38/1.68)	1.84 (1.43/2.25)
	WET	0.63 (0.55/0.69)	0.45 (0.40/0.50)	1.34 (1.21/1.46)	1.63 (1.30/1.99)
DMI	DRY	0.76 (0.70/0.80)	0.58 (0.51/0.60)	1.24 (1.14/1.37)	1.23 (1.00/1.50)
	WET	0.69 (0.64/0.74)	0.59 (0.54/0.63)	1.16 (1.07/1.31)	1.17 (0.95/1.42)
ETHZ	DRY	0.76 (0.70/0.80)	0.63 (0.59/0.69)	1.28 (1.18/1.40)	1.27 (1.05/1.53)
	WET	0.69 (0.64/0.73)	0.62 (0.58/0.69)	1.20 (1.12/1.34)	1.22 (1.02/1.49)
ICTP	DRY	0.74 (0.68/0.78)	0.58 (0.53/0.64)	1.34 (1.24/1.47)	1.70 (1.33/2.16)
	WET	0.67 (0.61/0.71)	0.57 (0.51/0.64)	1.22 (1.14/1.36)	1.48 (1.18/1.88)
KNMI	DRY	0.76 (0.70/0.80)	0.59 (0.56/0.65)	1.29 (1.17/1.43)	1.27 (1.01/1.48)
	WET	0.70 (0.64/0.75)	0.60 (0.55/0.67)	1.16 (1.06/1.30)	1.14 (0.93/1.35)
METNO	DRY	0.76 (0.71/0.80)	0.62 (0.58/0.67)	1.23 (1.12/1.39)	1.33 (1.08/1.60)
	WET	0.69 (0.63/0.74)	0.61 (0.56/0.66)	1.18 (1.07/1.32)	1.28 (1.05/1.53)
METO-HC1	DRY	0.73 (0.66/0.77)	0.56 (0.53/0.60)	1.36 (1.25/1.53)	1.35 (1.12/1.57)
	WET	0.66 (0.60/0.71)	0.55 (0.51/0.60)	1.24 (1.15/1.39)	1.28 (1.06/1.51)
MPI	DRY	0.76 (0.71/0.80)	0.59 (0.55/0.61)	1.24 (1.14/1.38)	1.35 (1.10/1.56)
	WET	0.70 (0.65/0.75)	0.59 (0.54/0.63)	1.15 (1.06/1.28)	1.27 (1.03/1.48)
SMHI	DRY	0.76 (0.71/0.81)	0.61 (0.56/0.67)	1.23 (1.13/1.39)	1.27 (0.97/1.53)
	WET	0.70 (0.65/0.75)	0.62 (0.55/0.67)	1.16 (1.07/1.32)	1.18 (0.83/1.45)
UCLM	DRY	0.68 (0.61/0.73)	0.54 (0.49/0.60)	1.53 (1.43/1.63)	1.52 (1.30/1.76)
	WET	0.63 (0.56/0.68)	0.52 (0.47/0.57)	1.30 (1.23/1.40)	1.34 (1.15/1.58)

Table 4. Accuracy scores for the MOS and RCM methods. Each cell shows the median and IQR (in parenthesis) of the spatial distribution of the Correlation (CORR) and the Mean Absolute Error weighted by the observed mean (MAEr), for the RCMs and the respective MOS, for the wet and dry test period.

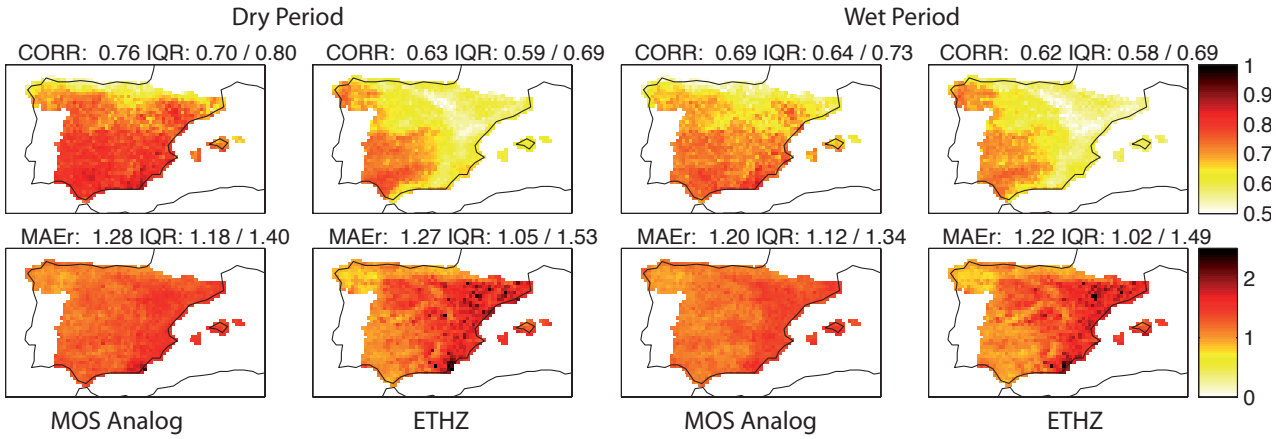

Figure 6. Spearman correlation ($CORR$) and Mean Absolute error relative to the mean precipitation ($MAEr$) comparing the ETHZ model with the *MOS analog* for the dry test period (left) and wet test period (right). The values on the top of each map are the median and the interquartile range (IQR) of the spatial distribution of the corresponding score.

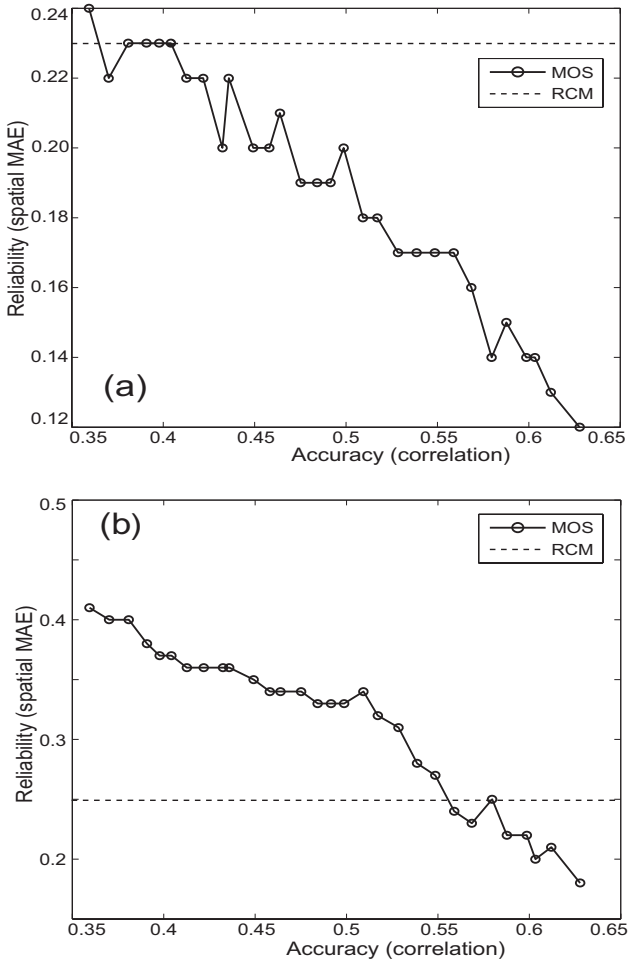

Figure 7. Spatial error (MAE) of the precipitation indices (a) PRCPTOT and (b) CDD for the *MOS analog* method as a function of the accuracy of the RCM, measured as the daily temporal correlation of the RCM training surrogate data and the observations. The dashed lines indicate the reliability (MAE) of the RCM; see running text for more details.