**Hydrology and
Earth System
Sciences**

Downscaling of surface moisture flux and precipitation in the Ebro Valley (Spain) using analogues and analogues followed by random forests and multiple linear regression

G. Ibarra-Berastegi[1], J. Saénz[2], A. Ezcurra[2], A. Elías[3], J. Diaz Argandoña[4], and I. Errasti[1]

[1]Dept. of Fluid Mechanics and Nuclear Engineering, University of the Basque Country, Spain
[2]Dept. of Applied Physics II, University of the Basque Country, Spain
[3]Dept. of Chemical and Environmental Engineering, University of the Basque Country, Spain
[4]Dept. of Applied Physics I, University of the Basque Country, Spain

Abstract. In this paper, reanalysis fields from the ECMWF have been statistically downscaled to predict from large-scale atmospheric fields, surface moisture flux and daily precipitation at two observatories (Zaragoza and Tortosa, Ebro Valley, Spain) during the 1961–2001 period. Three types of downscaling models have been built: (i) analogues, (ii) analogues followed by random forests and (iii) analogues followed by multiple linear regression. The inputs consist of data (predictor fields) taken from the ERA-40 reanalysis. The predicted fields are precipitation and surface moisture flux as measured at the two observatories. With the aim to reduce the dimensionality of the problem, the ERA-40 fields have been decomposed using empirical orthogonal functions. Available daily data has been divided into two parts: a training period used to find a group of about 300 analogues to build the downscaling model (1961–1996) and a test period (1997–2001), where models' performance has been assessed using independent data. In the case of surface moisture flux, the models based on analogues followed by random forests do not clearly outperform those built on analogues plus multiple linear regression, while simple averages calculated from the nearest analogues found in the training period, yielded only slightly worse results. In the case of precipitation, the three types of model performed equally. These results suggest that most of the models' downscaling capabilities can be attributed to the analogues-calculation stage.

1 Introduction

Global Climate Models (GCM) and Numerical Weather Forecast (NWF) models solve discretized versions of the primitive equations that govern the evolution of the climate system and, particularly, its atmospheric component. Due to technical limitations of current supercomputers, GCM are currently run using resolutions ranging from about 4 to 1.5 degrees for climate simulations and most recent NWF operate with a spatial resolution as high as T1279 L91 (at ECMWF), with a 16 km resolution globally (0.1 degree).

Statistical downscaling is one of the approaches that have been developed and used to regionalize the outputs from GCM and NWF i.e. to construct region- or site-specific scenarios. The rationale for this is based on the fact that the GCM and NWF are not able to simulate surface variables with enough accuracy on regional and local scales. Statistical downscaling consists in a search in the observed data for a statistical relationship or transfer function between the surface climate variable to be downscaled (predictand) and potential predictors, which are frequently the large-scale upper air variables. Statistical downscaling is very often used in the development of regional climate change, scenarios but it has also been used for NWF, particularly for short time ranges with (Fernandez-Ferrero et al., 2009, 2010; Hamill and Whitaker, 2006; Voisin et al., 2010 and references therein).

The analogue downscaling approach tries to identify from historical records, synoptic conditions that are similar to the current atmospheric state as described by GCM and NWF

models. The idea is that historical records at a given location obtained under past similar atmospheric conditions, can now be expected to be a good estimate for current observations at the same place. The Euclidean distance is usually employed (Matulla et al., 2008) to compare current and past synoptic conditions and then identify those which are close (similar) to the current state of the atmosphere. In order to reduce the dimensionality of the phase space and to help in the finding of good analogues, empirical orthogonal functions (EOF) are usually calculated from raw synoptic data. It is recommended to retain a number of EOF representing 80 %–90 % of the overall variability (Matulla et al., 2008).

The analogue technique has often been used as a downscaling method (Zorita and von Storch, 1999; Fernández and Sáenz, 2003; Timbal and Jones, 2008) and it performs similarly when compared with Canonical Correlation Analysis (CCA) while outperforming nonlinear approaches like Classification and Regression Trees (CART) and neural networks (NN) (Zorita and von Storch, 1999).

For downscaling purposes, several regression techniques – linear or not – have been also widely used. In this line, for surface temperature, a classical CCA seems to perform better than other linear methods (Huth, 2002). Among nonlinear methods is worth mentioning the use of different types of NN which are reported to outperform linear models (Weichert and Bürger, 1998; Miksovsky and Raidl, 2005; Davy et al., 2010). Other works suggest that for statistical downscaling of temperature, linear models cannot be beaten by NN (Huth et al., 2008).

A combination of analogues with other techniques like CCA, NN or a Bayesian model for precipitation downscaling but at 6-hourly level and in forecasting mode has also yielded good results (Fernández-Ferrero et al., 2009, 2010). In general, several studies exist which as a first step, have used an initial search of analogues followed by some calibration step (bayesian or not) (Benestad, 2008, 2010; Hamill and Whitaker, 2006; Voisin et al., 2010). The use of bayesian techniques for calibration purposes of ensemble forecasts is the focus of other studies focused on multivariate gaussian data (Stephenson et al., 2005). It is well known that analogue-based statistical downscaling models cannot produce non-observed record-breaking output values and some calibration techniques have been proposed to overcome this problem (Benestad, 2010).

In the last decade, a great number of machine learning algorithms have been developed (Witten and Frank, 2005). Among them, Random Forests (RF) has become increasingly popular for several reasons, being the most important one its ability to model nonlinear relationships.

When compared with other techniques also intended to deal with nonlinear regression, in the case of similar downscaling problems the literature shows that RF and a type of NN, the multilayer perceptron, perform similarly (Eccel et al., 2007).

RF is based on the CART technique, in which a tree is built to relate a set of inputs to a group of output variables. The relationship between inputs and outputs may be highly nonlinear if taken as a whole, but if analyzed at different ranges, the nature of the relationships inside each range may be modelled in a simpler way. For this reason, in a CART the feature space is recursively splitted into a set of regions in which a simple model like a constant (Hastie et al., 2001) or a simple linear regression can be fitted. As the splitting process progresses starting from the root, the tree is divided at subsequent nodes into branches and sub-branches until reaching the leafs (a more homogeneous region where a simple regression between inputs and outputs is more evident). The final stage implies going backwards to the tree obtained and prune it using cost-complexity pruning algorithms. (Hastie et al., 2001).

Random forests are built using CART but adding two layers of randomness (Liaw and Wiener, 2002):

1. If a number of bootstrap samples from original data are drawn, a regression tree can be obtained with each sample, thus obtaining a "forest".

2. In standard trees, each node is split using the best split among all variables. In a random forest, for each tree, each node is split using the best among a subset of m predictors randomly chosen at that node.

Each tree is grown to the largest extent possible. There is no pruning. The most relevant parameters in running the RF algorithm are (i) m the number of predictors made available to each node and (ii) the number of trees grown Being p the number of candidate predictors, values of m like $p/3$ or $p^{0.5}$ are usually adopted. Nevertheless, results do not strongly rely on the choice of m (Breiman, 2001). The number of trees necessary for good performance grows with the number of predictors. The best way to determine how many trees are necessary is to compare predictions made by a forest to predictions made by a subset of a forest. When the error obtained with the subset is the same as the error obtained with the full forest, this means that the forest has got enough trees (Breiman, 2001).

RF can be used for classification and regression purposes. In the case of regression, the output of the RF model will simply be the average of the outputs obtained in each tree. RF are reported to clearly outperform CART (Siroky, 2009), being one of RF's most important advantages when compared with other techniques, that overfitting never takes place (Siroky, 2009). The relative importance of each predictor on the predictand is estimated at each tree by calculating the increase in the mean square error due to permuting a given predictor. If that regressor has no predictive value for the response, it should not make a difference if its values are randomly permuted before the predictions are generated (Grömping, 2009). Then, differences due to these permutations in all the

trees are averaged and normalized by the standard error. This is known as mean decrease in accuracy of the predictor.

A thorough revision regarding the major mathematical aspects of RF, can be found in the literature (Breiman et al., 1984; Breiman, 2001, 2002; Hastie et al., 2001; Liaw and Wiener, 2002; Ishwaran, 2007; Siroky, 2009; Grömping, 2009).

In the case of the Iberian Peninsula (IP), precipitation shows an important spatial variability with typical yearly values ranging from 300 to 3000 mm depending on the location. The variability of precipitation is associated with several atmospheric patterns, depending on the location or the season of the year: the most important patterns influencing Iberian precipitation are East Atlantic (EA), North Atlantic Oscillation (NAO); Southern Oscillation Index (SOI) and Scandinavia (SCAND) (Rodriguez-Puebla et al., 1998; Fernández et al., 2003).

Although in precipitation downscaling there is little consensus as to the choice of atmospheric predictor variables (Wilby and Wigley, 2000), the most common input variables have been 850 and 500-hPa heights. In other works, additional variables like 500-hPa temperature, 700 hPa levels, 500-hPa relative humidity, 500-hPa wind components, 1000–500-hPa thickness and 850-hPa temperature fields to mention a few, have also been used (Cavazos, 2000; Wilby and Wigley, 2000; Schoof and Pryor, 2001). In more recent works, the predictors for the downscaling procedure include mean sea level pressure, geostrophic wind field, specific humidity and moisture flux (Wei et al., 2010) or total precipitable water (Timbal and Jones, 2008). A thorough analysis of different combinations of predictors applied to short range quantitative precipitation forecast was presented by Fernandez-Ferrero et al. (2009).

In the case of precipitation, using highly nonlinear techniques like radial basis function (RBF) NN instead of more simple linear models, only represents a marginal improvement (Weichert and Bürger, 1998). Other works also suggest that in the case of precipitation, a linear model of downscaling can perform equally or even better than highly nonlinear NN (Trigo et al., 2001). A comparison between NN and linear downscaling for daily temperature and precipitation showed that NN performed better, although the results with precipitation have been significantly poorer (Schoof and Pryor, 2001). As reported in the literature, performance tends to be better in winter (Timbal and Jones, 2008; Wei et al., 2010) and mid latitudes (Cavazos and Hewitson, 2005). Typical correlation coefficients between predictions and observations may reach values around 0.7 for downscaling of monthly averages and about 0.5 for daily precipitation.

The technique of analogues have been used for precipitation downscaling in the IP with similar results and are reported to outperform NN (Zorita and Von Storch, 1999). Analogues followed by a bias-correcting heuristic formula has also been used for precipitation downscaling (Timbal and Jones, 2008). Generally speaking, the combination

of classification schemes like Cluster Analysis (Schoof and Pryor, 2001), or self-organizing maps (Cavazos and Hewitson, 2005) followed by regression techniques (linear or not) seem to yield the best results for precipitation downscaling. A comprehensive review of methods and predictors used can be found in the literature (Wilby and Wigley, 2000).

At a hemispheric scale, changes in the vertically integrated moisture transport are connected to the different phases of the NAO (Hurrell, 1995). Regarding the moisture transport on the Iberian Peninsule (IP), the NAO and EA patterns are known to play a key role (Fernandez et al., 2003). The two most important moisture source regions affecting the IP are in a tropical–subtropical North Atlantic corridor that extends from the Gulf of Mexico to the IP, and the IP itself and the surrounding Mediterranean (Gimeno et al., 2010). The importance of these two source areas varies throughout the year, and also with respect to different climatic regions inside the IP.

Some works have also focused on the downscaling of different moisture variables. A set of downscaling models built on multiple linear regression showed that in the case of humidity variables, the most efficient predictors are the low-medium tropospheric air humidity variables (up to 500 hPa) and adding circulation and/or temperature variables to the predictors could only bring a marginal or even no improvement over the downscaling models based on the humidity variables only (Huth, 2005). The results show that in the case of water vapour pressure and dew point temperature, results are better than those obtained for relative humidity and dew point depression (Huth, 2005). Like in the case of precipitation, a comparison of water vapour downscaling methods using linear regression and RBF NN, showed that a nonlinear model like RBF outperfoms linear models, although not dramatically (Wilby and Wigley, 2000).

In this study, a combination of analogues, RF and linear regression techniques are applied, analyzed and compared for downscaling purposes in the Ebro Valley (Spain) (Fig. 1). The main objective of the study is to compare the performance of different techniques for the downscaling of moisture-related variables. All of the techniques share a common core, the search of analogues in the space spanned by the EOF of the predictor fields, but they differ in the way the output from the analogue phase is postprocessed.

The target variables are the moisture flux and precipitation. Changes in the variables involved in the water cycle in the area of interest (Ebro Valley) like moisture flux and precipitation, are likely to have an impact on future changes regarding the overall water cycle in the Mediterranean Sea. Gaining a better knowledge on downscaling techniques applied to the variables involved in the water cycle, may contribute to a more accurate regionalization of future projections as described by large-scale climatic models (Mariotti, 2010).

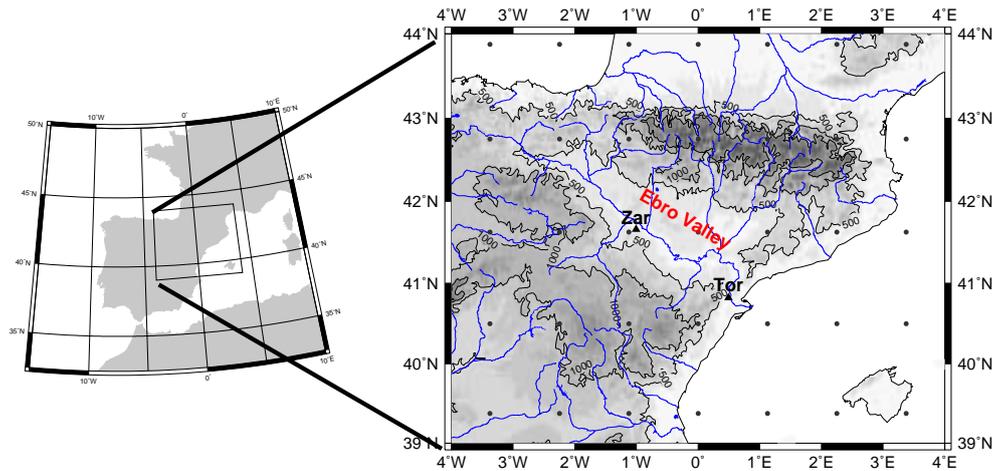Precipitation is one of the most commonly used variables for climate change studies.The reason is that the evolution

Fig. 1. Area of study. ERA-40 and ERA-Interim 35 gridpoints, Ebro Valley, Zaragoza (Zar) and Tortosa (Tor).

in the amount of precipitation over a country is crucial for a whole set of natural and economic processes.

In the case of the surface moisture flux (in this paper, denoted as the vector q_{10} with zonal and meridional components q_x and q_y), it is a variable closely related with the hydrological cycle over any area. Most of the humidity transport takes place through low-level atmospheric layers and is affected by several surface characteristics such as topography, land surface and others that vary in the smallest scales. In general, a complete study of the moisture cycle over the region would require the analysis of vertically integrated moisture transports and water column, which are the quantities related through conservation laws with moisture sinks and sources such as precipitation or evaporation (Berbery and Rasmusson, 1999).

Unfortunately, historical data from rawinsonde observations are quite scarce, and the study cannot be currently easily extended to vertically integrated moisture transports in order to close the atmospheric hydrological cycle through the whole troposphere. Thus, for this study the moisture fluxes at the surface are considered as an initial problem before further studies allow an analysis of tropospherically integrated transports. Additionally, moisture flux at the low levels of the troposphere is a variable obtained from the product of two original variables forecast by a NWF or a GCM model. It is interesting to check the ability of the downscaling model to forecast it, since there are other variables, such as the wind speed that involve multiplication of original variables and are of special interest, for instance in wind energy.

The previous experiences gained by the research group in the combination of analogues and linear/nonlinear techniques (Fernández and Sáenz, 2003; Fernández-Ferrero et al., 2009, 2010) has inspired the methodology used in this paper. In this work, analogues, linear regression and random forests have been used to build statistical downscaling models for surface moisture flux and precipitation at two observatories (Zaragoza and Tortosa, Ebro Valley, Spain)

corresponding to the period 1961–2001. The models have been built at a daily time scale.

Section 2 presents the data used in the study and the methodology applied to perform and evaluate the forecasts. Results are presented in Sect. 3. Finally, conclusions and prospects for future work are presented in Sect. 4.

2   Data and methodology

2.1   Data

For q_{10} and precipitation downscaling purposes, the large-scale observed predictors have been derived from the European Centre Medium-Range Weather Forecast (ECMWF), corresponding to the ERA-40 (Uppala et al., 2005) and ERA-Interim reanalysis, downloaded from ECMWF's MARS server with a $1.125° \times 1.125°$ resolution. ERA-40 and ERA-Interim reanalysis are projected onto the same grid. The studied area is a rectangle of 35 gridpoints with latitudes in the range (39.375° N, 43.875° N) and longitudes between $-3.375°$ E and 3.375° E, that is, the North-Eastern part of the Iberian Peninsule, and more specifically, the Ebro valley (Fig. 1). In order to check the sensitivity of the results to the size of the domain, two additional domains with 90 and 9 gridpoints have been used, and additional information on them can be found in the discussion paper, whilst results from these domains will only briefly presented here. Considering the spatial resolution of the original dataset, the daily frequency (that is the large sample used) and the high number of vertical levels used, larger domains could not be used due to memory constraints in the available computers. However, previous results from the authors suggest (Fernandez and Saenz, 2003; Figs. 13 and 14; and Table 1) that very big domains do not imply a real advantage when using analogues in the EOF space. The reason is that the EOF select the directions of the phase space with the maximum variability, which, for very large domains, can happen in areas very

Table 1. The 79 leading EOF for ERA-40.

| Variable | Number of EOF retained | Percentage of overall variability in: | | | |
		All retained EOF	1st EOF	2nd EOF	3rd EOF
Z	5	97.6	77.7	15.6	2
T	5	93.9	84.2	4.4	2.8
U	7	80.6	55.9	9.2	5.3
V	7	83.4	59.9	10.3	5
RH	28	80	27.3	10	5.4
MSL	5	98.8	89.6	4.5	2.5
$U10$	7	80.5	49.9	10.4	6.3
$V10$	5	80	56.4	9.7	7.2
$T2$	5	97.6	91.2	2.9	1.7
$D2$	5	94.6	86.4	3.2	2.2

Table 2. The 79 leading EOF for ERA-Interim.

| Variable | Number of EOF retained | Percentage of overall variability in: | | | |
		All retained EOF	1st EOF	2nd EOF	3rd EOF
Z	5	97.6	74.7	18.2	2.2
T	5	93.7	83.8	4.4	2.9
U	7	84.5	58.2	9.7	5.5
V	7	85.8	61.2	11.2	5.3
RH	28	78.3	25.6	10.6	5.5
MSL	5	98.9	90	4.3	2.5
$U10$	7	86.5	56.1	13.7	6.2
$V10$	5	83.3	58.7	10.4	7.6
$T2$	5	97.1	90.7	3.1	1.5
$D2$	5	94.3	86.4	3.1	3.1

remote to the placement of the observatories where the forecast is done.

The original predictors as defined in ERA-40 and ERA-Interim reanalysis grid were as follows: geopotential (Z), temperature (T), zonal wind speed (U), meridional wind speed (V) and relative humidity (RH) defined at the following 5 levels: 300, 500, 700, 850 and 1000 hPa, Additional variables considered were mean sea level pressure (MSL), surface pressure (SP), zonal wind speed at 10 m ($U10$), meridional wind speed at 10 m ($V10$), temperature at 2 m ($T2$) and dew point temperature at 2 m ($D2$). The selection of predictors is consistent with previous results of the authors regarding the optimal set of predictors (Fernandez-Ferrero et al., 2009). Finally, for all these variables at the 35 gridpoints, data were available at 00:00, 06:00, 12:00 and 18:00 GMT. That made a total of 4200 daily predictor variables in the area considered. Due to their different nature and ranges, all predictors were rescaled to have a mean of zero and a standard deviation of one (Imbert and Benestad, 2005)

With the aim to reduce the dimensionality of ERA-40 and ERA-Interim data, Empirical Orthogonal Functions (EOF) were calculated, selecting for each variable a varying number of EOF under the condition that the retained fraction of variance was at least 80 %. With this criterium, the number of EOF for each variable ranged from 5 to 28, thus notoriously reducing the number of predictors from 4200 to a final global amount of 79. These 79 leading EOF were calculated for ERA-40 (Table 1) and ERA-Interim (Table 2). The four values corresponding to the same days ($00Z$, $06Z$, $12Z$ and $18Z$) are considered as four samples at different times during the computation of the EOF in the same way that data with time delays is used during the computation of extended EOF (Weare and Nasstrom, 1982).

Surface observations at two locations (Zaragoza and Tortosa) were obtained from the European Climate Assessment (ECA) dataset http://eca.knmi.nl (Klein Tank et al., 2002). The variables obtained from ECA repository were temperature, relative humidity, mean sea level pressure, precipitation and wind speed and direction from which zonal and meridional components (U_s and V_s) were computed. Mean sea level pressures at Tortosa (48 m altitude.) and Zaragoza (247 m altitude) were corrected to surface pressure (SP) by assuming that the vertical temperature profile corresponds to the adiabatic lapse rate. Combining SP with observed temperature and relative humidity, specific humidity (q) values (kg water vapour/kg air) at surface level were calculated at both locations using the Clausius-Clapeyron equation. Zonal and meridional moisture fluxes (q_x and q_y expressed as kilograms of water vapour per square meter per second) were calculated as follows:

$$q_x = q\rho U_s \tag{1}$$

$$q_y = q\rho V_s \tag{2}$$

where ρ represents the density of air as a function of temperature.

In Zaragoza, the meridional component of q_{10} (q_y) accounts for 29.3 % of the overall variance associated to the q_{10} vector, while in Tortosa, the meridional component represents as much as 67.3 %. The most important values of the variables measured in Zaragoza and Tortosa can be seen in Table 3.

In a similar way as described above, (Clausius-Clapeyron followed by Eqs. (1) and (2) fed with $U10$ and $V10$ instead of U_s and V_s), zonal and meridional components of q_{10} were calculated on the ERA-40 and ERA-Interim gridpoints using $T2$, $D2$, SP and ρ. For comparison purposes, zonal and

Table 3. Observed values in Zaragoza and Tortosa (1961–2001).

		Average	5th percentil	95th percentil
Zaragoza	Temp (K)	288.42	277.15	300.15
	Rel. Hum. (%)	64.46	43.00	91.00
	Prec. (mm day^{-1})	0.90	0.00	5.30
	q_x (kg vapour m^{-2} s^{-1})	0.02	−0.04	0.07
	q_y (kg vapour m^{-2} s^{-1})	−0.01	−0.05	0.02
Tortosa	Temp (K)	290.68	281.15	300.35
	Rel. Hum. (%)	66.61	41.00	89.00
	Prec. (mm)	1.55	0.00	8.30
	q_x (kg vapour m^{-2} s^{-1})	−0.0016	−0.034	0.03
	q_y (kg vapour m^{-2} s^{-1})	0.001	−0.037	0.045

meridional components of q_{10} (q_x and q_y) at the nearest gridpoints to Zaragoza (13.6 km away) and Tortosa (65.1 km away) were considered. The reason is that any downscaling effort on q_{10} should yield better results that the values of q_{10} as calculated from both reanalyses at the geographically closest gridpoints.

Additionally, daily precipitation data were also retrieved from Global Precipitation Climatology Project (GPCP) http://jisao.washington.edu/data/gpcp/daily/ which provides gridded precipitation data with a 1° × 1° resolution (Huffman et al., 2001, Adler et al., 2003).

The precipitation data from the GPCP dataset (Adler et al., 2003; Huffmann et al., 2001) is used because previous studies (Lucarini et al., 2007) have already found inadequacies in precipitation data from ECMWF reanalyses. Precipitation observations are not directly assimilated during the preparation of reanalyses and therefore, they are produced by the model used in the data assimilation system. Therefore, it is better to use precipitation data from other sources like GPCP.

In the GPCP grid, the closest gridpoints to Zaragoza and Tortosa are located respectively 59 and 36 km away from the observatories. Persistence and precipitation data as given by the GPCP dataset at these two gridpoints were used as additional reference values. Again, the idea behind the use of these data is that any downscaling effort on precipitation could only be justified if better results than local persistence and/or raw GPCP data at these two nearest gridpoints were obtained. That is, it can not be assumed that observations are error free, particularly for a magnitude such as precipitation. If two different observational estimations of precipitation (ECA and GPCP) differ in a given amount, it can not be expected that a particular downscaling model could be more accurate than any of the observational datasets with respect to the other is.

2.2 Building the models

The original database consisted of 14 975 daily cases spanning from 1961 through 2001 and was splitted into a training (years 1961–1996, 13 149 cases) and a test period (years 1997–2001, 1826 cases). In both the training and the test periods, daily ERA-40 and surface ECA values were available. Additionaly, for the test period ERA-Interim data were also available.

In this study, downscaling models have been built for two locations (Zaragoza and Tortosa) and three variables: zonal and meridional components of q_{10} and precipitation. The steps followed in all cases have been:

1. For each of the 1826 days belonging to the test dataset, the 300 nearest cases among the 13 149 days corresponding to the training database are selected. The nearest cases are those with the smallest Euclidean distance to the current case as defined in the 79-dimensional hyperspace corresponding to the historical ERA-40 data (Table 1). The reason for choosing a number of 300 analogues was to allow for a reasonable number of cases (4–5) for each of the 79 candidate predictors, thus avoiding overfitting when applying linear regression at the following regression stage.

2. With these 300 analogues, two downscaling regression models are built using as candidate predictors the 79 ERA-40's principal components and, as predictand, the chosen variable (any of the the two components of q_{10} or precipitation in Zaragoza or Tortosa). Two techniques are used to build the models: random forests (RF, with $m = 9$ predictors) and multiple linear regression (MLR, stepwise regression).

   Being 79 the number of predictors, several values of m as suggested by the literature were tested with RF: $m = 9$ ($79^{0.5}$, usually employed for classification), $m = 26$ ($79/3$, usually employed for regression), $m = 79$

plus the sequence of $m = 10, 20, 30, 40, 50, 60, 70$. For these ten candidate values of m, the correlation coefficients between observations and predictions were computed for all the variables and also, the most influential inputs were identified. The correlation coefficients obtained with the different candidate values of m were not different at a 95 % confidence level for all the variables studied. Also, the relevant inputs identified were the same – as could be expected from CART-based RF (Grömping, 2009) –, so for the sake of simplicity $m = 9$ was the final value selected for this study.

At a preliminary stage of this work, several values of the number or trees were considered and it could be seen that the test set error did not decrease beyond approximately 100 trees. As suggested by the literature (Breiman, 2002; Peters et al., 2007) and since RF produces trees very rapidly, 1000 was the number of trees finally selected for all predictands. The predicted values was the average of the outputs from the 1000 trees (Breiman, 2001).

As mentioned above, two regression models for each predictand are fitted on the 300 nearest cases. In all cases, the candidate predictors are the 79 EOF from ERA-40 reanalysis and the predictand, the surface variable (q_x, q_y or precipitation in Zaragoza or Tortosa as derived from ECA observations). One of the models is fitted using a MLR and the other using random forests. Due to the gaps in the ECA database, some historical records of the predictands corresponding to the most similar 300 days identified in the atmospheric circulation analogues (reanalysis), are not present. For this reason, regression is carried out using a set of cases in which predictors and predictand are present that typically ranges between 250 and 300. For each case belonging to the test dataset, two models (RF and MLR) are fitted in this way.

3. Using the 79 EOF corresponding to every day of the test dataset as inputs, the two models previously fitted (RF and MLR) on the 250–300 most similar historic records were used to calculate an estimated value of the chosen variable for the same day in Zaragoza or Tortosa. To test the sensitivity of both techniques (RF and MLR) to the use of ERA-40 or ERA-Interim analyses, once the two models have been fitted on ERA-40 data, both models are run with two different sets of inputs: (i) the 79 EOF from ERA-40 (models denoted as RF ERA-40 and MLR ERA-40) and (ii) the 79 EOF obtained with ERA-Interim (RF ERA-Interim and MLR ERA-Interim). Additionally, a plain average obtained from ECA values corresponding to the 250–300 most similar daily cases identified, is used to build an dditional analogue-type downscaling model (denoted as "Analogues" model).

4. Finally, as mentioned above, the most evident estimations of q_{10} and precipitation were also considered. In the case of zonal and meridional components of q_{10}, (q_x, q_y,) the values directly calculated using ERA-40 and ERA-Interim reanalyses raw data at the geographically nearest points to Zaragoza and Tortosa. In the case of precipitation, the idea is the same but two other references were used. The first one was the GPCP satellite and rain gauge merged precipitation data set. The second one was just to consider the persistence of levels from the previous day.

It is worth mentioning that the sensitivity of this methodology to (i) the domain size and (ii) the use of "mixed EOF" (Benestad et al., 2002) instead of independent EOF, was intensively tested. The results show that for this area of the Iberian Peninsule, changing the area covered by the domain by an order of magnitude (in km^2) does not have an important impact on results. This is due to the fact that the most influential variables used for downscaling exhibit a small spatial variability. As a consequence, using independent EOF or mixed EOF obtained from variables that do not have an important spatial variability, does not have an important effect on results either (for in-depth details, please see the discussion forum of HESS corresponding to this paper).

2.3 Evaluation of models

All the models have been evaluated and intercompared using the 1826 observations corresponding to the test dataset. Being the objective of this paper to assess the overall performance of the models mentioned above, a set of statistical indicators has been chosen to compare observations and models' predictions:

  i. Correlation coefficient (R).

 ii. Ratio of standard deviations (RSD) between observations and model's standard deviations. A good model should show values near one.

iii. Root mean square error (RMSE).

With the aim to summarize in a visual manner models' performance according to this group of three statistical indicators (R, RSD and RMSE), Taylor diagrams (Taylor, 2001) have been plotted. Additionally, three more statistical indicators have also been considered:

 iv. Fraction of two (FA2). FA2 indicates the fraction of cases in which the ratio between observations and model's predictions falls in the range [0.5–2].

  v. Ratio of means (RM) between observations and model's averages.

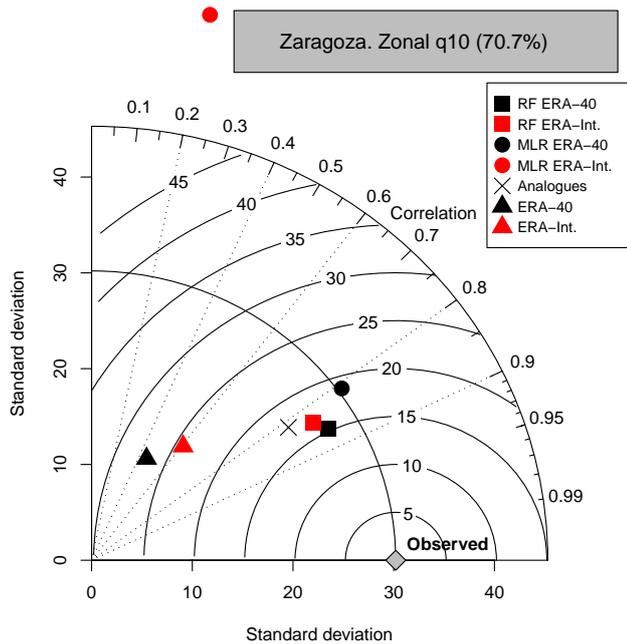 vi. Index of agreement D. In order to overcome some of the widely reported problems associated to the plain use of

Fig. 2. Taylor's diagram of the models for daily zonal q_{10} in Zaragoza. 1826 test cases (1997–2001). In brackets, percentage of the overall q_{10} variability corresponding to the zonal component.

Fig. 3. Taylor's diagram of the models for daily meridional q_{10} in Zaragoza. 1826 test cases (1997–2001). In brackets, percentage of the overall q_{10} variability corresponding to the meridional component.

the correlation coefficient, an additional indicator called index of agreement has been proposed (Wilmott, 1981, 1982; Wilmott et al., 1985) and adopted for this study. It is an indicator of the overall agreement between model and observations that ranges between 0 and 1 (perfect model).

In all cases, bootstrap resampling (1000 samples extracted from the test data set) has been used to calculate 95 % confidence intervals corresponding to these statistical indicators. Likewise, for model intercomparison purposes, 95 % confidence intervals have also been calculated to assess differences between two models' performances as described by the set of indicators metioned above.

The relative importance of input variables is to be assessed in two different ways, depending on the nature of the model. In the case of RF models, input relevance is related to the increase of the mean square error MSE (%) due to the random permutation of the input variable's values. For linear regression models, the most important variables will be the ones that after being incorporated into the equation, are responsible for the highest increases in the overall determination coefficient R^2 (Grömping, 2009). The most influential predictors with one technique or another (MLR or RF) are roughly the same, as could be expected from the literature (Grömpig, 2009).
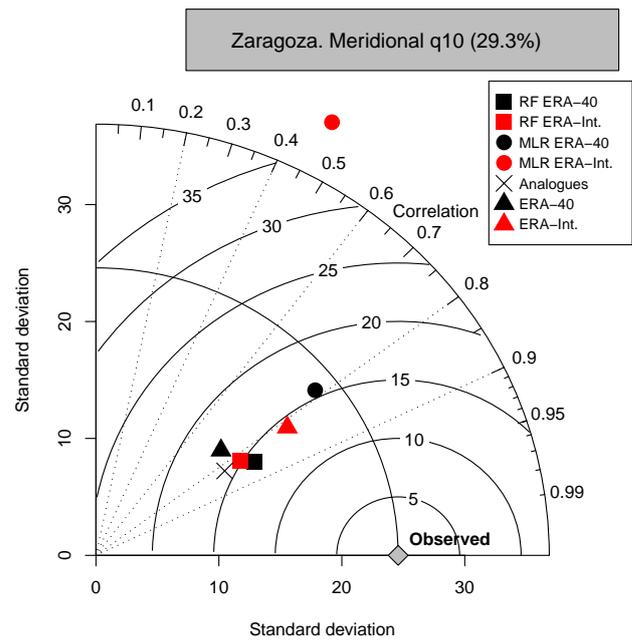
3 Results

The main results can be seen in Tables 4–5 and Figs. 2–7. In both locations (Zaragoza and Tortosa) models' best performance for precipitation ($R \sim 0.5$) are notoriously poorer than for any of the two components of q_{10} ($R \sim [0.8, 0.85]$). The reason for this is that the nature of precipitation is much more intermittent and dependent of very local factors than in the case of q_{10}.

It is worth mentioning that models fitted with ERA-40, when they are fed with ERA-Interim data, experiment a deterioration in performance which is more evident in analogues followed by MLR. In the case of analogues+RF, the degradation is much smaller being a reasonable option for the future fitting RF models on ERA-40 data (covers the period from mid-1957 to mid-2002) and use them with the more recently available ERA-Interim (from 1989 onwards). For zonal and meridional components of q_{10} at both locations, the comparison of the downscaling using raw ERA-40 or ERA-Interim data at the nearest gridpoints, indicates that ERA-Interim values tend to be more accurate than those yielded by ERA-40. The lack of continuity between both reanalyses suggests that the procedures and algorithms used in ERA-Interim provide a better description of at least q_{10}. Even though they perform worse than RF ERA-40 and MLR ERA40 models, both reanalyses' direct outputs yield reasonable estimates of q_{10}, particularly for the meridional component. The overall performance of raw ERA-Interim predictions, particularly, in

Table 4. Models' results in Zaragoza when applied to the 1826 cases corresponding to the test dataset (1997–2001).

		RF ERA-40	RF ERA-Interim	MLR ERA-40	MLR ERA-Interim	Analogues	ERA-40*/ GPCP**	ERA-Interim*/ Persistence**
ZARAGOZA	Zonal q_{10}*							
	R	0.86	0.84	0.81	0.20	0.82	0.46	0.61
	RSD	0.90	0.87	1.01	1.93	0.79	0.40	0.50
	RMSE	16.35	17.89	19.41	61.80	18.42	27.56	24.72
	FA2	0.70	0.66	0.63	0.35	0.63	0.28	0.36
	RM	1.46	1.54	1.40	2.21	1.45	0.54	0.62
	D	0.74	0.71	0.70	0.38	0.69	0.37	0.47
	Meridional q_{10}*							
	R	0.85	0.82	0.78	0.46	0.82	0.75	0.82
	RSD	0.62	0.58	0.93	1.70	0.52	0.55	0.77
	RMSE	14.59	15.39	15.65	37.62	16.00	21.40	17.48
	FA2	0.56	0.53	0.57	0.35	0.53	0.34	0.53
	RM	0.78	0.83	1.01	1.25	0.86	0.14	0.32
	D	0.68	0.65	0.70	0.47	0.63	0.55	0.66
	Precipitation**							
	R	0.55	0.43	0.47	0.15	0.53	0.38	0.16
	RSD	0.33	0.37	0.68	1.20	0.26	1.17	1.00
	RMSE	2.97	3.10	3.13	5.12	3.05	4.24	4.43
	FA2	0.08	0.08	0.07	0.05	0.09	0.05	0.03
	RM	0.66	1.13	1.48	2.56	0.80	2.02	1.00
	D	0.60	0.45	0.52	0.32	0.53	0.49	0.51

Table 5. Models' results in Tortosa when applied to the 1826 cases corresponding to the test dataset (1997–2001).

		RF ERA-40	RF ERA-Interim	MLR ERA-40	MLR ERA-Interim	Analogues	ERA-40*/ GPCP**	ERA-Interim*/ Persistence**
TORTOSA	Zonal q_{10}*							
	R	0.80	0.75	0.71	0.13	0.76	0.53	0.55
	RSD	0.63	0.55	0.83	1.60	0.55	0.76	0.72
	RMSE	12.12	13.37	13.88	34.14	13.31	18.99	18.58
	FA2	0.49	0.41	0.48	0.24	0.45	0.34	0.34
	RM	0.90	0.68	1.02	0.53	1.48	−2.26	−2.35
	D	0.65	0.59	0.63	0.35	0.60	0.50	0.49
	Meridional q_{10}*							
	R	0.83	0.79	0.77	0.22	0.79	0.71	0.74
	RSD	0.93	0.82	1.02	1.70	0.82	0.66	0.79
	RMSE	12.71	13.87	15.47	40.51	14.31	16.30	15.65
	FA2	0.65	0.57	0.57	0.30	0.60	0.46	0.52
	RM	1.02	0.73	0.48	−1.12	2.61	−0.14	−0.61
	D	0.76	0.70	0.70	0.39	0.71	0.61	0.66
	Precipitation**							
	R	0.49	0.40	0.46	0.21	0.50	0.33	0.26
	RSD	0.41	0.47	0.75	1.50	0.34	0.75	1.00
	RMSE	4.70	4.96	5.05	9.20	4.72	5.56	6.54
	FA2	0.07	0.06	0.06	0.03	0.07	0.03	0.02
	RM	0.74	1.34	1.50	3.31	0.94	1.38	1.00
	D	0.61	0.43	0.53	0.29	0.53	0.52	0.55

the meridional component of q_{10} at Zaragoza and Tortosa, compares well with that obtained by analogues. Zaragoza (13.6 km) is closer than Tortosa (65.1 km) to its corresponding nearest gridpoint and this can explain why results at Zaragoza are somewhat better than those at Tortosa.

Analyzing in detail the downscaling of q_{10} in Zaragoza, it can be seen that RF ERA-40 gives the best results with values of R of 0.86 (zonal) and 0.85 (meridional). However, the comparison with MLR-ERA40 (Table 4) indicates only a marginal improvement in favour of RF ERA-40. For the zonal component of q_{10} (70.7 % of variance), R, RMSE, FA2, and D skill scores exhibit better values with RF ERA-40 model than with MLR ERA-40. RM values are not different at a 95 % confidence level while RSD is closer to 1 for
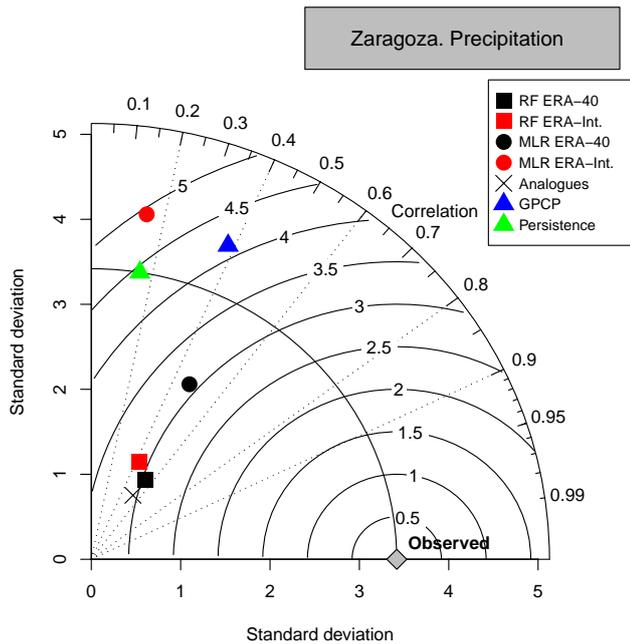
Fig. 4. Taylor's diagram of the models for daily precipitation in Zaragoza. 1826 test cases (1997–2001).

Fig. 5. Taylor's diagram of the models for daily zonal q_{10} in Tortosa. 1826 test cases (1997–2001). In brackets, percentage of the overall q_{10} variability corresponding to the zonal component.

MLR ERA-40 (Fig. 2). In the case of the meridional component of q_{10} (29.3 % of variance), RMSE, RM and D are not different at a 95 % confidence level, while RSD and RM are better for the linear MLR ERA40 model. Only the correlation coefficient is higher for RF ERA-40 (Fig. 3).

In the case of the zonal component of q_{10} in Tortosa (Fig. 4) (Table 5), RF ERA-40 and MLR ERA-40 at a 95 % confidence level do not have different values of FA2, RM, and D. R and RMSE are better for RF ERA-40 while MLR ERA-40 has got a better RSD indicator. For the meridional component of q_{10}, (Fig. 5) (Table 5), RF ERA40 slightly outperforms MLR ERA40 in all the statistical indicators, except for RSD. In all the cases mentioned above, it can be seen that analogues only have a little worse performance when compared with RF ERA-40 and MLR ERA-40. Therefore, it can be concluded that obtaining analogues represents the most efficient step in the model building process and accounts for the most important part of the goodnes of fit. Using RF or MLR at a next stage adds a much smaller improvement. The employment at this second stage of RF, a recently developed machine learning algorithm intended to deal with highly non linear mechanisms like the ones known to be involved in atmospheric processes, does not represent a net improvement when compared with the classical MLR model.

The most influential variables in the downscaling of q_{10} at both locations, have been identified, as mentioned above, by computing the increase (%) of MSE due to the permutation of their data when acting as inputs for the RF. In the 1826 RF models fitted, the most frequently selected as influential
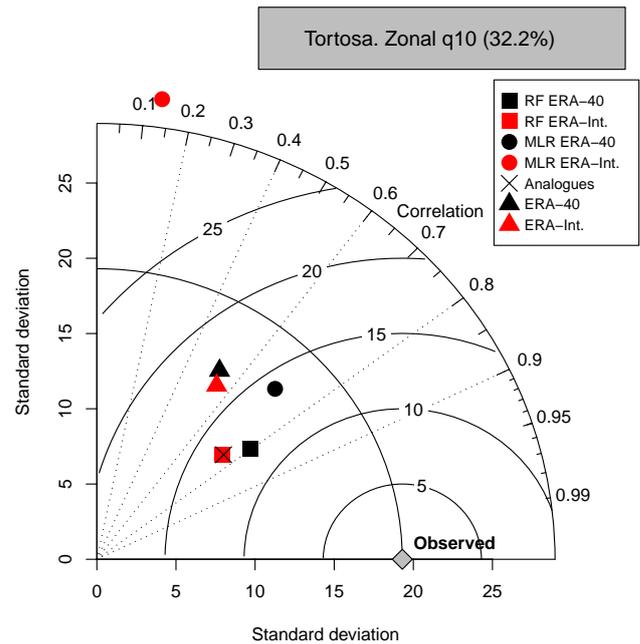
inputs have been the first EOF of $V\,10$, the first EOF of T2 and the first and second EOF of D2 The interpretation is as follows. As can be seen in q_{10} roses at both locations (not shown) the direction of q_{10} at both locations is mainly zonal. The meridional component of surface q_{10} affects moisture fluxes due to the fact that meridionally flowing air must cross mountain ranges over central Iberian Peninsula, thus suffering a strong Föhn effect. Similarly, warmer air leads to a higher moisture flux as explained by the Clausius-Clapeyron equation. The variability in the dew-point temperature enclosed in the two leading EOF reflect the main temporal (1st EOF) and spatial changes (2nd EOF) of the humidity in the air.

Regarding precipitation at both locations, for RF ERA-40, MLR ERA-40 and analogues, the statistical indicators R, RMSE and FA2 do not differ at a 95 % confidence level. In both locations, the model of analogues can represent better the observed mean, MLR ERA-40 exhibits the best RSD and RF ERA-40 has the best D value. Therefore, it can be concluded that, if globally considered, RF ERA-40 and MLR ERA-40 models do not represent any significant improvement over the plain use of analogues (Tables 4–5). At both locations, these three models outperform GPCP predictions and also persistence. The results are similar to those found in the literature (see Introduction) and all models tend to underpredict extreme values (please, see discussion forum on HESS website). The most important input variables for precipitation downscaling have been the first EOF of relative humidity RH, the first EOF of meridional wind speed V
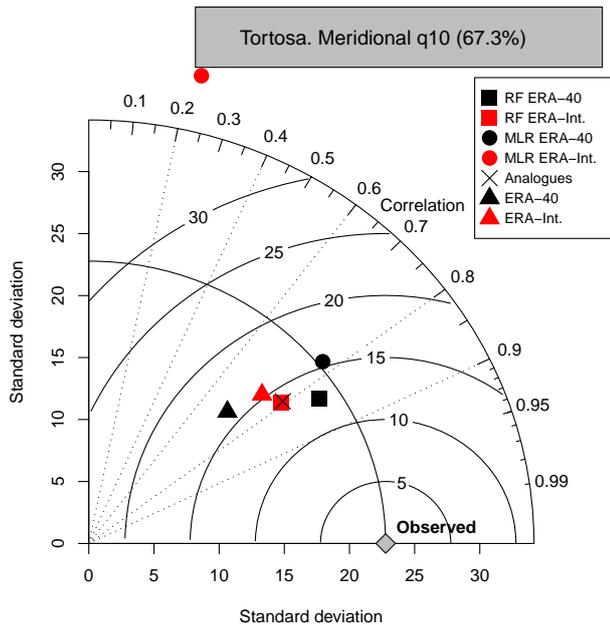
Fig. 6. Taylor's diagram of the models for daily meridional q_{10} in Tortosa. 1826 test cases (1997–2001). In brackets, percentage of the overall q_{10} variability corresponding to the meridional component.

and the second EOF of dew-point temperature $D2$. Similarly, precipitation depends on the availability of moisture and the saturation of air (leading EOF of relative humidity), the direction (northern/southern) from which the air flows (1st EOF of V) and local effects accounted for by the second EOF of $D2$.

4 Conclusions and future outlook

The results of the statistical downscaling carried out for q_{10} and precipitation indicate that analogues represent a powerful and at the same time, easy-to-use tool. In most cases, their results outperform raw predictions derived from reanalysis data (ERA-40, ERA-Interim and GPCP) at the nearest gridpoints or persistence of daily precipitation. Incorporating a second regression stage represents a clear though not overwhelming improvement. However, using at this second stage either a classical linear model or a more sophisticated tool like RF, does not make a net difference. As mentioned in the Introduction, similar effects with other nonlinear techniques like NN have been described in the literature. The explanation for this may lay in the fact that the combination of a great number of highly nonlinear mechanisms involved in the downscaling might result in a linearization of the overall effect. A further explanation already pointed out in the literature (Zorita and Von Storch, 1999) might be that most of the overall nonlinearity involved, has been already captured and described at the stage of the selection of analogues. As a conclusion, the combination of two techniques which are quite easy to use and implement like analogues and multi-
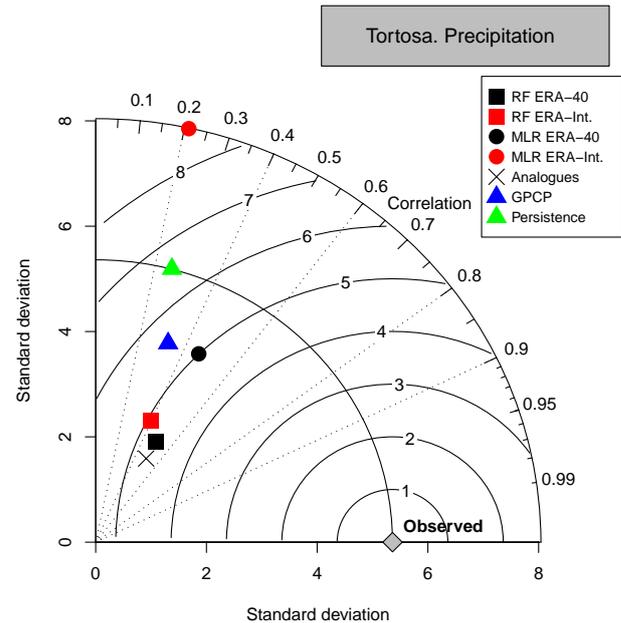
Fig. 7. Taylor's diagram of the models for daily precipitation in Tortosa. 1826 test cases (1997–2001).

ple linear regression, can be the best compromise between accuracy and simplicity in similar downscaling problems. However, an important shortcome of MLR is the lack of continuity between ERA-40 and ERA-Interim which introduces a heavy degradation in model's performance, thus not making possible the use of an important set of historical records and learning periods, currently available only as ERA-40. Instead, RF models fitted on ERA-40 data and fed with ERA-Interim, experiment a much smaller deterioration. The reason might be that MLR models heavily rely on the coefficients of an equation and the differences between ERA-40 and ERA-Interim input data tend to be amplified by the values of the coefficientes. However, the nature of the regression with RF is different and is based on CART where homogeneous areas in the final leafs are sought as the trees split at the different stages. In this sense, it can be expected that similar homogeneous areas can be described either using ERA-40 or ERA-Interim EOF (Tables 1–2) as inputs, since leading EOF from both reanalysis are likely to be describing the same physical effects on the studied area. Anyway, before considering it as definitive, this result should be confirmed for other variables, since it might happen that it cannot be generalized to other cases with, for instance, larger biases, as could happen when using GCM output in a perfect-prog approach

Further research is currently being carried out by this group in two directions: (i) to find a relationship between ERA-Interim reanalysis and moisture flux at several locations on the Ebro Valley but at different upper levels (ii) to apply the methodology followed so far to other variables involved in the water cycle.

All the calculations have been carried out using the freely available software R (R core development team, 2009) and apart from the core module, four specific packages have been used: (i) "FactoMineR" for calculation of EOF (Husson et al., 2008), (ii) "MASS" for multiple linear regression (Venables and Ripley, 2002), (iii) "randomForest" for RF (Liaw and Wiener, 2002) and (iv) "plotrix" for Taylor diagrams (Lemon, 2006).

References

Adler, R. F., Huffman, G. J., Chang, A., Ferraro, R., Xie, P., Janowiak, J., Rudolf, B., Schneider, U., Curtis, S., Bolvin, D., Gruber, A., Susskind, J., Arkin, P., Nelkin, E.: The Version 2 Global Precipitation Climatology Project (GPCP) Monthly Precipitation Analysis (1979–Present), J. Hydrometeor., 4, 1147–1167, 2003.

Benestad, R. E.: A Simple Test for Changes in Statistical Distributions, Eos, 89(41), 389–390, 2008.

Benestad, R. E.: Downscaling Precipitation Extremes: Correction of Analog Models through PDF Predictions', Theor. Appl. Clim., 100(1), 1–21, doi:10.1007/s00704-009-0158-1, 2010.

Benestad, R. E., Førland, E. J., and Hanssen-Bauer, I.: Empirically downscaled temperature scenarios for Svalbard, Atmos. Sci. Lett., 3(2–4), 71–93, doi:10.1006/asle.2002.0051, 2002.

Berbery, E. H. and Rasmusson, E. M.: Mississippi moisture budgets on regional scales, Mon. Weather Rev., 127, 2654–2673, 1999.

Breiman, L.: Random forests, Mach. Learn., 45, 5–32, 2001.

Breiman, L.: Manual on setting up, using, and understanding random forests v3.1, available at: http://oz.berkeley.edu/users/breiman/Using_random_forests_V3.1.pdf (last access: 18 February 2011), 2002.

Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J.: Classification and Regression Trees. Monterey, CA: Wadsworth. Wadsworth Statistics/Probability Series, Wadsworth Advanced Books and Software, Belmont, CA, 1984.

Cavazos, T.: Using Self-Organizing Maps to investigate extreme climate events: an application to wintertime precipitation in the Balkans, J. Clim., 13, 1718–1732, 2000.

Cavazos, T. and Hewitson, B. C.: Performance of NCEP–NCAR reanalysis variables in statistical downscaling of daily precipitation, Clim. Res., 28, 95–107, 2005.

Davy, J. R., Woods, M. J., Russelll, C. J., and Coppin, P. A.: Statistical downscaling of wind variability from meteorological fields, Bound.-Lay. Metorol., 135, 161–175, 2010.

Eccel, E., Ghielmi, L., Granitto, P., Barbiero, R., Grazzini, F., and Cesari, D.: Prediction of minimum temperatures in an alpine region by linear and non-linear post-processing of meteorological models, Nonlin. Processes Geophys., 14, 211–222, doi:10.5194/npg-14-211-2007, 2007.

Fernández, J. and Sáenz, J.: Improved field reconstruction with the analog method: Searching the CCA space, Clim. Res., 24, 199–213, 2003.

Fernandez, J., Sáenz, J., and Zorita, E.: Analysis of wintertime atmospheric moisture transport and its variability over southern Europe in the NCEP Reanalyses, Clim. Res., 23, 195–215, 2003.

Fernández-Ferrero, A., Sáenz, J., Ibarra-Berastegi, G., and Fernández, J.: Evaluation of statistical downscaling in short range precipitation forecasting, Atmos. Res., 94, 448–461, 2009.

Fernández-Ferrero, A., Sáenz, J., and Ibarra-Berastegi, G.: Comparison of the performance of different analogue-based Bayesian probabilistic precipitation forecasts over Bilbao, Mon. Weather. Rev., 138, 3107–3119, 2010.

Gimeno, L., Nieto, R., and Trigo, R. M.: Where does the Iberian Peninsula moisture come from? An answer based on a Lagrangian approach, J. Hydrometeor., 11, 421–436, 2010.

Grömping, U.: Variable Importance Assessment in Regression: Linear Regression versus Random Forest, Amer. Stat., 63, 308-319, 2009.

Hamill, T. M. and Whitaker, J. S.: Probabilistic Quantitative Precipitation Forecasts Based on Reforecast Analogs: Theory and Application, Mon. Weather Rev., 134, 3209–3229, 2006.

Hastie, T., Tibshirani, R., and Friedman, J.: The elements of statistical learning. Data mining, inference and prediction, Springer, 2001.

Huffman, G. J., Adler, R. F., Morrissey, M., Bolvin, D. T., Curtis, S., Joyce, R., McGavock, B., and Susskind, J.: Global precipitation at one-degree daily resolution from multi-satellite observations, J. Hydrometeor., 2, 36–50, 2001.

Husson, F., Josse, J., Le, S., and Mazet, J.: FactoMineR: factor analysis and data mining with R, R package version 1.10., available at: http://factominer.free.fr, http://www.agrocampus-rennes.fr/math (last access: 18 February 2011), 2008.

Huth, R.: Statistical downscaling of daily temperature in Central Europe, J. Climate, 15, 1731–1742, 2002.

Huth, R.: Downscaling of humidity variables: a search for suitable predictors and predictands, Int. J. Climatol., 25(2), 243–250, 2005.

Huth, R., Kliegrova, S., and Metelka, L.: Non-linearity in statistical downscaling: does it bring any improvement for daily temperature in Europe?, Int. J. Climatol., 28, 465–477, 2008.

Hurrell, J. W.: Decadal trends in the North Atlantic Oscillation: regional temperatures and precipitation, Science, 269, 676–679, 1995.

Imbert, A. and Benestad, R. E.: An improvement of analog model strategy for more reliable local climate change scenarios, Theor. Appl. Clim., 82, 245–255, doi:10.1007/s00704-005-0133-4, 2005.

Ishwaran, H.: Variable Importance in Binary Regression Trees and Forests, Electronic Journal of Statistics, 1, 519–537, 2007.

Klein Tank, A. M. G., Wijngaard, J. B., Können, G. P., Böhm, R.,

Demarée, G., Gocheva, A., Mileta, M., Pashiardis, S., Hejkrlik, L., Kern-Hansen, C., Heino, R., Bessemoulin, P., Müller-Westermeier, G., Tzanakou, M., Szalai, S., Pálsdóttir, T., Fitzgerald, D., Rubin, S., Capaldo, M., Maugeri, M., Leitass, A., Bukantis, A., Aberfeld, R., van Engelen, A. F. V., Forland, E., Mietus, M., Coelho, F., Mares, C., Razuvaev, V., Nieplova, E., Cegnar, T., López, J. Antonio, Dahlström, B., Moberg, A., Kirchhofer, W., Ceylan, A., Pachaliuk, O., Alexander, L. V., and Petrovic, P.: Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment, Data and metadata available at: http://eca.knmi.nl, Int. J. Climatol., 22, 1441–1453, 2002.

Lemon, J.: Plotrix: a package in the red light district of R, R-News, 6(4), 8–12, 2006.

Liaw, M. and Wiener, M.: Classification and Regression by randomForest, R News, 2(3), 18–22, 2002.

Lucarini, V., Calmanti, S., Della Aquila, A., Ruti, P. M., and Speranza, A.: Intercomparison of the northern hemisphere winter mid-latitude atmospheric variability of the IPCC models, Clim. Dynam., 28(7–8), 829–849, 2007.

Mariotti, A.: Recent changes in the Mediterranean water cycle: A pathway toward long-term regional hydroclimatic change?, J. Climate, 23, 1513–1525, 2010.

Matulla, C., Zhang, X., Wang, X.L., Wang, J., Zorita, E., Wagner, S., and von Storch, H.: Influence of similarity measures on the performance of the analog method for downscaling daily precipitation, Clim. Dynam., 30, 133–144, 2008.

Miksovsky, J. and Raidl, A.: Testing the performance of three nonlinear methods of time seriesanalysis for prediction and downscaling of European daily temperatures, Nonlin. Processes Geophys., 12, 979–991, doi:10.5194/npg-12-979-2005, 2005.

Peters, J., De Baets, B., Verhoest, N. E. C., Samson, R., Degroeve, S., De Becker, P., and Huybrechts, W.: Random forests as a tool for ecohydrological distribution modelling, Ecol. Model., 304–318, 2007.

R Development Core Team.: R: A language and environment for statistical computing, R Foundation for Statistical Computing,Vienna, Austria, ISBN 3-900051-07-0, available at: http://www.R-project.org (last access: 18 February 2011), 2009.

Rodríguez-Puebla, C., Encinas, A. H., Nieto, S., and Garmendia, J.: Spatial And Temporal Patterns Of Annual Precipitation Variability Over The Iberian Peninsula, Int. J. Climatol., 18, 299–316, 1998.

Schoof, J. T. and Pryor, S. C.: Downscaling temperature and precipitation: A comparison of regression-based methods and artificial neural networks, Int. J. Climatol., 21, 773–790, 2001.

Siroky, D. S.: Navigating Random Forests and related advances in algorithmic modeling, Stat. Surv., 3, 147–163, 2009.

Stephenson, D. B., Coelho, C. A. S., Doblas-Reyes, F. J., and Balmaseda, M.: Forecast assimilation: a unified framework for the combination of multi-model weather and climate predictions, Tellus A, 57, 253–264, 2005.

Taylor, E. K.: Summarizing multiple aspects of model performance in a single diagram, J. Geo. Res., 106(D7), 7183–7192, 2001.

Timbal, B. and Jones, D. A.: Future projections of winter rainfall in southeast Australia using a statistical downscaling technique, Clim. Change, 86, 165–187, 2008.

Trigo, R. M. and Palutikof, J. P.: Precipitation scenarios over Iberia: A comparison between direct GCM output and different downscaling techniques, J. Climate, 14, 4422–4446, 2001.

Uppala, S., Kallberg, P., Simmons, A., Andrae, U., da Costa Bechtold, V., Fiorino, M., Gibson, J., Haseler, J., Hernandez, A., Kelly, G., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R., Andersson, E., Arpe, K., Balmaseda, M., Beljaars, A., van de Berg, L., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Holm, E., Hoskins, B., Isaksen, L., Janssen, P., Jenne, R., McNally, A., Mahfouf, J.-F., Morcrette, J.-J., Rayner, N., Saunders, R., Simon, P., Sterl, A., Trenberth, K., Untch, A., Vasiljevic, D., Viterbo, P., and Woollen, J.: The ERA-40 re-analysis, Quart. J. Roy. Meteor. Soc., 131, 2961-3012, 2005.

Venables, W. N. and Ripley, B. D.: Modern Applied Statistics with S, 4th Edn., Springer, New York, ISBN 0-387-95457-0, 2002.

Voisin, N., Schaake, J. C., and Lettenmaier, D. P.: Calibration and Downscaling Methods for Quantitative Ensemble Precipitation Forecasts, Weather Forecast., 25(6), 1603–1627, 2010.

Weare, B. C. and Nasstrom, J. S.: Examples of extended empirical orthogonal function analyses, Mon. Weather Rev., 110, 481–485, 1982.

Weichert, A. and Bürger, G.: Linear versus nonlinear techniques in downscaling. Clim. Res., 10, 83–93, 1998.

Wei, Y., Bárdossy, A., and Caspary, H. J.: Downscaling daily precipitation time series using a combined circulation- and regression-based approach, Theor. Appl. Climatol., 102, 439–454, 2010.

Wilby, R. L. and Wigley, T. M. L.: Precipitation predictors for downscaling: observed and general circulation model relationships, Int. J. Climatol., 20, 641–661, 2000.

Wilmott, C. J.: On the validation of models, Phys. Geo., 2, 184–194, 1981.

Wilmott, C. J.: Some comments on the evaluation of model performance, Bull. Am. Meteorol. Soc., 63(11), 1309–1313, 1982.

Wilmott, C. J., Ackleson, S. G., Davis, R. E., Feddema, J. J., Klink, K. M., Legates, D. R., O'Donnell, J., and Rowe, M. C.: Statistics for the evaluation and comparison of models, J. Geo. Res., 90, 8995–9005, 1985.

Witten, I. H. and Frank, E.: Data mining. Practical machine learning tools and techniques, 2nd Edn., ISBN-13: 978-0-12-08840-0, Elsevier Inc., 2005.

Zorita, E. and von Storch, H.: The analog method, as a simple statistical downscaling technique: comparison with more complicated methods, J. Climate, 12, 2474–2489, 1999.