



# Validation of a new SAFRAN-based gridded precipitation product for Spain and comparisons to Spain02 and ERA-Interim

Pere Quintana-Seguí<sup>1</sup>, Marco Turco<sup>2</sup>, Sixto Herrera<sup>3</sup>, and Gonzalo Miguez-Macho<sup>4</sup>

<sup>1</sup>Observatori de l'Ebre (OE), Universitat Ramon Llull – CSIC, Horta Alta 38, 43520 Roquetes, Spain

<sup>2</sup>Department of Applied Physics, University of Barcelona (UB), Av. Diagonal 647, 08028 Barcelona, Spain

<sup>3</sup>Meteorology Group, Department of Applied Mathematics and Computer Science, Universidad de Cantabria, Avenida de los Castros s/n, 39005 Santander, Spain

<sup>4</sup>Grupo de Física No Lineal, Universidade de Santiago Compostela, Calle Xosé María Suárez Núñez s/n., 15782 Santiago de Compostela, Spain

*Correspondence to:* Pere Quintana-Seguí (pquintana@obsebre.es)

Received: 13 July 2016 – Discussion started: 15 September 2016

Revised: 5 March 2017 – Accepted: 24 March 2017 – Published: 21 April 2017

**Abstract.** Offline land surface model (LSM) simulations are useful for studying the continental hydrological cycle. Because of the nonlinearities in the models, the results are very sensitive to the quality of the meteorological forcing; thus, high-quality gridded datasets of screen-level meteorological variables are needed. Precipitation datasets are particularly difficult to produce due to the inherent spatial and temporal heterogeneity of that variable. They do, however, have a large impact on the simulations, and it is thus necessary to carefully evaluate their quality in great detail.

This paper reports the quality of two high-resolution precipitation datasets for Spain at the daily time scale: the new SAFRAN-based dataset and Spain02. SAFRAN is a meteorological analysis system that was designed to force LSMs and has recently been extended to the entirety of Spain for a long period of time (1979/1980–2013/2014). Spain02 is a daily precipitation dataset for Spain and was created mainly to validate regional climate models. In addition, ERA-Interim is included in the comparison to show the differences between local high-resolution and global low-resolution products. The study compares the different precipitation analyses with rain gauge data and assesses their temporal and spatial similarities to the observations.

The validation of SAFRAN with independent data shows that this is a robust product. SAFRAN and Spain02 have very similar scores, although the latter slightly surpasses the former. The scores are robust with altitude and throughout the year, save perhaps in summer when a diminished skill is

observed. As expected, SAFRAN and Spain02 perform better than ERA-Interim, which has difficulty capturing the effects of the relief on precipitation due to its low resolution. However, ERA-Interim reproduces spells remarkably well in contrast to the low skill shown by the high-resolution products. The high-resolution gridded products overestimate the number of precipitation days, which is a problem that affects SAFRAN more than Spain02 and is likely caused by the interpolation method. Both SAFRAN and Spain02 underestimate high precipitation events, but SAFRAN does so more than Spain02. The overestimation of low precipitation events and the underestimation of intense episodes will probably have hydrological consequences once the data are used to force a land surface or hydrological model.

## 1 Introduction

Good knowledge of the continental water cycle and its quantification is critical for society, mainly because it represents both a resource and a hazard. As a consequence, a better understanding of the continental water cycle is a key issue for climate change impact and adaptation studies, and it has become a strategic topic in national and international climate programs, such as the Global Energy and Water Cycle Exchanges Project (GEWEX; <http://www.gewex.org/>) at the global scale and the Hydrological cycle in the Mediterranean Experiment (HyMeX; <http://www.hymex.org/>) at the

Mediterranean scale. Simulating the continental water cycle is difficult. For example, precipitation is a challenging variable due to its spatial heterogeneity, spatial variability and time variability. Evapotranspiration is a very complex process that is difficult to measure and depends on the atmosphere, soil and vegetation in a complex manner (e.g., Peters-Lidard et al., 2011; Nasonova et al., 2011; Long et al., 2014).

There are different approaches to simulating the continental cycle. These range from easy-to-calibrate simple water balance models (e.g., Orth and Seneviratne, 2015) to the more complex land surface models (LSM). LSMs simulate the physical processes at the interface among the soil, vegetation and the atmosphere. These models may be run offline and forced by a gridded dataset of screen-level meteorological variables or online and coupled to an atmospheric model.

The main advantage of using offline LSMs is that they avoid atmospheric model biases because they are forced by gridded observational datasets, but they miss many feedback processes. Such systems have been applied at the global, continental, national and basin scales (Rodell et al., 2004; Decharme et al., 2012; Balsamo et al., 2012; Cosgrove, 2003; Mitchell, 2004; Chen et al., 2007; Habets et al., 2008; Artinyan et al., 2008; Szczypta et al., 2012; Barbu et al., 2014; Martínez de la Torre, 2014) in a wide range of work that includes the study of water resources, the initialization of meteorological models and the study of the continental water cycle.

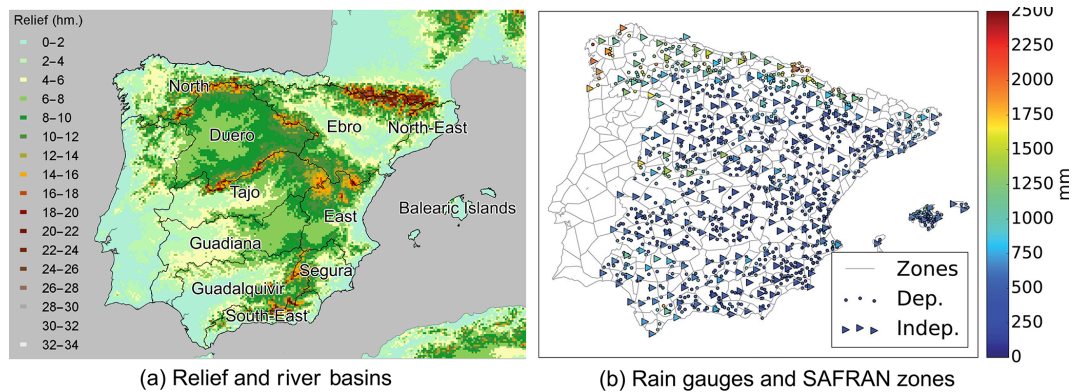
The performance of offline LSM simulations depends strongly on the quality of the meteorological datasets used to force the models. In addition, the nonlinear nature of the LSMs, which may amplify the errors, makes it essential to ensure the good quality of the meteorological forcing datasets for land surface and hydrological modeling. Unfortunately, it is not an easy task to build such gridded datasets that include all of the meteorological parameters and reach the spatial (at least 5 km) and temporal (intra-daily) resolutions needed to force an LSM. Meteorological analysis systems such as CANARI (Taillefer, 2002), SPAN (Rodríguez et al., 2003; Navascués et al., 2003; Cansado et al., 2003), MESAN (Häggmark et al., 2000) MESCAN (Soci et al., 2016) and SAFRAN (Durand et al., 1993, 1999; Quintana-Seguí et al., 2008; Vidal et al., 2010; Quintana-Seguí et al., 2016) are well suited for this task.

Concerning precipitation, which is the variable examined in this paper, there are other interesting reference datasets that might not be able to force an LSM because they may not have sufficient temporal or spatial resolution (they are often daily products with resolutions that range from 50 to 12 km); however, they are well known and widely used for other applications, such as the validation (e.g., Gómez-Navarro et al., 2012; Turco et al., 2013a) and post-processing of regional climate models (Turco et al., 2011; Casanueva et al., 2015) or the statistical downscaling of global climate models (Casanueva et al., 2016).

It is interesting to compare the analysis systems mentioned in the previous paragraph with these daily reference datasets. At the European scale, E-OBS (Haylock et al., 2008) is probably the best known of such products. However, in some countries, the station density considered when building E-OBS was very low, thus limiting the capability of that dataset to properly reproduce the climatic variability in those regions. As a result, some products have been developed for specific regions or countries, such as the gridded dataset EURO4M-APGD (Isotta et al., 2014) for the greater Alpine region and Spain02 (Herrera et al., 2012, 2016) in Spain, which consider a denser station network.

A careful validation of these products should be performed to ensure that they correctly reproduce the important characteristics of the climate in the region under study. The validation should test both the temporal and spatial accuracies of the products, not only for the mean values but also for the extremes. Some examples of such studies are listed in the following. Ensor and Robeson (2008) analyzed the statistics of precipitation in gridded products in the USA and found that the interpolation process significantly increased the number of low precipitation events while greatly reducing the intensity and frequency of heavy precipitation episodes. Prein and Gobiet (2016) evaluated many regional datasets in Europe and demonstrated that the differences among them have the same magnitude as the precipitation errors found in regional climate models. Flaounas et al. (2012) validated the E-OBS dataset in the Mediterranean region and found that those data must be used with care in complex environments such as coastal areas and mountainous regions. Turco et al. (2013b) compared three different gridded datasets in the greater Alpine region (E-OBS and two local grids) and found that specific regional products are often more trustworthy than E-OBS. Herrera et al. (2012) compared Spain02 and E-OBS in Spain and found that the station density severely affects the results. Turco and Llasat (2011) confirmed the high quality of the Spain02 dataset in northeastern Spain. Belo-Pereira et al. (2011) compared IB02, an Iberian daily precipitation dataset built by joining two methodologically equivalent gridded products for Portugal (PT02) and Spain (Spain02 v2), to global datasets and found that the global products produce better results in western Iberia than on the Mediterranean side. Finally, Katsanos et al. (2016) validated the CHIRPS global remote-sensing product (Funk et al., 2015) in Cyprus and compared it to E-OBS and local rain gauge data and found that at the monthly scale, the results showed a good correlation between the CHIRPS values and the recorded precipitation.

The objective of this paper is 2-fold. Firstly, this paper validates SAFRAN's precipitation as reproduced in the most recent Spanish SAFRAN dataset (Quintana-Seguí, 2015), which is a temporal and spatial extension of the dataset presented in Quintana-Seguí et al. (2016). The new dataset spans a period of 35 years (1979/1980–2013/2014) and geographical covers mainland Spain and the Balearic Islands.



**Figure 1.** Maps of the study area. The left panel shows the relief (using the same 5 km resolution grid that SAFRAN uses) and the main river basins in Spain (note that for the sake of simplification, some of the smaller river basins have been grouped). The right panel shows SAFRAN's climatically homogeneous zones, the rain gauges used for the study and their mean yearly precipitation. The dependent and independent stations are shown as dots and triangles, respectively.

Secondly, this paper compares the SAFRAN precipitation dataset with Spain02, which is a well-known, high-quality daily precipitation dataset for Spain, and ERA-Interim, the ECMWF's reanalysis dataset.

ERA-Interim is not the same kind of dataset as SAFRAN or Spain02, in the sense that it is a reanalysis product into which gauge observations are not assimilated. Precipitation in ERA-Interim is the direct result of a model forecast, even though the dynamics conducive to it are indeed constrained by other assimilated observational data in the reanalysis procedure. Nevertheless, due to its high quality, ERA-Interim is of special interest in hydrology because it is the basis for other derived products that are tailored for hydrological applications such as WFDEI (Weedon et al., 2014). Moreover, ERA-Interim is a dataset suitable to force LSMs, so it is introduced in the comparison to show how a low-resolution global product compares to local higher resolution ones, which is one of the objectives of the earth2Observe project and thus of this paper. E-OBS has not been included for several reasons. On the one hand, E-OBS can be considered a different kind of product in comparison to SAFRAN and Spain02 due to its much lower density of stations and the lower resolution of the final product. On the other hand, the E-OBS station density is especially low in Spain; as a consequence, this dataset cannot correctly reproduce the Iberian precipitation regimes, which primarily affects the variability and the extremes of the precipitation events; these are well-known problems reflected in previous studies (e.g., Herrera et al., 2012). Finally, in contrast to ERA-Interim, it is not suitable to force LSMs, which is the main objective of SAFRAN.

## 2 Study area

The geographical domain of this study is peninsular Spain and the Balearic Islands, located in southwestern Europe. Two main factors modulate the climate in this area: the in-

fluence of both the Atlantic Ocean and the Mediterranean Sea, and the complex orography (Fig. 1a). On the one hand, frontal systems from the Atlantic Ocean swipe the region regularly bringing wet and cold air, whilst wet and warm air arrives from the Mediterranean Sea. On the other hand, the complex orography and the configuration of the mountain ranges drive the wet air masses distributing the precipitation across the region. The combination of both factors leads to a marked NW–SE precipitation gradient ranging from large amounts of precipitation (from 900 to 2500 mm yr<sup>-1</sup>) without a dry season on the North Atlantic coast to semiarid and desert regions in the southeast with less than 100 mm yr<sup>-1</sup> concentrated in some severe events; there are variants of Atlantic and Mediterranean climates in between (AEMET, 2011).

The impact of a surrounding water mass and a complex orography is particularly relevant in the Ebro valley in the northeast of Spain. With both Atlantic and Mediterranean influences, the region has a low mean precipitation that is caused by the shadowing effect of the surrounding orography. This is an example of the important role of Spain's marked relief on the distribution of precipitation, enhancing it in some areas and decreasing it in others, which in turn has an important effect on runoff generation. This is an added difficulty for gridded precipitation products. Figure 1a shows the study area, its relief and the main river basins.

## 3 Datasets and methods

### 3.1 SAFRAN

The SAFRAN (Système d'Analyse Fournissant des Renseignements Atmosphériques à la Neige) meteorological analysis system (Durand et al., 1993, 1999) was initially created with the objective of forcing the CROCUS snow model (Brun et al., 1989) in the French Alps to improve avalanche

prediction. Currently, it is extensively used in a wide range of applications. SAFRAN uses an optimal interpolation algorithm (OI) (Gandin, 1966) that combines observations and a first guess (e.g., the outputs of a meteorological model) to produce a gridded dataset of precipitation, temperature, wind speed, relative humidity and cloudiness. SAFRAN is also able to calculate downward visible and infrared radiation (Ritter and Geleyn, 1992).

SAFRAN is currently used for operational and research purposes in France (Quintana-Seguí et al., 2008; Vidal et al., 2010; Habets et al., 2008). It is also used for research purposes in Spain (Quintana-Seguí et al., 2016) and Morocco (Szczypta et al., 2015). In Spain, SAFRAN was first implemented and tested in a pilot study for the northeastern Iberian Peninsula using only 1 year of data (Quintana-Seguí et al., 2016). This study is based on the recent extension of SAFRAN to the whole of Spain over a period of 35 years (1979/1980–2013/2014) (Quintana-Seguí, 2015).

One of the main characteristics of SAFRAN is the use of climatically homogeneous zones to divide the space into the analysis areas shown in Fig. 1b. These zones have irregular shapes and cover an area that is generally smaller than 1000 km<sup>2</sup>. Ideally, they should have no horizontal gradient because, within the zone, SAFRAN cannot reproduce it. As it is impossible to design zones without horizontal gradients, the design of the zones should minimize them as much as possible, especially for precipitation. The zones have several vertical levels spaced at 300 m. The lowest level is on the ground; the other levels are located at 300 m, 600 m etc. (provided that they are above the ground). The highest level of each zone is the first level above the highest point of the relief. SAFRAN estimates the value of the analyzed variable on each level. These values are subsequently horizontally interpolated to the regular grid, which is how the data are presented to the users. In our implementation, the grid has a resolution of 5 km (Fig. 1a shows the relief of Spain using the exact same grid as SAFRAN). Each grid point has a location and an altitude according to the relief and belongs to a given zone (each zone has, on average, 40 grid points). To interpolate the variable to the altitude of the grid point, the values of the two adjacent vertical levels of the zone are used. A consequence is that, even though the zones are initially homogeneous, the grid points within a zone have different values if they are at different altitudes. By means of the vertical levels, SAFRAN accounts for variations in precipitation with altitude using the real gradients at the time of analysis. A downside of the climatically homogeneous zones is that they create artificial discontinuities at the borders of the zones. For precipitation, SAFRAN analyzes daily observations, and all other variables are analyzed every 6 h. The data are then time interpolated to the hourly scale using different methods for each variable. For precipitation, relative humidity is currently used to hourly distribute the daily precipitation. For each analysis, SAFRAN uses as much data as possible after performing a quality control based on an iterative procedure

that compares the observed and analyzed quantities at the observation location. The number of stations used in the analysis thus changes with time; this optimizes the quality of the daily precipitation but makes the dataset untrustworthy for trend analysis. We note that in this paper we do not study the hourly distribution of the precipitation, but only the quality of the daily data. For more details on SAFRAN, please see Quintana-Seguí et al. (2008, 2016).

For this study, SAFRAN has been extended to mainland Spain and the Balearic Islands; for this purpose, a new set of climatological zones has been designed. The new zones, which are shown in Fig. 1b, also cover Portugal for a possible future extension. Quintana-Seguí et al. (2016) tested two methods to define the zones, one based on the contours of the river basins and the other based on the meteorological alert zones used by the Spanish Meteorological State Agency (Agencia Estatal de Meteorología; AEMET). The results showed that using the meteorological alert zones worked slightly better, but the differences were small. In other words, SAFRAN is quite robust regardless of the zone map used, provided it has physical sense and the zones have the right sizes. In order to expand the zone map to the whole of mainland Spain and the Balearic Islands, we first tried to use the same meteorological alert zones, but found that, in general, the zones were too big. As a consequence, we decided to create new zones following a subjective method based on the meteorological alert zones, the basin limits and our expert knowledge. In flat areas, the divisions created were more subjective than on the relief because there is no clear objective climatic delimitation of the terrain. However, it is precisely in these flat areas where SAFRAN is less sensitive to the shape of the zones, as the horizontal gradients are very weak.

In addition, this newer version of SAFRAN uses ERA-Interim as a first guess for most variables, excluding precipitation (the variable analyzed in this paper), for which the first guess is deduced from the observations. The ERA-Interim dataset is available from 1979, which determines that the start date of SAFRAN be the same. The meteorological station data come from the AEMET network (see Sect. 3.5). Our dataset finishes in August 2014, which is the date that the data were requested from AEMET for this study.

### 3.2 Spain02

Spain02 (Herrera et al., 2012, 2016) is a series of high-resolution daily precipitation and temperature gridded datasets developed for peninsular Spain and the Balearic Islands. Version 4 of the product has been produced using the standard EURO-CORDEX (European Coordinated Regional Climate Downscaling Experiment) grids at resolutions of 0.44, 0.22 and 0.11° and different interpolation approaches (Herrera et al., 2016). To compare Spain02 and SAFRAN, we chose the AA-3D version for this work, which considers the orography to be a covariable (3D) in the in-

terpolation process. It is areally representative (AA); more specifically, the AA-3D version is constructed by calculating the areal average of the interpolation, which is performed by applying thin plate splines (TPS) with the orography at the monthly scale and ordinary kriging (OK) on the daily anomalies. In fact, this is the same method employed by the widely used European-scale product E-OBS (Haylock et al., 2008). As a result, both SAFRAN and the version considered for Spain02 are representative of the areas and account for the relief in the interpolation process. In addition, Spain02 presents the same problems for reproducing trends as described for SAFRAN due to the station network considered (2756 rain gauges), which prioritizes spatial density over temporal consistency.

We will use Spain02, or SP02, to refer to the AA-3D version of Spain02, which has a resolution of  $0.11^\circ$ .

### 3.3 ERA-Interim

ERA-Interim (Dee et al., 2011) is the current global atmospheric reanalysis produced by the European Centre for Medium-Range Weather Forecasts (ECMWF). It covers the period from 1979 onward. ERA-Interim has a spatial resolution of  $\sim 79$  km (T255).

### 3.4 Time periods

Figure 2 shows the time periods covered by the SAFRAN and the Spain02 datasets. The latter covers the period from 1971 to 2010, whereas the current 35-year SAFRAN dataset has been produced within two projects and can be divided into three subperiods, as shown in Fig. 2. The central subperiod (labeled SAFRAN1 in the figure) covers the hydrological years (a hydrological year starts on 1 September and finishes on 31 August of the next calendar year) from 1995/1996 to 2006/2007. The other subperiod (SAFRAN2) was developed later in another research project and includes the remaining years in the series between 1979/1980 and 2013/2014.

For each of the SAFRAN subperiods, a set of randomly selected stations was used a set of randomly selected stations was excluded from the analysis in order to obtain an independent validation dataset. The excluded stations were selected randomly, but making sure that the spatial separation between the selected stations was larger than a minimum distance in order to guarantee an even spatial distribution of the validation stations. Unfortunately for the SAFRAN1 and SAFRAN2 subperiods, the same validation datasets were not considered, and a common set of independent stations for the whole 35-year period is thus not available for SAFRAN. Furthermore, Spain02 also has its own set of independent stations. The common set of stations that are independent for both SAFRAN and Spain02 is not useful for the validation because there are too few stations that are poorly spatially distributed. Nevertheless, Spain02 has already been validated

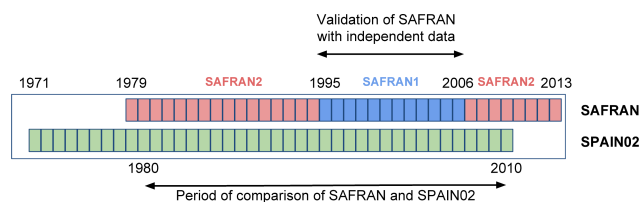


Figure 2. The data availability for SAFRAN and Spain02.

with independent data by Herrera et al. (2012, 2016); thus, in this study, the comparisons between SAFRAN and Spain02 are performed using dependent data.

As a consequence, in this study

1. SAFRAN is validated with independent data for the 1995/1996–2006/2007 period (SAFRAN1),
2. and the other comparisons, which also included Spain02 and ERA-Interim, are performed over a common period (1980–2010; natural years) using dependent stations.

### 3.5 Station data

Figure 1b shows the locations of the daily precipitation stations used for this study, which were provided by the Spanish Meteorological State Agency (Agencia Estatal de Meteorología, AEMET). The figure only depicts the 1237 stations for which the time series have at least 90 % of the data, which are the stations that were used for our analysis. The triangles correspond to the independent dataset used to validate SAFRAN for the 1995/1996–2006/2007 time period (249 stations). The smaller dots correspond to the remaining stations, which are used for all other validations (988 stations).

### 3.6 Comparison measures

The methodology used in this paper is based on the approach used by Turco et al. (2013b) with some modifications to account for the hydrological context in which the SAFRAN dataset has been developed.

All of the metrics used in this study compare the products with rain gauge data, considering for each station the nearest grid points in terms of the Euclidean distance. The dual nature of precipitation in terms of its occurrence and amount makes it necessary to explicitly consider both components in the analysis and to include appropriated validation measures (e.g., the ROC skill area) and indicators (e.g., the wet-day frequency). First, the temporal agreement between the time series and the binary (the occurrence or non-occurrence of precipitation) sequences were studied using the relative mean absolute error (MAEr) and the Pearson's correlation (CORR), and the relative operating characteristic (ROC) skill area (RSA) and its components (HIR: hit rate; FAR: false alarm rate; and CAR: correct alarm rate), respectively. In a second stage, the spatial agreement is studied

**Table 1.** The climatic mean and extreme indices for the precipitation used in this work based on ETCCDI (Expert Team on Climate Change Detection and Indices; <http://ccma.seos.uvic.ca/ETCCDI>).

Label	Description	Units
PRCPTOT	Total precipitation	mm
R1	Number of days with precipitation over 1 mm day <sup>-1</sup> (i.e., wet days)	day
SDII	Mean precipitation on a wet day	mm
R10	Number of days with precipitation over 10 mm day <sup>-1</sup>	day
R20	Number of days with precipitation over 20 mm day <sup>-1</sup>	day
RX1D	Maximum precipitation in 1 day	mm
RX5D	Maximum precipitation in 5 days	mm
CWD	Consecutive wet days (> 1 mm)	day
CDD	Consecutive dry days (< 1 mm)	day

**Table 2.** The annual and seasonal correlation (CORR), relative mean absolute error (MAEr) and ROC skill area (RSA) of SAFRAN compared to the independent (ind.) and dependent (dep.) observations for the period 1995/96–2005/06.

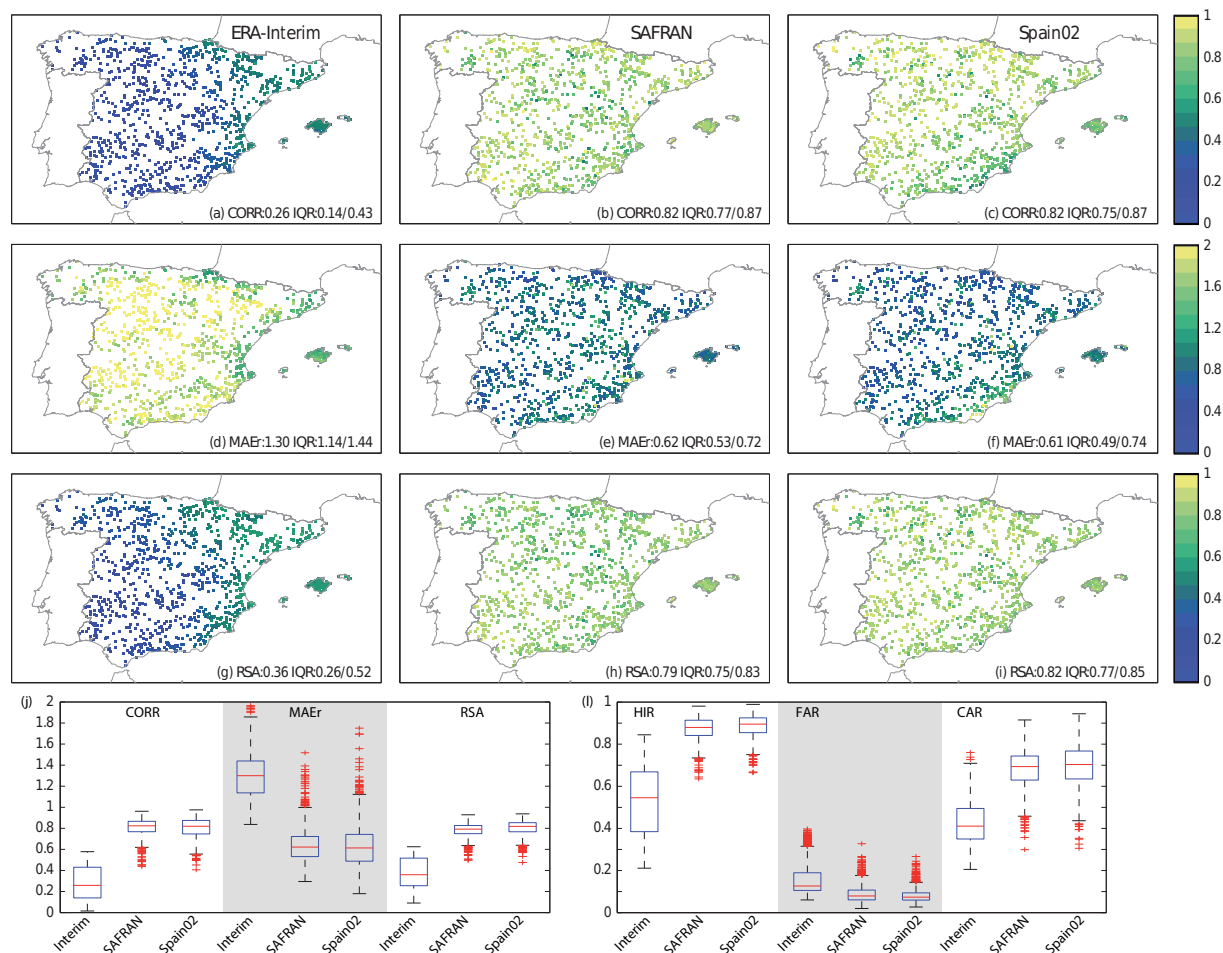
	CORR			MAEr			RSA		
	Mean	Q <sub>25</sub>	Q <sub>75</sub>	Mean	Q <sub>25</sub>	Q <sub>75</sub>	Mean	Q <sub>25</sub>	Q <sub>75</sub>
Annual									
Ind. obs.	0.82	0.75	0.86	0.62	0.54	0.75	0.77	0.73	0.81
Dep. obs.	0.82	0.77	0.87	0.62	0.53	0.72	0.79	0.75	0.83
Winter									
Ind. obs.	0.84	0.77	0.90	0.57	0.48	0.71	0.78	0.72	0.82
Dep. obs.	0.85	0.79	0.89	0.58	0.48	0.69	0.79	0.74	0.83
Spring									
Ind. obs.	0.82	0.76	0.86	0.62	0.54	0.74	0.77	0.73	0.80
Dep. obs.	0.83	0.78	0.87	0.61	0.53	0.71	0.78	0.74	0.82
Summer									
Ind. obs.	0.71	0.62	0.79	0.88	0.74	1.05	0.70	0.64	0.77
Dep. obs.	0.74	0.66	0.80	0.82	0.71	0.94	0.75	0.70	0.80
Autumn									
Ind. obs.	0.83	0.76	0.87	0.60	0.51	0.71	0.78	0.73	0.81
Dep. obs.	0.83	0.77	0.88	0.60	0.51	0.71	0.79	0.75	0.83

using the standard precipitation indicators shown in Table 1. All indices are calculated on an annual scale in the common period (see Sect. 3.4), although some indices have also been calculated for each season (PRCPTOT, R1 and RX1D) or at a monthly scale (PRCPTOT and RX1D) for each river basin (see the river basin map in Fig. 1). Finally, the spatial pattern obtained for each index and gridded dataset is compared to the observations using the correlation (CORR), the variability (relative standard deviation, or STDEVr), the relative centered root mean square error (CRMSEr) and the relative bias (BIASr). To make the different indices comparable, the “relative” statistics (STDEVr and CRMSEr) have been normalized by dividing both the CRMSEr and STDEVr by the standard deviation of the observations (the reference dataset).

Therefore, an STDEVr of 1 means that the standard deviation of the product is the same as that of the observations. The BIASr is relative with respect to the mean of the observations. The Taylor diagram (Taylor, 2001) summarizes these measures on a single plot, which has been produced to show the seasonal statistics.

Note that there are differences between the spatial representativity of the considered datasets, ranging from local (stations) to low-resolution ( $\sim 79 \text{ km}^2$ ) areal averages (ERA-Interim). These differences have been taken into account by performing, on the one hand, a comparison among the validation scores of the different gridded datasets with respect to rain gauges and, on the other, by defining the global scores based on the spatial averages over large enough areas





**Figure 3.** The correlation (a–c), relative mean absolute error (d–f) and ROC skill area (g–i) of ERA-Interim, SAFRAN and Spain02 with regard to the observations. The lower row shows the box plots of the (j) CORR, MAEr, and RSA as well as the (l) HIR, FAR, and CAR for the three datasets, reflecting the spatial variability of the scores.

(e.g., river basins). Finally, the aforementioned problem has been considered in the following sections in the interpretation and discussion of the results.

## 4 Results

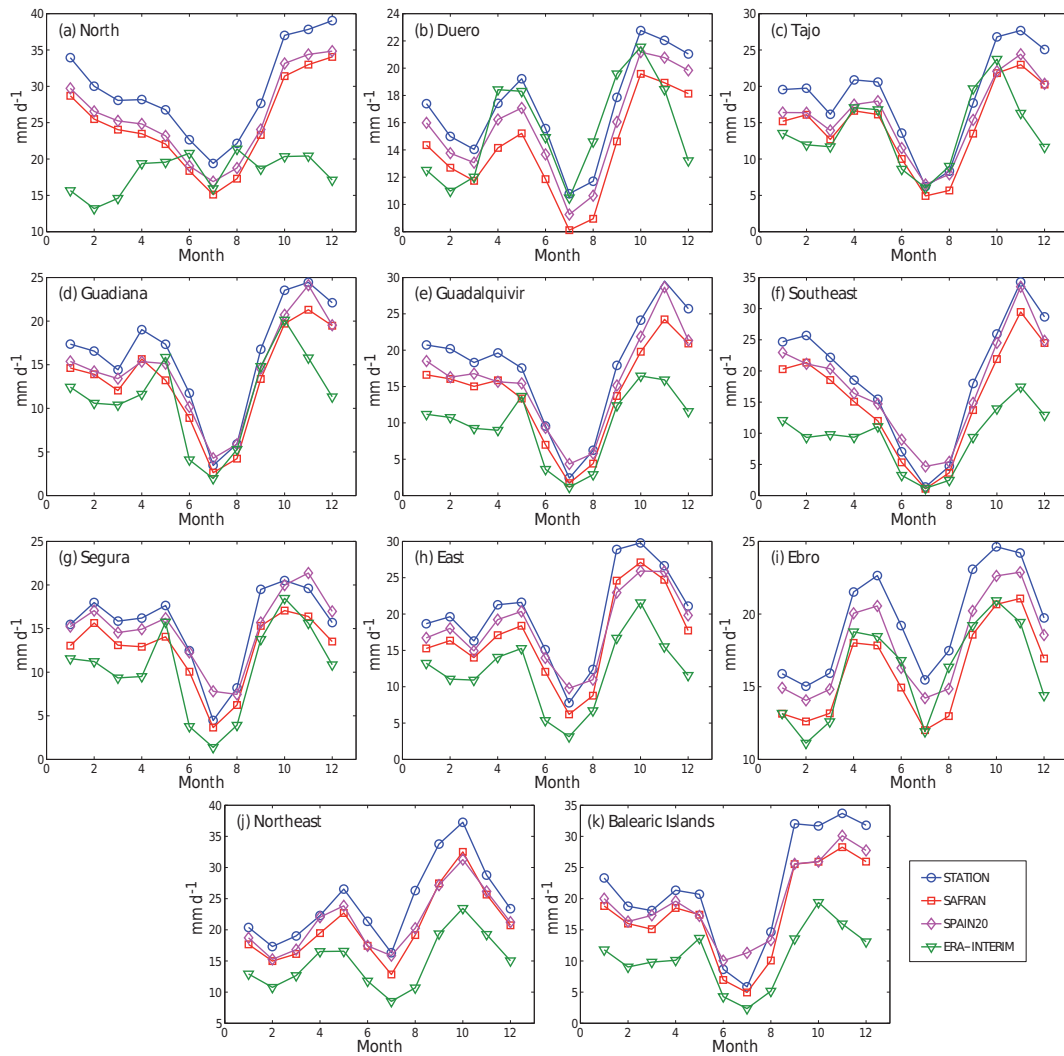
### 4.1 Validation of SAFRAN using independent data

The first step in our analysis is to validate the new SAFRAN dataset for Spain using independent stations for the period 1995/1996–2006/2007 (SAFRAN1). Table 2 summarizes the results at the annual and the seasonal scale, showing that the CORR, MAEr and RSA are slightly degraded when independent stations are employed (not used to perform the analysis) but have nevertheless a rather similar mean and interquartile range, which is an indicator of the robustness of SAFRAN. SAFRAN is quite robust across seasons with summer as the season with the worst performance.

### 4.2 Evaluation of the time similarity

The next step in the analysis is to compare the three studied products (SAFRAN, Spain02 and ERA-Interim) to a common set of observations considering the same period (1980 to 2010).

Figure 3 shows the correlation (CORR), relative mean absolute error (MAEr) and ROC skill area (RSA) between the three products and the observations. The SAFRAN and Spain02 correlations are very similar, and the main difference is located on the southeastern coast where SAFRAN has larger correlations. This similarity is also reflected in the statistics in Table 3, with both products having the same mean annual correlation (0.82) and an almost identical interquartile range (approximately 0.1). As expected, the ERA-Interim correlation is lower. In this case, there is a strong east–west gradient with the correlations higher in the Mediterranean due to the binary component (rain or no rain) of the precipitation, which is moderately well represented in



**Figure 4.** The annual cycle of maximum precipitation in 1 day (RX1D) for the Spanish river basins. The y axis range is different for each figure to highlight the differences between the datasets.

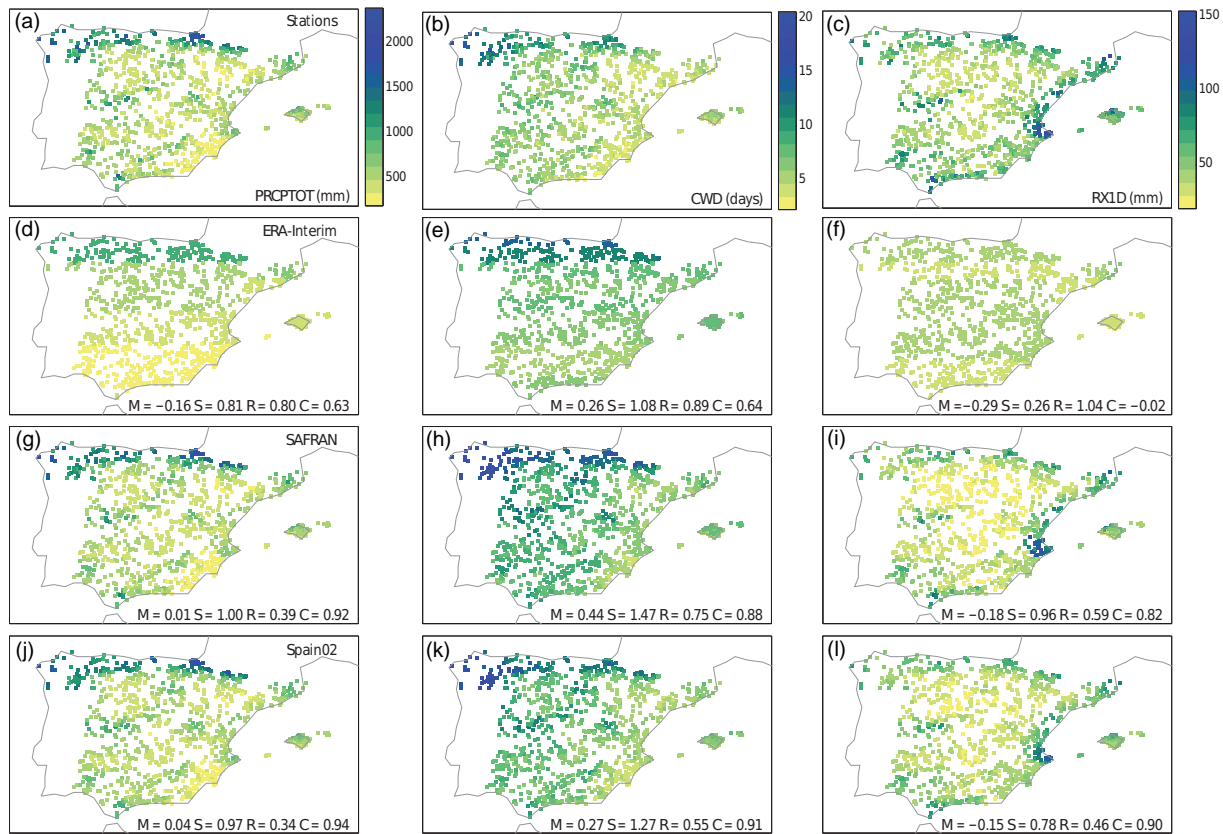
the reanalysis. Precipitation is a variable with many occurrences of zero, which may suggest the use of the Spearman’s rank correlation coefficient instead of Pearson’s. In order to verify that both measures give the same qualitative information, we calculated the Spearman’s rank correlation coefficient and found that it is lower for both SAFRAN (0.73) and Spain02 (0.75). The corresponding interquartile ranges are 0.68 to 0.78 for SAFRAN and 0.69 to 0.80 for Spain02. Therefore, all results are shifted when the Spearman’s coefficient is used instead of Pearson’s. Notwithstanding, the width of the interquartile range remains almost the same for SAFRAN, and it is narrower by only 0.01 for Spain02; thus, we conclude that both Spearman’s and Pearson’s provide the same outlook when used to assess the relative performance of either product with respect to the observations.

The results obtained with the correlation are, in general, in agreement with those obtained for the RSA. However, for the

latter score, Spain02 shows a more homogeneous spatial pattern with similar values on the southeastern coast and in the rest of the area. For ERA-Interim, the spatial pattern shown in the correlation is extended for the RSA, with an improvement in the Ebro valley and on the southeastern coast.

Concerning the MAEr, again SAFRAN and Spain02 show very similar results, as Table 3 corroborates. Spain02’s larger number of errors are mainly concentrated in the southeast, as seen in the correlation, whereas SAFRAN shows a more random spatial pattern. In the case of Spain02, the affected region (close to Murcia) has the highest station density for the entire peninsula, so the interpolation process obtains an areally averaged precipitation value that differs from the local observations of each particular station. This is more relevant when intense local precipitation events occur, which are very frequent in the Mediterranean region, leading to an overestimation of the interpolated values in the surrounding





**Figure 5.** The PRCPTOT (a, d, g, j), CWD (b, e, h, k) and RX1D (c, f, i, l) for the stations (a–c), ERA-Interim (d–f), SAFRAN (g–i) and Spain02 (j–l).

area. The more spatially distributed nature of SAFRAN’s errors points to a close relation with the position of the zone borders, which produces a homogeneous spatial distribution over the study area. ERA-Interim’s errors are more numerous and quite spatially homogeneous, with the exception of the Mediterranean coast and an area located in the Basque Country (on the eastern side of the Bay of Biscay).

The box plots shown in Fig. 3j and l corroborate the similarity between SAFRAN and Spain02, not only in terms of the mean but also in the spatial variability of the scores. They also corroborate the worse results and the higher spatial variability in terms of the box altitude for ERA-Interim. Note that there is remarkable agreement between SAFRAN and Spain02 on all scores related to the binary sequence. In addition to the correlation, this reflects the very similar temporal structure of both datasets.

Table 3 also shows the results by season, clearly indicating that the seasonal behavior of SAFRAN and Spain02 is the same. They are both very robust across seasons with the exception of summer when the scores are degraded, especially for CORR and MAEr.

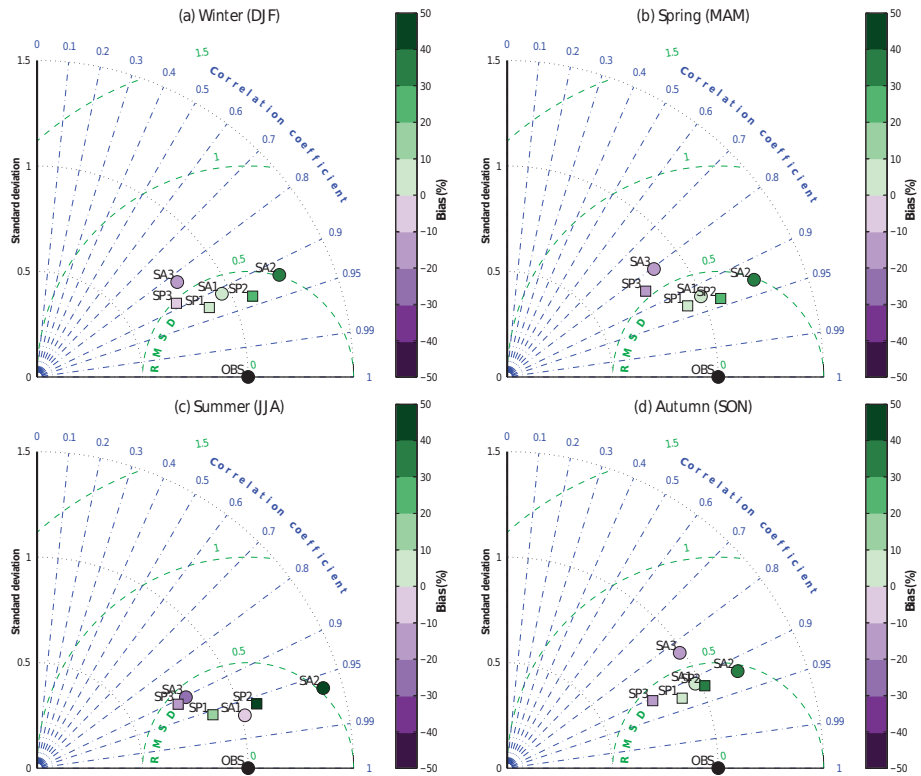
Both SAFRAN and Spain02 (the AA-3D version used in this paper) consider the relief in their algorithms, and the same statistics were thus calculated for a subset of high-

altitude stations located at 1000 m or higher. The results, which are shown in Table 3, do not reflect significant differences in terms of the correlation and MAEr between high- and low-altitude stations. Both SAFRAN and Spain02 perform equally well with only some slight differences shown for the RSA, which are more relevant for ERA-Interim. The trend in the errors with altitude for all stations was calculated, and it was not significant in any case, even for ERA-Interim. However, this result does not account for the fact that high-altitude areas are not well represented by the observations because there is a lack of stations at the highest elevations.

**Evaluation of the spatial similarity by means of precipitation indices**

In this section, we evaluate how the different analyzed products are able to reproduce a standard set of precipitation indices. First, the annual cycles of the monthly mean and the daily maximum precipitation are studied for the 11 river basins shown in Fig. 1. The global statistics of each index are then summarized and analyzed.

Figure 4 shows the annual cycle of the maximum daily precipitation (RX1D) for each of the main Spanish river basins. A first look at the figure shows that, as expected, all



**Figure 6.** Taylor diagrams for the seasonal precipitation climatology over Spain. The points most similar to the station data are closest to the point indicated as OBS. The squares labeled SP are used for Spain02, whereas the circles labeled SA are for SAFRAN. The colors indicate the bias (in percentage with respect to the station data) of the dataset. The numbers correspond to the different indices: 1 is PRCPTOT, 2 is R1 and 3 is RX1D.

products underestimate this index and that ERA-Interim is the dataset with the lowest RX1D values for most basins. Another interesting result is that Spain02 is closer to the observations than SAFRAN, although they are generally close to each other. Despite a general underestimation of RX1D, in some cases ERA-Interim is closer to the observations (e.g., Duero) than SAFRAN and Spain02 or has higher values than either gridded dataset for some months. This is probably due to the relief surrounding these basins (see Fig. 1), which creates a strong rain shadow effect on the leeward side that cannot be reproduced by ERA-Interim because of the coarse resolution.

Concerning the monthly mean (not shown), SAFRAN and Spain02 are closer, with the latter overestimating the monthly precipitation in the east, southeast and Segura basins, especially in autumn; this is in agreement with the results described in the previous section.

Figure 5 shows how both SAFRAN and Spain02 are able to reproduce the spatial patterns and magnitudes for the three indices, although they mostly overestimate the CWD in the high-relief areas and underestimate RX1D, which is a well-known effect of the interpolation process. ERA-Interim misses the effects of the relief with a strong north–south gradient for the PRCPTOT and CWD indices and a homoge-

neous spatial pattern in almost all of the study area for RX1D. ERA-Interim is unable to reproduce the intensity of the precipitation, which leads to very low values for both PRCPTOT and RX1D.

In addition to Fig. 5, Table 4 summarizes the scores of the three products for the indices defined in Table 1. Both SAFRAN and Spain02 obtain correlations greater than 0.75 with comparable values, whereas ERA-Interim is unable to reproduce the spatial pattern of most of the indices with several correlations below 0.65. In general, the three datasets nevertheless underestimate the indices related to the magnitude and intensity of the precipitation (PRCPTOT, SDII, RX1D and RX5D), whereas they overestimate R10, R20 and the frequency of wet days. In addition, there is a common tendency to overestimate CWD and underestimate CDD. There are some exceptions, however, such as the slight overestimation of SAFRAN and Spain02 of PRCPTOT or the overestimation of the CDD index given by SAFRAN. Finally, in terms of variability, there is a general underestimation for the PRCPTOT, SDII, R20, RX1D and RX5D indices, which is as high as 30 % for SDII. There is an overestimation for the rest, mainly for the CWD index.

The scores shown in Table 4 are summarized in the Taylor diagram. Figure 6 shows a summary of the seasonal statistics

**Table 3.** The correlation (CORR), relative mean absolute error (MAEr) and ROC skill area (RSA) of the three compared products considering all stations (the upper half of the table) and those located at altitudes higher than 1000 m (the lower half of the table).

	Correlation			MAEr			RSA		
	Mean	Q <sub>25</sub>	Q <sub>75</sub>	Mean	Q <sub>25</sub>	Q <sub>75</sub>	Mean	Q <sub>25</sub>	Q <sub>75</sub>
Annual									
ERA	0.26	0.14	0.43	1.30	1.14	1.44	0.36	0.26	0.52
SFR	0.82	0.77	0.87	0.62	0.53	0.72	0.79	0.75	0.83
SP02	0.82	0.75	0.87	0.62	0.49	0.75	0.82	0.77	0.85
Annual (> 1000 m)									
ERA	0.21	0.04	0.36	1.23	1.11	1.43	0.33	0.27	0.40
SFR	0.82	0.75	0.86	0.64	0.55	0.77	0.77	0.72	0.81
SP02	0.82	0.75	0.88	0.61	0.49	0.76	0.80	0.73	0.85
Winter									
ERA	0.32	0.15	0.51	1.27	1.01	1.34	0.37	0.24	0.56
SFR	0.85	0.79	0.89	0.58	0.48	0.69	0.79	0.74	0.83
SP02	0.83	0.76	0.89	0.58	0.45	0.70	0.81	0.76	0.85
Spring									
ERA	0.30	0.16	0.48	1.21	1.06	1.34	0.36	0.25	0.48
SFR	0.83	0.78	0.87	0.61	0.53	0.71	0.78	0.74	0.82
SP02	0.82	0.76	0.88	0.59	0.47	0.71	0.82	0.77	0.86
Summer									
ERA	0.17	0.10	0.26	1.52	1.30	1.87	0.27	0.17	0.37
SFR	0.74	0.66	0.80	0.82	0.71	0.94	0.75	0.70	0.80
SP02	0.72	0.57	0.81	0.89	0.67	1.30	0.80	0.73	0.85
Autumn									
ERA	0.23	0.10	0.40	1.37	1.13	1.55	0.36	0.25	0.53
SFR	0.83	0.77	0.88	0.60	0.51	0.71	0.79	0.75	0.83
SP02	0.84	0.78	0.89	0.57	0.45	0.71	0.81	0.76	0.85

of PRCPTOT, R1 and RX1D. In agreement with the results shown in the table, a general pattern in all four diagrams is that SAFRAN and Spain02 are generally close to each other. In addition, the figures show that, in general, the results are quite robust across seasons. The only exceptions are R1 in summer when the distance between SAFRAN and Spain02 increases in favor of the latter, and in autumn when the distance of RX1D increases, also in favor of Spain02. Another interesting difference is that in summer, the sign of the bias of PRCPTOT is different for SAFRAN and Spain02. The former has a negative bias, and the latter has a positive one.

### 5 Discussion

A general analysis of the results shows that SAFRAN and Spain02 have very similar scores, although the latter slightly surpasses the former. The similarities are very consistent across the scores used and the seasons analyzed. This means

that in Spain, SAFRAN’s daily precipitation, for which the algorithm was designed in the late 20th century, is very close to the precipitation estimated by more recent and more specialized state-of-the-art products. As expected, SAFRAN and Spain02 have better scores than ERA-Interim. The most important limitation of the ERA-Interim reanalysis is its inability to capture the effects of the relief due to its low resolution, as noted in previous studies (e.g., Belo-Pereira et al., 2011), and to reproduce the temporal structure of the precipitation in terms of both the occurrence and the amount. However, ERA-Interim reproduces length spells remarkably well (although not timely) in contrast to the difficulties of high-resolution products.

The high-resolution gridded products overestimate the number of precipitation days. This is probably caused by the interpolation method and affects SAFRAN more than Spain02, likely because the former uses climatically homogeneous zones. SAFRAN generates precipitation for a whole zone once there is a localized precipitation event in some

**Table 4.** A comparison of the performance of ERA-Interim (ERA), SAFRAN (SFR) and Spain02 (SP02) in reproducing the indices of the mean and extreme precipitation.

	BIASr	STDEvr	CRMSEr	CORR
Total precipitation (PRCPTOT)				
ERA	-0.16	0.81	0.80	0.63
SFR	0.01	1.00	0.39	0.92
SP02	0.04	0.97	0.34	0.94
Consecutive dry days (CDD)				
ERA	-0.01	1.28	0.69	0.85
SFR	0.04	1.04	0.49	0.88
SP02	-0.03	0.91	0.50	0.87
Consecutive wet days (CWD)				
ERA	0.26	1.08	0.89	0.64
SFR	0.44	1.47	0.75	0.88
SP02	0.27	1.27	0.55	0.91
Number of rainy days (R1)				
ERA	0.39	1.66	1.06	0.79
SFR	0.40	1.27	0.50	0.93
SP02	0.30	1.18	0.41	0.94
Mean precipitation on a wet day (SDII)				
ERA	-0.40	0.16	1.03	-0.11
SFR	-0.26	0.68	0.66	0.76
SP02	-0.18	0.67	0.59	0.82
Number of days of $P > 10$ mm (R10)				
ERA	-0.31	0.71	0.78	0.64
SFR	-0.04	1.08	0.40	0.93
SP02	0.05	1.06	0.34	0.95
Number of days of $P > 20$ mm (R20)				
ERA	-0.57	0.25	0.94	0.38
SFR	-0.22	0.94	0.46	0.89
SP02	-0.09	0.94	0.40	0.92
Max. precip in 1 day (RX1D)				
ERA	-0.29	0.26	1.04	-0.02
SFR	-0.18	0.96	0.59	0.82
SP02	-0.15	0.78	0.46	0.90
Max. precip in 5 days (RX5D)				
ERA	-0.27	0.26	1.01	0.10
SFR	-0.07	0.94	0.54	0.84
SP02	-0.06	0.85	0.42	0.91

of the stations within, thereby wrongly assigning the event to unaffected stations. After many occurrences, this process causes an overestimation of the number of wet days for all stations. Similarly, localized high precipitation events may be missed by SAFRAN because it tends to average the values

of all stations in a zone. SAFRAN may also completely miss events in zones with few stations; this is in agreement with Ensor and Robeson (2008). The overestimation of low precipitation events and the underestimation of intense episodes is likely to have consequences when the data are used to force a land surface or hydrological model. In the future, it will be necessary to investigate how the aforementioned problems affect the quality of LSM and hydrological simulations.

The comparison has shown a known problem of SAFRAN, which is the higher number of errors often found at the borders of the zones (Quintana-Seguí et al., 2008, 2016). However, Spain02 also has bugs that are induced by the algorithm, as shown by the higher number of errors in the Murcia region (southeast). This shows that each algorithm has its own set of limitations, which need to be well documented and explained to the users of the data.

The comparison between the results obtained using dependent and independent data shows that the procedure used to build SAFRAN is robust. Part of this robustness is due to SAFRAN's algorithm, but it is also due to the high density of the stations found in the area. Prein and Gobiet (2016) state that "the information content of a gridded dataset is mainly related to its underlying station density and not to its grid spacing". Because the station density in Spain is so high, the algorithms can leverage a lot of information, thereby producing similar results. Spain, which is a data-rich country, is thus a relatively easy target for such algorithms. Vidal et al. (2010) and Herrera et al. (2016) have shown that the station density has an impact on the resulting SAFRAN and Spain02 datasets. In a future study, it would be interesting to see how differently the results of SAFRAN and Spain02 degrade with station density.

Another interesting result concerning the robustness of the products is the similarity between the scores obtained from SAFRAN and Spain02 across seasons, with the exception of summer. This decrease in the performance in summer is expected due to the small scale of the systems that produce the precipitation during this season.

One limitation of this study is that there was no common set of independent stations for the different periods (see Sect. 3.4 and Fig. 2). As a consequence, the dependent station datasets used to perform the analysis are not exactly the same. At first glance, it might seem that this negatively affects the temporal homogeneity of the product; however, the approach in our study minimizes this effect. There is always a trade-off between the number of stations included in an analysis and the homogeneity of these stations. The approach used in SAFRAN favors station density, not temporal homogeneity; thus, if a station is missing 1 day, it will not be included in the analysis for that day, but it will be considered when there are data again. As a result, the loss of homogeneity resulting from not excluding the same independent stations for each period is diluted by the fact that the stations are not exactly the same day to day, together with the relatively small number of independent stations. The temporal

homogeneity is not the first priority of SAFRAN, which mitigates the effect of this methodological limitation.

Mountain areas are an exception because the station density in those areas is low and insufficient to reproduce the climatic variability of those regions. Our study has shown that the results of SAFRAN and Spain02 are robust with altitude, as the scores were similar for the stations situated at altitudes higher than 1000 m, but these results do not account for the fact that the mountain areas are not well observed. This has hydrological consequences because high-altitude areas are the generators of most basin runoff and thus of most of the water resources used in Spain. Further inspection of these aspects is necessary in the future. In addition, it is important to underline the fact that ERA-Interim is unable to reproduce the precipitation regimes of most basins due to its inability to reproduce the effects of the surrounding relief.

## 6 Conclusions

SAFRAN and Spain02 have very similar scores, although the latter generally slightly surpasses the former. As expected, SAFRAN and Spain02 are able to reproduce the main characteristics of the precipitation in Spain and perform better than ERA-Interim, which has difficulty capturing the effects of relief on precipitation due to its low resolution. SAFRAN is robust because the scores obtained using dependent and independent data are similar. Furthermore, both SAFRAN and Spain02 are also robust with altitude and across seasons (except in summer when the scores decrease for both products). ERA-Interim reproduces spells remarkably well in contrast to the low skill shown by the high-resolution products. The high-resolution gridded products overestimate the number of precipitation days; this affects SAFRAN more than Spain02 and is likely caused by the interpolation method. Both SAFRAN and Spain02 underestimate high precipitation events, but SAFRAN more so than Spain02.

*Data availability.* Four datasets were used for this paper. The SAFRAN dataset for Spain is available for research purposes from the Mistrals-HyMeX database (Quintana-Seguí, 2015). Spain02 v4 is freely distributed for research purposes at the Santander MetGroup THREDDS CATALOG (<http://meteo.unican.es/en/datasets/spain02>). ERA-Interim is distributed by the ECMWF on its own website (<http://www.ecmwf.int/en/research/climate-reanalysis/era-interim>). The observational data were obtained from AEMET using the standard forms available on their website (<https://sede.aemet.gob.es/>).

*Author contributions.* Pere Quintana-Seguí applied SAFRAN to Spain. He gathered the necessary data, obtained and modified the SAFRAN code, wrote the necessary scripts to run it properly, defined the climatically homogeneous zones and performed the product quality assessment. Marco Turco, Pere Quintana-Seguí and

Sixto Herrera designed the validation and comparison methodology. Marco Turco and Sixto Herrera performed the analysis and generated the figures that show the results. Pere Quintana-Seguí wrote most of the text and generated some figures. Marco Turco and Sixto Herrera provided feedback and improvements to the text. Sixto Herrera helped choose the right version of Spain02 for this comparison and also edited some figures for publication. Gonzalo Miguez-Macho provided feedback, which improved the experiment and the document, and helped with the language editing.

*Competing interests.* The authors declare that they have no conflict of interest.

*Acknowledgements.* We are grateful to the French National Centre for Meteorological Research (CNRM UMR3539, Météo-France CNRS) for allowing us to use the code of the SAFRAN analysis system for our studies, the Spanish State Meteorological Agency (AEMET) for sharing their very valuable observational data with us and the European Centre for Medium-Range Weather Forecasts (ECMWF) for making their ERA-Interim product openly available. This is a contribution to the FP7 earth2Observe project (<http://www.earth2observer.eu>), which received funding from the European Union's Seventh Programme for research, technological development and demonstration under grant agreement no. 603608. This work has been funded by the Spanish Economy and Competitiveness Ministry and the European Regional Development Fund through grant CGL2013-47261-R. This work has been supported by the Metropolitan Area of Barcelona Project (no. 308321; flood evolution in the metropolitan area of Barcelona from a holistic perspective: past, present and future) and the Spanish Project HOPE (CGL2014-52571-R) supported by the Ministry of Economy and Competitiveness. This work is a contribution to the HyMeX program (Hydrological cycle in the Mediterranean EXperiment; <http://www.hymex.org>).

Edited by: J. Seibert

Reviewed by: two anonymous referees

## References

- AEMET: Atlas climático ibérico/Iberian climate atlas, Agencia Estatal de Meteorología, Madrid, Spain, 2011.
- Artinyan, E., Habets, F., Noilhan, J., Ledoux, E., Dimitrov, D., Martin, E., and Le Moigne, P.: Modelling the water budget and the riverflows of the Maritsa basin in Bulgaria, *Hydrol. Earth Syst. Sci.*, 12, 21–37, doi:10.5194/hess-12-21-2008, 2008.
- Balsamo, G., Albergel, C., Beljaars, A., Boussetta, S., Brun, E., Cloke, H., Dee, D., Dutra, E., Pappenberger, F., de Rosnay, P., Muñoz-Sabater, J., Stockdale, F., and Vitart, F.: ERA-Interim/Land: A global land-surface reanalysis based on ERA-Interim meteorological forcing, Tech. rep., ECMWF, Reading, UK, 2012.
- Barbu, A. L., Calvet, J.-C., Mahfouf, J.-F., and Lafont, S.: Integrating ASCAT surface soil moisture and GEOV1 leaf area index into the SURFEX modelling platform: a land data assimilation

- application over France, *Hydrol. Earth Syst. Sci.*, 18, 173–192, doi:10.5194/hess-18-173-2014, 2014.
- Belo-Pereira, M., Dutra, E., and Viterbo, P.: Evaluation of global precipitation data sets over the Iberian Peninsula, *J. Geophys. Res.-Atmos.*, 116, 1–16, doi:10.1029/2010JD015481, 2011.
- Brun, E., Martin, E., Simon, V., Gendreau, C., and Coleou, C.: An energy and mass model of snow cover suitable for operational avalanche forecasting, *J. Glaciol.*, 35, 333–342, 1989.
- Cansado, A., Martín, C., Navascués, B., Martín, C., and Navascués, B.: Optimum interpolation new snow depth analysis in HIRLAM, *HIRLAM Newsletter*, 58–64, available at: [http://hirlam.org/index.php/component/docman/doc\\_download/198-hirlam-newsletter-no-43-paper-10-cansado-and-navascues-2003?Itemid=70](http://hirlam.org/index.php/component/docman/doc_download/198-hirlam-newsletter-no-43-paper-10-cansado-and-navascues-2003?Itemid=70) (last access: April 2017), 2003.
- Casanueva, A., Kotlarski, S., Herrera, S., Fernández, J., Gutiérrez, J. M., Boberg, F., Colette, A., Christensen, O. B., Goergen, K., Jacob, D., Keuler, K., Nikulin, G., Teichmann, C., and Vautard, R.: Daily precipitation statistics in a EURO-CORDEX RCM ensemble: added value of raw and bias-corrected high-resolution simulations, *Clim. Dynam.*, 47, 719–737, doi:10.1007/s00382-015-2865-x, 2015.
- Casanueva, A., Herrera, S., Fernández, J., and Gutiérrez, J.: Towards a fair comparison of statistical and dynamical downscaling in the framework of the EURO-CORDEX initiative, *Climatic Change*, 137, 411–426, doi:10.1007/s10584-016-1683-4, 2016.
- Chen, F., Manning, K. W., Lemone, M. A., Trier, S. B., Alfieri, J. G., Roberts, R., Tewari, M., Niyogi, D., Horst, T. W., Oncley, S. P., Basara, J. B., and Blanken, P. D.: Description and evaluation of the characteristics of the NCAR high-resolution land data assimilation system, *J. Appl. Meteorol. Clim.*, 46, 694–713, doi:10.1175/JAM2463.1, 2007.
- Cosgrove, B. A.: Real-time and retrospective forcing in the North American Land Data Assimilation System (NLDAS) project, *J. Geophys. Res.*, 108, 8842, doi:10.1029/2002JD003118, 2003.
- Decharme, B., Alkama, R., Papa, F., Faroux, S., Douville, H., and Prigent, C.: Global off-line evaluation of the ISBA-TRIP flood model, *Clim. Dynam.*, 38, 1389–1412, doi:10.1007/s00382-011-0154-9, 2012.
- Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, *Q. J. Roy. Meteorol. Soc.*, 137, 553–597, doi:10.1002/qj.828, 2011.
- Durand, Y., Brun, E., Merindol, L., Guyomarc'h, G., Lesaffre, B., and Martin, E.: A meteorological estimation of relevant parameters for snow models, *Ann. Glaciol.*, 18, 65–71, 1993.
- Durand, Y., Giraud, G., Brun, E., Merindol, L., and Martin, E.: A computer-based system simulating snowpack structures as a tool for regional avalanche forecasting, *J. Glaciol.*, 45, 469–484, 1999.
- Ensor, L. A. and Robeson, S. M.: Statistical Characteristics of Daily Precipitation: Comparisons of Gridded and Point Datasets, *J. Appl. Meteorol. Clim.*, 47, 2468–2476, 2008.
- Flaounas, E., Drobinski, P., Borga, M., Calvet, J.-C., Delrieu, G., Morin, E., Tartari, G., and Toffolon, R.: Assessment of gridded observations used for climate model validation in the Mediterranean region: the HyMeX and MED-CORDEX framework, *Environ. Res. Lett.*, 7, 024017, doi:10.1088/1748-9326/7/2/024017, 2012.
- Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., and Michaelsen, J.: The climate hazards infrared precipitation with stations – a new environmental record for monitoring extremes, *Scientific Data*, 2, 150066, doi:10.1038/sdata.2015.66, 2015.
- Gandin, L. S.: Objective analysis of meteorological fields. By L. S. Gandin. Translated from the Russian. Jerusalem (Israel Program for Scientific Translations), Q. J. Roy. Meteorol. Soc., 92, 447–447, doi:10.1002/qj.49709239320, 1966.
- Gómez-Navarro, J. J., Montvez, J. P., Jerez, S., Jiménez-Guerrero, P., and Zorita, E.: What is the role of the observational dataset in the evaluation and scoring of climate models?, *Geophys. Res. Lett.*, 39, L24701, doi:10.1029/2012GL054206, 2012.
- Habets, F., Boone, A., Champeaux, J. L., Etchevers, P., Franchistéguy, L., Leblois, E., Ledoux, E., Le Moigne, P., Martin, E., Morel, S., Noilhan, J., Seguí, P. Q., Rousset-Regimbeau, F., and Viennot, P.: The SAFRAN-ISBA-MODCOU hydrometeorological model applied over France, *J. Geophys. Res.-Atmos.*, 113, D06113, doi:10.1029/2007JD008548, 2008.
- Häggmark, L., Ivarsson, K.-I., Gollvik, S., and Olofsson, P.-O.: Mesan, an operational mesoscale analysis system, *Tellus A*, 52, 2–20, doi:10.3402/tellusa.v52i1.12250, 2000.
- Haylock, M. R., Hofstra, N., Klein Tank, A. M. G., Klok, E. J., Jones, P. D., and New, M.: A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006, *J. Geophys. Res.*, 113, D20119, doi:10.1029/2008JD010201, 2008.
- Herrera, S., Gutiérrez, J. M., Ancell, R., Pons, M. R., Frías, M. D., and Fernández, J.: Development and analysis of a 50-year high-resolution daily gridded precipitation dataset over Spain (Spain02), *Int. J. Climatol.*, 32, 74–85, doi:10.1002/joc.2256, 2012.
- Herrera, S., Fernández, J., and Gutiérrez, J. M.: Update of the Spain02 gridded observational dataset for EURO-CORDEX evaluation: assessing the effect of the interpolation methodology, *Int. J. Climatol.*, 36, 900–908, doi:10.1002/joc.4391, 2016.
- Isotta, F. A., Frei, C., Weigluni, V., Perčec Tadić, M., Lassègues, P., Rudolf, B., Pavan, V., Cacciamani, C., Antolini, G., Ratto, S. M., Munari, M., Micheletti, S., Bonati, V., Lussana, C., Ronchi, C., Panettieri, E., Marigo, G., and Vertačnik, G.: The climate of daily precipitation in the Alps: development and analysis of a high-resolution grid dataset from pan-Alpine rain-gauge data, *Int. J. Climatol.*, 34, 1657–1675, doi:10.1002/joc.3794, 2014.
- Katsanos, D., Retalis, A., and Michaelides, S.: Validation of a high-resolution precipitation database (CHIRPS) over Cyprus for a 30-year period, *Atmos. Res.*, 69, 459–464, doi:10.1016/j.atmosres.2015.05.015, 2016.
- Long, D., Longuevergne, L., and Scanlon, B. R.: Uncertainty in evapotranspiration from land surface modeling, remote sensing, and GRACE satellites, *Water Resour. Res.*, 50, 1131–1151, doi:10.1002/2013WR014581, 2014.
- Martínez de la Torre, A.: Groundwater influence on soil moisture memory and land-atmosphere interactions over the Iberian



- Peninsula, PhD thesis, Universidade de Santiago de Compostela, Santiago de Compostela, 2014.
- Mitchell, K. E.: The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system, *J. Geophys. Res.*, 109, 1–32, doi:10.1029/2003JD003823, 2004.
- Nasonova, O. N., Gusev, Y. M., and Kovalev, Y. E.: Impact of uncertainties in meteorological forcing data and land surface parameters on global estimates of terrestrial water balance components, *Hydrol. Process.*, 25, 1074–1090, doi:10.1002/hyp.7651, 2011.
- Navascués, B., Rodríguez, E., Ayuso, J. J., and Jarvenoja, S.: Analysis of surface variables and parameterization of surface processes in HIRLAM. Part II: Seasonal assimilation experiment, Tech. rep., HIRLAM Technical Report 59, [http://www.hirlam.org/index.php/component/docman/doc\\_download/259-hirlam-technical-report-no-59?Itemid=71](http://www.hirlam.org/index.php/component/docman/doc_download/259-hirlam-technical-report-no-59?Itemid=71) (last access: April 2017), 2003.
- Orth, R. and Seneviratne, S. I.: Introduction of a simple-model-based land surface dataset for Europe, *Environ. Res. Lett.*, 10, 044012, doi:10.1088/1748-9326/10/4/044012, 2015.
- Peters-Lidard, C. D., Kumar, S. V., Mocko, D. M., and Tian, Y.: Estimating evapotranspiration with land data assimilation systems, *Hydrol. Process.*, 25, 3979–3992, doi:10.1002/hyp.8387, 2011.
- Prein, A. F. and Gobiet, A.: Impacts of uncertainties in European gridded precipitation observations on regional climate analysis, *Int. J. Climatol.*, 37, 305–327, doi:10.1002/joc.4706, 2016.
- Quintana-Seguí, P.: SAFRAN analysis over Spain, doi:10.14768/MISTRALS-HYMEX.1388, 2015.
- Quintana-Seguí, P., Le Moigne, P., Durand, Y., Martin, E., Habets, F., Baillon, M., Canellas, C., Franchisteguy, L., and Morel, S.: Analysis of near-surface atmospheric variables: Validation of the SAFRAN analysis over France, *J. Appl. Meteorol. Clim.*, 47, 92–107, doi:10.1175/2007JAMC1636.1, 2008.
- Quintana-Seguí, P., Peral, M. C., Turco, M., Llasat, M.-C., and Martin, E.: Meteorological Analysis Systems in North-East Spain: Validation of SAFRAN and SPAN, *J. Environ. Inform.*, 27, 116–130, doi:10.3808/jei.201600335, 2016.
- Ritter, B. and Geleyn, J.-F.: A Comprehensive Radiation Scheme for Numerical Weather Prediction Models with Potential Applications in Climate Simulations, *Mon. Weather Rev.*, 120, 303–325, doi:10.1175/1520-0493(1992)120<0303:ACRSFN>2.0.CO;2, 1992.
- Rodell, M., Houser, P. R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C.-J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., Entin, J. K., Walker, J. P., Lohmann, D., and Toll, D.: The Global Land Data Assimilation System, *B. Am. Meteorol. Soc.*, 85, 381–394, doi:10.1175/BAMS-85-3-381, 2004.
- Rodríguez, E., Navascués, B., Ayuso, J. J., and Jarvenoja, S.: Analysis of surface variables and parameterization of surface processes in HIRLAM. Part I: Approach and verification by parallel runs, Tech. rep., HIRLAM Technical Report 58, [http://hirlam.org/index.php/component/docman/doc\\_download/247-hirlam-technical-report-no-58?Itemid=71](http://hirlam.org/index.php/component/docman/doc_download/247-hirlam-technical-report-no-58?Itemid=71) (last access: April 2017), 2003.
- Soci, C., Bazile, E., Besson, F., and Landelius, T.: High-resolution precipitation re-analysis system for climatological purposes, *Tellus A*, 68, 29879, doi:10.3402/tellusa.v68.29879, 2016.
- Szczypta, C., Decharme, B., Carrer, D., Calvet, J.-C. C., Lafont, S., Somot, S., Faroux, S., and Martin, E.: Impact of precipitation and land biophysical variables on the simulated discharge of European and Mediterranean rivers, *Hydrol. Earth Syst. Sci.*, 16, 3351–3370, doi:10.5194/hess-16-3351-2012, 2012.
- Szczypta, C., Gascoïn, S., Habets, F., Saadi, A., and Berjamy, B.: Evaluation of the SAFRAN-ISBA-RAPID hydrometeorological chain on a mountainous catchment in a semi-arid region. Case of the Rheraya (Marrakech, Morocco), *Geophys. Res. Abstr.*, 17, EGU2015-14149, 2015.
- Taillefer, F.: CANARI Technical Documentation based on ARPEGE cycle CY25T1 (AL25T1 for ALADIN), Tech. rep., Météo-France, Toulouse, France, 2002.
- Taylor, K. E.: Summarizing multiple aspects of model performance in a single diagram, *J. Geophys. Res.-Atmos.*, 106, 7183–7192, doi:10.1029/2000JD900719, 2001.
- Turco, M. and Llasat, M. C.: Trends in indices of daily precipitation extremes in Catalonia (NE Spain), 1951–2003, *Nat. Hazards Earth Syst. Sci.*, 11, 3213–3226, doi:10.5194/nhess-11-3213-2011, 2011.
- Turco, M., Quintana-Seguí, P., Llasat, M. C., Herrera, S., and Gutiérrez, J. M.: Testing MOS precipitation downscaling for ENSEMBLES regional climate models over Spain, *J. Geophys. Res.-Atmos.*, 116, D18109, doi:10.1029/2011JD016166, 2011.
- Turco, M., Sanna, A., Herrera, S., Llasat, M.-C., and Gutiérrez, J. M.: Large biases and inconsistent climate change signals in ENSEMBLES regional projections, *Climatic Change*, 120, 859–869, doi:10.1007/s10584-013-0844-y, 2013a.
- Turco, M., Zollo, A. L., Ronchi, C., De Luigi, C., and Mercogliano, P.: Assessing gridded observations for daily precipitation extremes in the Alps with a focus on northwest Italy, *Nat. Hazards Earth Syst. Sci.*, 13, 1457–1468, doi:10.5194/nhess-13-1457-2013, 2013b.
- Vidal, J. P., Martin, E., Franchistéguy, L., Baillon, M., and Soubeyrou, J. M.: A 50-year high-resolution atmospheric reanalysis over France with the Safran system, *Int. J. Climatol.*, 30, 1627–1644, doi:10.1002/joc.2003, 2010.
- Weedon, G. P., Balsamo, G., Bellouin, N., Gomes, S., Best, M. J., and Viterbo, P.: The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data, *Water Resour. Res.*, 50, 7505–7514, doi:10.1002/2014WR015638, 2014.