Electronic Supplementary Material

The utility of land-surface model simulations to provide drought information in a water management context using global and local forcing datasets.

Water Resources Management.

Pere Quintana-Seguí[1], Anaïs Barella-Ortiz[1], Sabela Regueiro-Sanfiz[2], and Gonzalo Míguez-Macho[2]

[1]Observatori de l'Ebre, (Universitat Ramon Llull - CSIC), Roquetes (Spain).

[2]Universidade de Santiago de Compostela, Santiago de Compostela (Spain).

2018

1 Introduction

In this electronic supplement we include a comparison of the forcing datasets in both absolute (Sec. 3.1.1) and drought terms (Sec. 3.1.2) and we validate the simulated streamflows (Sec. 3.2).

2 Methodology

To validate streamflow we have used the Kling-Gupta efficiency (KGE) (Gupta et al, 2009):

$$KGE = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2} \quad (1)$$

where r is the correlation coefficient, $\alpha = \frac{\sigma_m}{\sigma_e}$ is the ratio of the standard deviation of the model to that of the evaluation dataset, and $\beta = \frac{\mu_m}{\mu_e}$ is the ratio between the means of the model and evaluation datasets. The best possible value of KGE is 1.

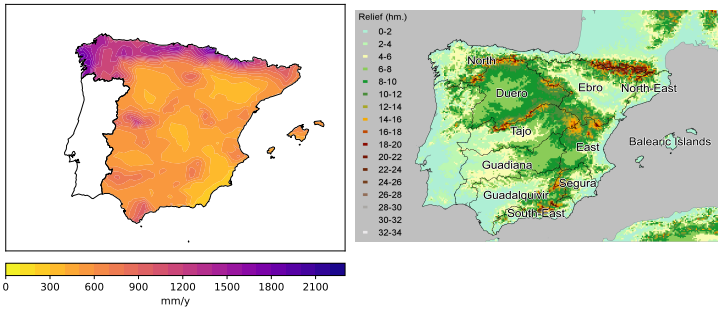
3 Results

3.1 Analysis of forcing datasets

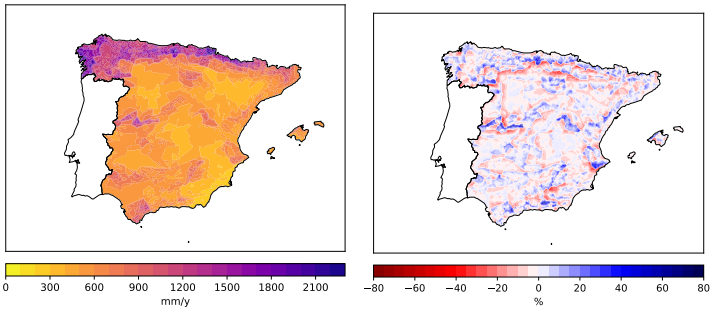
In this section we analyze the forcing datasets. Using SFR as a reference, we validate how the different forcing datasets reproduce precipitation in the area of study. We first evaluate annual means (Sec. 3.1.1) and then we focus on variability using SPI (Sect. 3.1.2). For simplicity and brevity, we limit our comparison to precipitation, even when Land-surface models are also forced with other variables, such as temperature, relative humidity, wind speed and downward short and long wave radiation.

3.1.1 Comparison of annual mean precipitation

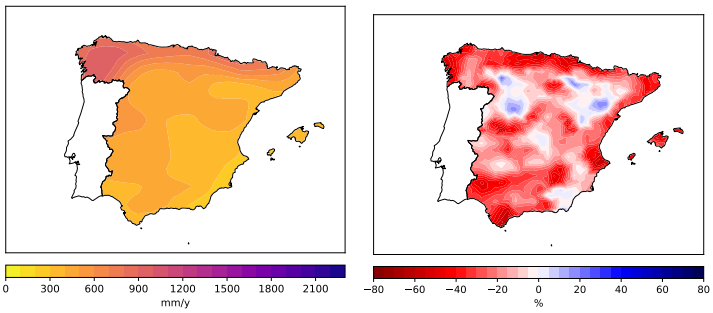
Fig. 1 shows the mean annual precipitation in the area of study, as reproduced by the forcing datasets used in this study. SFR (Fig. 1a), our reference dataset, captures the complexity of the spatial structure of precipitation in Spain. The most important factors determining its distribution are relief (increased precipitation in the main mountain ranges due to orographic lift enhancement and a decrease in inland valleys, due to rain shadowing effect) and the proximity to the Atlantic ocean, upwind from the prevailing westerly flow. SLR (Fig. 1c) is a lower resolution version (30km) of SFR, with a smoother spatial structure retaining nonetheless its main patterns and features. The difference between both datasets (Fig. 1b) shows the error due to resolution and, as expected, is largest on the relief. SLR is similar in resolution to E2O and MSW, thereby allowing for a fairer comparison than SFR. E2O (Fig. 1d) resolves less features in the precipitation pattern than SLR (and hence, SFR), and presents a strong negative bias (Fig. 1e) in most of the area of study. The exceptions are concentrated in the inland valleys of the Ebro and the Duero basins, where precipitation is overestimated due to the misrepresentation of the shadowing effect of the relief by E2O. MSW (Fig. 1f) , also underestimates precipitation (Fig. 1g), but to a lesser degree than E2O, which attests to the incorporation of other sources of precipitation information (from satellite, rain gauges, etc.) representing an important improvement. Notwithstanding, MSW still overestimates precipitation in the Ebro valley (but not in the Duero) and also in a small area of the South East. These results are in accordance with those found by Belo-Pereira et al (2011), who validated the precipitation of different global datasets in the Iberian peninsula.



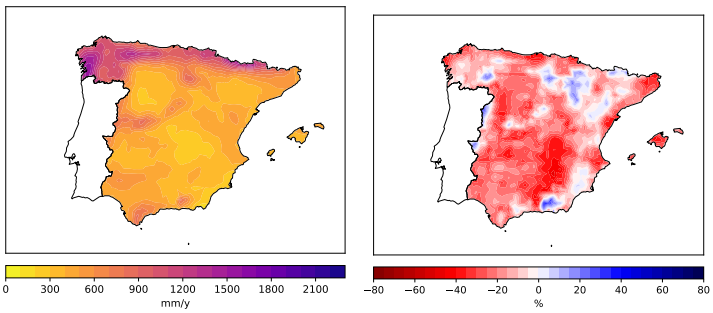
(a) SLR (b) Relief



(c) SFR (d) SFR-SLR

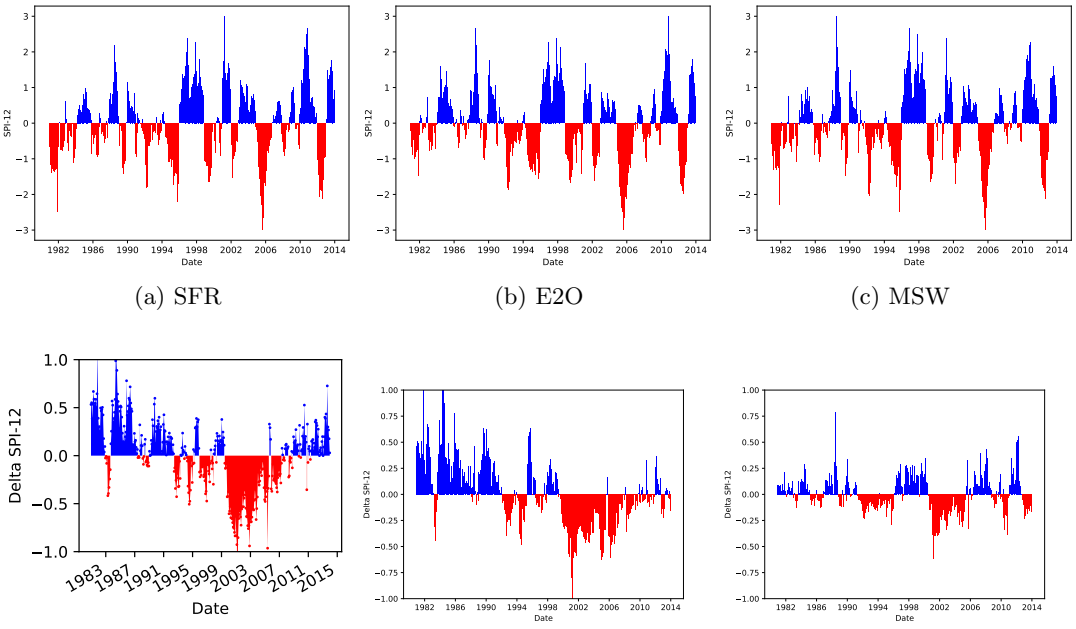


(e) E2O (f) E2O-SLR



(g) MSW (h) MSW-SLR

Figure 1: Mean annual precipitation in the area of study for the period 1980-2013 from the four products used in this study. Panel (b) shows the relief at the resolution of the SFR grid (5 km) and the main river basins (small ones have been aggregated). All products have been regrided to the SFR grid, which is the one used in the LSM simulations. The first column shows means and the second column, differences using SLR as reference.



(a) SFR (b) E2O (c) MSW

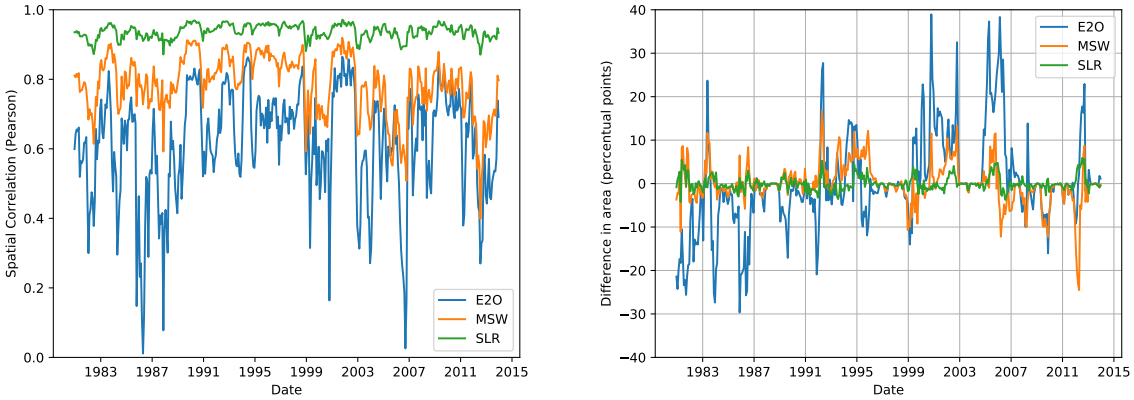
(d) ERA-SFR (e) E2O-SFR (f) MSW-SFR

Figure 2: SPI-12 time series. First row: SPI-12 calculated with the spatially aggregated time series of precipitation of Spain as reproduced by SFR, E2O and MSW. Second row: Difference between the aggregated SPI-12 time series calculated with ERA, MSE and E2O compared to SFR.

3.1.2 Comparison of precipitation drought

In the previous section we found something that was expected, as it is well known that global products underestimate precipitation, specially on the relief. However, in the context of this study, it is more interesting to study the variability of the products, that is, their capacity to reproduce drought. This is the objective of this section.

Figure 2 shows time series of SPI-12 computed from area averaged precipitation over Spain. SPI-12 from SFR, our reference dataset ,(Fig. 2a) indicates that several drought events have occurred during the period, the most severe in 2004-2005. These drought spells coincide with those detected, for instance, by Belo-Pereira et al (2011). Panels (b) and (c) show the same index estimated from E2O and MSW data, respectively (we should not forget that MSW integrates rain gauge data, when comparing these two products). At first sight, the time series look similar, with the main drought periods well detected by the different products. However, there are relevant differences in both drought duration and intensity. This is better seen in panels (e) and (f), which depict the differences between E2O and SFR and MSW and SFR, respectively.



(a) (b)

Figure 3: Panel(a): Time evolution of the spatial correlation (Pearson) between SPI-12 derived from SLR, MSW and E2O compared to SFR. Panel (b): Differences in the proportion of the area under drought (SPI-12< -1) estimated by E2O, MSW and SLR, compared with SFR.

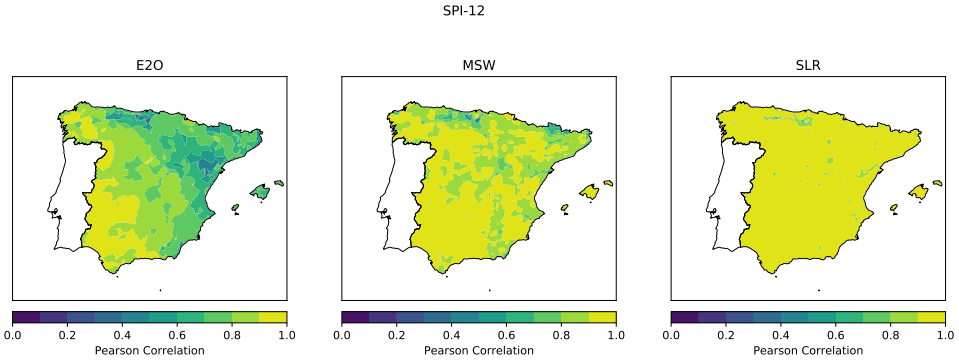


Figure 4: Map of the temporal correlation (Pearson) between the SPI-12 time series produced by E2O, MSW and SLR compared to SFR.

The differences between E2O and SFR present a noticeable trend, which can affect drought studies, as it modifies the intensity of droughts with time. To determine whether this trend develops from the original ERA-Interim data (ERA from now on), on which E2O is based, or it is due to corrections applied by the algorithms used to produce E2O, panel (d) shows the difference between ERA and SFR. The plot reveals that the problem is mainly inherited from ERA. For MSW (f) the differences with SFR are smaller, which means that the MSW algorithm effectively combines diverse sources of data, including ERA, other reanalyses, satellite and rain gauge observations, in order to better estimate real precipitation. However, even when the differences are reduced in this case, they remain relevant. An analysis of the monthly maps of drought for the three products (not shown) reveals that E2O and MSW are able to reproduce the main drought spatial patterns, but miss the finer details. MSW is nevertheless an important improvement, comparing with E2O. This is summarized in Fig. 3 (a), which shows the spatial correlation (the similarity of the drought maps for each month) of the SPI-12 derived from E2O, MSW and SLR, with that of SFR. The green line corresponds to SLR. As expected, its correlation is very high, around 0.9; the deviation from 1. reflecting the error due to resolution loss. The other products, add further sources of error, making their correlations lower. MSW has better correlations and less variability in time (more robustness) than E2O, which sometimes completely misses the spatial structure of drought, with very low correlations that approach zero for some months. Another perspective on product performance is given by Fig. 3 (b), which shows the differences in the percent area under drought (SPI-12< -1). In general, and as expected, the errors of SLR are minimal, but still reaching 6% for some months. MSW does not present a systematic bias, but it overestimates the area under drought in the first part of the 1990s and it underestimates it from year 2006 on. E2O shows the trend seen before (increasing drought with time). In addition, we have also studied the temporal correlation of the drought indices, shown in Fig. 4 for each grid point. As the most interesting result, this comparison denotes the improvement that MSW represents with respect to E2O. E2O has lower correlations in the Ebro basin and the Mediterranean coast, which is a known limitation of global products in the Iberian Peninsula (Belo-Pereira et al, 2011; Andrade and Belo-Pereira, 2015). The incorporation of satellite rainfall data into MSW ameliorates the detection of the small scale structures that produce precipitation in the Mediterranean area. SLR has high

Code	E2O-DIF	MSW-DIF	MSW-LHD	SFR-3L	SFR-DIF	SFR-LHD	SLR-DIF	SMP-SMP
1359	-0,06	0,45	0,43	0,58	0,50	0,42	0,41	0,61
2029	0,18	0,89	-0,19	0,34	0,73	-0,20	0,69	0,75
2030	-0,29	0,12	0,24	0,44	0,26	0,31	0,13	0,76
2031	-0,12	0,21	0,16	0,57	0,39	0,24	0,30	0,75
2036	0,07	0,47	-1,28	0,64	0,68	-1,23	0,57	0,68
2042	0,33	0,65	0,35	0,34	0,70	0,26	0,74	0,74
2043	0,26	0,72	0,09	0,35	0,83	0,08	0,87	0,76
2074	0,1	0,54	0,41	0,57	0,78	0,47	0,74	0,81
2087	0,55	0,54	-0,26	0,57	0,73	-0,17	0,57	0,71
2097	0,29	0,72	0,09	0,32	0,81	0,08	0,84	0,76
3005	0,02	0,09	0,05	0,51	0,57	0,21	0,43	0,75
3169	0,6	0,52	-1,02	0,66	0,55	-0,99	0,57	0,77
8091	-0,02	0,07	0,19	0,61	0,65	0,1	0,5	0,63
9002	0,04	0,82	-0,18	0,76	0,85	0,02	0,84	0,6
9003	-0,06	0,35	0,15	0,5	0,87	-0,02	0,76	0,65
9004	-0,13	0,44	0,36	0,68	0,52	0,38	0,63	0,86
9065	-0,13	0,45	0,36	0,75	0,63	0,44	0,59	0,71
9120	-0,04	0,67	0,04	0,56	0,69	0,29	0,68	0,79

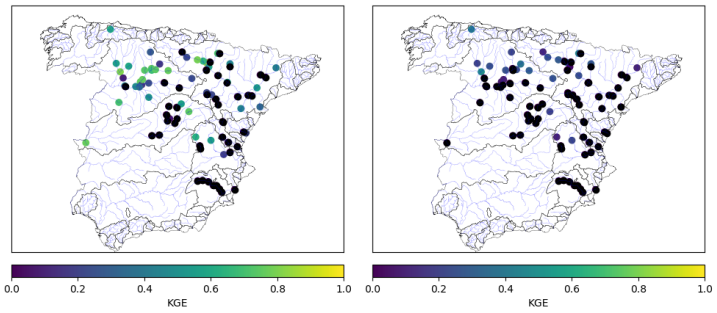
Table 1: Performance of the different simulations on stations with near natural flow. KGE is the criterion used to evaluate the models and observations are the reference. Stations with near-natural flow have been defined as those where SMP-SMP has a KGE$>$ 0.6.

temporal correlations with SFR, as expected.

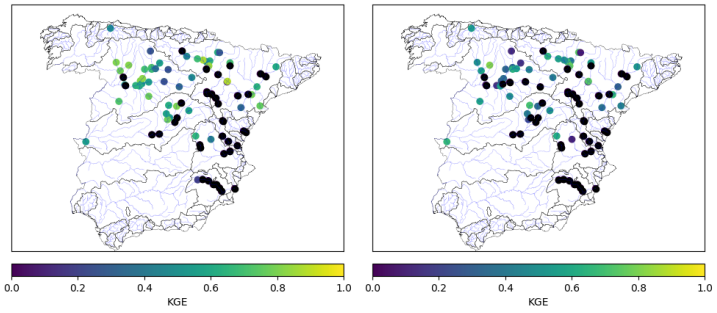
3.2 Validation of simulated streamflow

First, we validate the models by comparing simulated flows with observations in terms of KGE. Fig. 5 (a) shows results for SMP-SMP (SMP uses its own forcing dataset, which is close to SFR as it uses mostly the same observational dataset, even though the methodologies are different) and panels (b), (c) and (d) correspond to the LSMs in this study, forced by SFR. The maps depict numerous black dots corresponding to negative values of KGE, which indicate a remarkably bad performance likely due to the high level of human intervention in the Spanish river systems. The worst KGEs for our benchmark simulated flows (SMP-SMP, a), are found toward the South and to the East of Spain, indicating more departed from natural regime flows. This is in coherence with the maps of mean precipitation (see Fig. 1) and also with the areas of higher consumption.

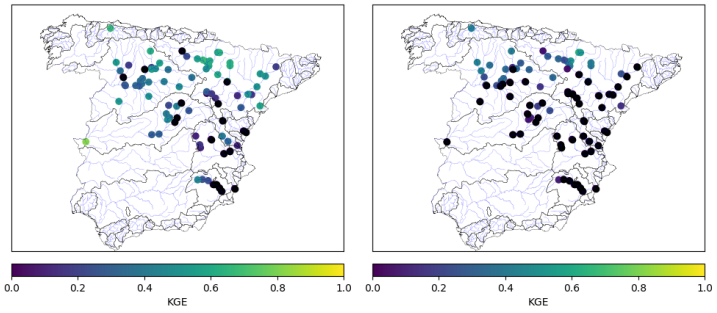
More interesting is to compare the performance of the models at gauge stations with near



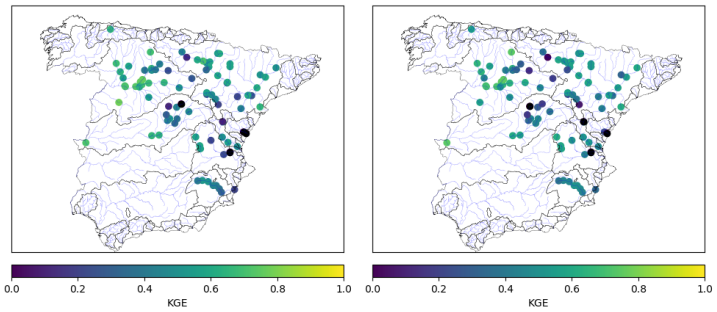
(a) SMP-SMP vs OBS (b) SFR-LHD vs OBS



(c) SFR-DIF vs OBS (d) SFR-3L vs OBS



(e) SFR-3L vs SMP-SMP (f) SFR-LHD vs SMP-SMP



(g) SFR-DIF vs SMP-SMP (h) SLR-DIF vs SMP-SMP

Figure 5: Performance of the modeled streamflow, measured according to the KGE criterion, as compared with observations (OBS) and naturalized streamflow (SMP-SMP). Negative values of KGE are depicted in black. Note that observations reflect water management procedures, while none of the models simulate those. All simulations, except for SMP, are forced by SFR. SMP uses its own forcing dataset.

natural flow regime. In order to single out these stations, we have used a simple and objective method: we select those where $KGE > 0.6$ for SMP-SMP, considering that the latter is the best estimate of natural flow currently available, even though SMP is a model and thus it has its own issues. This value of KGE is arbitrary, but results would not change considerably if we chose other reasonable numbers, such as 0.5 or 0.7. Table 1 shows the KGE scores of the simulations at the resulting subset of near natural regime stations. The location of these stations can be found in panel (a) of Fig. 5 (the green and yellowish dots on the map). Most of them are located in the Duero river basin, but there are a few in the Ebro, Tajo, the eastern basins and even one station in the North.

Table 1 shows that the simulation with the best results, apart from SMP-SMP, is SFR-DIF. SLR-DIF, using the same model and a degraded in resolution version of the same forcing, has similar results. In fact, SLR-DIF is better for some stations. This means that SLR provides roughly the same information as SFR, whose spatial structure is sometimes too constrained by the fact that it uses homogeneous zones as its main spatial unit to perform the optimal interpolation (see Quintana-Seguí et al (2017) for more details on the effect of the SAFRAN methodology on the final spatial structure of precipitation). In other words, the mean distance between rain gauges used by SFR is greater than its nominal 5km resolution. Another interesting result found in Table 1 is that MSW-DIF has a notably good performance. This suggests that the precipitation data sources integrated in this product are accurate and greatly improve E20, which, in turn, produces low skill streamflow results that include the spurious trend discussed in Section 3.1 .

Concerning the differences among models, Table 1 shows that in terms of streamflow, DIF is superior to 3L. The performance of LHD is lower due to a general overestimation of baseflow, which is highly influenced by the water table. The overestimation of baseflow by LHD is likely related to some deficiency in its river conductance formulation, which regulates the interaction between river and water table. At the other extreme, DIF and 3L cannot sustain summer flows in some basins and their recession curves are often too steep as they do not have any physics to reproduce groundwater lateral flows or any interaction with the water table. In all cases, SMP is the model with the best skill scores. Another interesting result is the great improvement that MSW represents compared with E2O in these rather small basins, and how well simulations with

SLR perform, suggesting that the key issue is not resolution but the quality of the information provided by the forcing dataset.

To evaluate model performance in human influenced stations, we use SMP-SMP for comparison, instead of observations. SMP-SMP provides a reasonable estimate of natural flows, thus making this comparison valuable. The results are shown in Figure 5 (panels e,f,g and h). As before for stations with natural flow regimes, now with SMP-SMP as reference, SFR-DIF is the best performing simulation and SLR-DIF is close to it. SFR-3L is clearly less skillful than DIF, and results from LHD have the smallest KGEs, due to the aforementioned baseflow problems.

References

Andrade C, Belo-Pereira M (2015) Assessment of droughts in the Iberian Peninsula using the WASP-Index. Atmospheric Science Letters 16(3):208–218, DOI 10.1002/asl2.542, URL `http://doi.wiley.com/10.1002/asl2.542`

Belo-Pereira M, Dutra E, Viterbo P (2011) Evaluation of global precipitation data sets over the Iberian Peninsula. Journal of Geophysical Research: Atmospheres 116(July):1–16, DOI 10.1029/2010JD015481

Gupta HV, Kling H, Yilmaz KK, Martinez GF (2009) Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. Journal of Hydrology 377(1-2):80–91, DOI 10.1016/j.jhydrol.2009.08.003, URL `http://dx.doi.org/10.1016/j.jhydrol.2009.08.003`

Quintana-Seguí P, Turco M, Herrera S, Miguez-Macho G (2017) Validation of a new SAFRAN-based gridded precipitation product for Spain and comparisons to Spain02 and ERA-Interim. Hydrology and Earth System Sciences 21(4):2187–2201, DOI 10.5194/hess-21-2187-2017, URL `http://www.hydrol-earth-syst-sci.net/21/2187/2017/`