Supplement of

Hydrological assessment of atmospheric forcing uncertainty in the Euro-Mediterranean area using a land surface model

E. Gelati et al.

Correspondence to: J.-C. Calvet (jean-christophe.calvet@meteo.fr)

S1 River discharge score maps

We show and discuss the maps of river discharge scores (Fig. S1). The main outcomes are reported in Sect 5.3.

To analyse the spatial patterns of river discharge scores, Fig. S1 maps the values summarised in Fig. 7. To have a unique colour scale for all maps, relative errors are represented as $1-|RE_\mu|$ and $1-|RE_\sigma|$.

The geographical distributions of both MC and AC are very well correlated across the simulations: map correlations range between 0.94 and 0.99. Thus, we find not only strong similarities in the frequency distributions of correlation scores (Fig. 7), but also very similar spatial patterns. This indicates that the simulated timing and shape of monthly discharge time series are robust to forcing uncertainty.

MC and AC values above 0.6 prevail in catchments of central Europe (Danube, Vistula, Rhine, Elbe, Oder, Weser, Po, Meuse, and downstream Rhone), and Iberian Peninsula (Ebro, Tejo, Duero, Guadiana, and Guadalquivir). Lower values are scattered across the northern and eastern parts of the domain, where 6 of the 10 largest catchments considered are located (Volga, Dnepr, Don, Severnaya Dvina, Neva and Ural).

The $|RE_\mu|$ maps are slightly less correlated than those of AC and MC (between 0.68 and 0.91). Absolute error values are below 0.4 at most gauges of central Europe, Scandinavia, the eastern Baltic basin, and northern Russia. Larger values are generally found in the south-western Iberian Peninsula (Guadiana, Guadalquivir and upstream Tejo) and in the south-eastern part of the domain (Dnepr and Ural).

Also $|RE_\sigma|$ maps are well correlated spatially, with correlation coefficients ranging from 0.83 to 0.97. The poorer values taken systematically by the P-ERA over several catchments in central Europe (Danube, Rhine, Elbe, Oder and Weser) and over the Iberian Peninsula reflect the inferior performance shown by the $|RE_\sigma|$ exceedance frequency curve (Fig. 7).

As KGE is function of relative errors and MC, its maps are well correlated across simulations (0.77–0.96). Moreover, KGE maps are well correlated with absolute relative error maps (0.60–0.80 for $|RE_\mu|$ and 0.84–0.95 for $|RE_\sigma|$). Instead, they are negatively correlated with MC maps. As expected, KGE spatial patterns are largely influenced by those of the relative error scores, in particular $|Re_\sigma|$.

S2 Catchment averaged annual cycles

We present the average annual cycles of forcing and LSM output variables over the Danube, Rhone, Ebro and Po river catchments (Fig. S2 to S5). These results are summarised in Sect. 5.4.

Forcing input and simulated output are compared to the datasets described in Sect. 3.2 to 3.4: downward LW and SW radiation are computed from SRB and CERES data, referred to as benchmark 1 and 2 in Fig. S2 to S5; E-OBS (benchmark 1) and CRU (benchmark 2) data are used for air temperature; E-OBS (benchmark 1) and GPCC (benchmark 2) data for precipitation; ESA-CCI, GIMMS and GRDC are the reference datasets for LAI, SSM and river discharge.

All variables are averaged over monthly time steps where river discharge measurements are available. Subsets of such time steps are used to compute the annual cycles of some datasets that are not available over the whole simulation period (1979–2012): SRB (1983–2007) and CERES (2000–2014) LW and SW radiation; and GIMMS LAI (1981-2011).

ERA-based forcing datasets show close (sometimes identical) annual cycles of LW and SW radiation, air temperature, wind speed, and relative air humidity. ERA-I and P-ERA forcing data are identical except for precipitation. WFDEI differs from ERA-I for precipitation, downward SW radiation, and air temperature, which affects relative air humidity. However the ERA-I and WFDEI annual cycles of LW radiation and wind speed are slightly different (Fig. 8 to 11). The differences are caused by a version discrepancy between the ERA-I reanalysis underlying the WFDEI data and the one used as forcing, which is the most recent.

Compared to ERA-based datasets, the annual cycles of PGF wind speed have larger amplitudes and also generally larger values, especially in winter. PGF relative humidity is systematically the largest, with shapes and amplitudes that are similar to those of the other forcing datasets.

In terms of air temperature, PGF is identical to CRU, whose annual cycles are also undistinguishable from E-OBS. Also the temperature annual cycles of ERA-based forcing datasets are nearly overlapping with the benchmarks: the largest discrepancies occur in summer values over the Ebro and Po catchments, where ERA-I and WFDEI values are slightly smaller than the rest.

For both SW and LW radiation, SRB and CERES annual cycles are overlapping over 3 of the 4 considered catchments. An exception is the Danube, where SRB summer SW is lower and SRB LW is slightly larger than the corresponding CERES values. The SW radiation annual cycles of all forcing datasets are very close to the benchmarks during autumn and winter. In spring and summer, forcing SW cycles are generally larger than the benchmarks except in the Danube catchment, where CERES values are larger than PGF and only slightly smaller than ERA-based data. Conversely, forcing LW cycles are generally below the benchmarks, with the following exceptions: PGF in the Danube and Po catchments; and WFDEI in the Ebro.

Precipitation is the variable for which catchment averaged annual cycles differ the most across datasets. Forcing values are often larger than benchmarks, among which GPCC is higher than E-OBS for 3 of the 4 chosen catchments. Cycle timings are generally in good agreement within each catchment.

All simulated annual cycles of RSSM are close. The similarity is partly due to the pixel-wise rescaling (Sect. 4.1) and is particularly strong among ERA-based simulations, which are characterised by the same precipitation occurrences. PGF, which is independent from ERA-I, yields few slightly different values, as for instance in the Po catchment during winter. Although remotely sensed RSSM are also rescaled, their annual cycles are significantly different from the simulated ones: while there is a fairly good phase agreement, the amplitude of ESA-CCI cycles is systematically smaller. Thus there may be structural differences between modelled and remotely sensed SSM that are beyond first-order soil property discrepancies, which should have beeen filtered out by the rescaling, and deserve further investigation. Moreover these results suggest that SSM simulations are robust to uncertainties in atmospheric forcing data, in particular precipitation amounts. Thus they are consistent with the correlation analysis presented in Sect. 5.1.

The simulated LAI annual cycles display common patterns in the Danube, Rhone and Po catchments: the spread between the curves is very small at the beginning of the growing season for the majority of vegetation types in the simulation domain (February-April); it increases as the maximum phenological development is approached (May-July); it is approximately constant during the beginning of the LAI decline phase (August-October); finally it shrinks during late autumn and winter (November-January). In these three catchments simulated cycles are in good agreement with GIMMS when LAI is increasing at the fastest rates (from March to May or June). They overestimate both the cycle maxima, which occur between June and August, and the values when LAI is declining. In the Ebro catchment, simulated cycles are systematically above the benchmark and feature larger ranges of variation: the differences with the GIMMS cycle are largest shortly before and after the maximum phenological development (May–July); are approximately constant from late summer to autumn; and progressively decrease in winter.

In all four catchments PGF yields the largest simulated LAI cycle values. Maximum differences occur in spring and early summer in the Ebro, and at the beginning of LAI decline in the other catchments. Precipitation, SW radiation and air temperature, which are among the primary drivers of phenological development, seem not to be accountable for this. Indeed PGF values are generally inside the inter-forcing spread or very close to the other datasets cycles. Instead, this might be explained by PGF relative humidity, which is by far the largest. The consequently reduced evaporative demand of the atmosphere may favour larger leaf stomata aperture and thus higher photosynthetic activity. The PGF evapotranspiration cycles, which are lower than the others especially in spring and summer, may constitute another signature of such reduction. For this hypothesis to be likely, however, we should also assume that the larger PGF wind speed does not counterbalance the effect of relative humidity on atmospheric water demand. Although the PGF wind speed cycle is almost systematically above the others, its positive bias is minimised during late spring and summer, i.e. when the phenological development is most intense. This hypothesis should be tested with a dedicated sensitivity study designed to isolate the effects of each forcing variable on the land surface water balance and phenology.

The maxima of LAI and evapotranspiration annual cycles of ERA-based forcing datasets appear weakly linked to precipitation. The datasets characterised by larger precipitation tend to yield higher annual peaks of LAI and evapotranspiration. An exception is the Po catchment, where P-ERA produces the largest LAI and evapotranspiration values (among ERA-based datasets), while WFDEI precipitation is the largest in spring and summer. Furthermore the ERA-based atmospheric water demand cycles are likely to be similar, since the relevant forcing variables are either very close or identical across the datasets. However, as stated above for PGF, further testing would be needed to validate this hypothesis. Evapotranspiration annual cycles are generally in phase, except for PGF annual maxima that are delayed 1 month compared to ERA-based forcing datasets in 3 of 4 catchments (Danube, Rhone and Po). The evapotranspiration spread between annual cycles is always below 20 mm/month.

At most of the 35 GRDC downstream gauges, simulations reproduce fairly well shape and timing of the measured discharge cycles. However, most simulated discharge cycles have larger amplitudes than GRDC measurements (see e.g. Fig. S2 to S4). Average summer-autumn discharge is systematically underestimated by the simulations. In winter and spring values are generally overestimated and the inter-simulation spread is maximum, consistently with the precipitation cycles.

Reservoir regulation and water abstractions for irrigation, which are not simulated by the model, may be partly accountable for the discharge seasonality mismatches. Reservoirs generally retain water in wet seasons to release it during dry periods, for several purposes (irrigation, water supply, flood protection, hydropower production, navigation, etc.). Irrigation also affects the catchment water balance by increasing evapotranspiration, thus generating complex interplays between human alterations of the streamflow regime. However, evaluating these feedbacks is beyond the scope of this study.

In the following subsections, we describe how the simulated annual cycles of river discharge compare to GRDC data over the Danube, Rhone, Ebro and Po catchments.

S2.1 Danube

In the Danube catchment, all simulations anticipate and overestimate the spring discharge peak (Fig. S2). The PGF and WFDEI simulations best approximate the dicharge annual cycle, by minimising winter-spring overestimation and summer-autumn underestimation. This is consistent with the discharge scores displayed in Table 2.

5   The PGF precipitation annual cycle is the lowest among forcing datasets. It is also the closest to GPCC, which is above E-OBS by an almost constant offset. At the same time, PGF spring-summer evapotranspiration is the lowest. In contrast, the P-ERA forcing features the largest winter precipitation and produces the largest overestimation of winter and spring discharge. Although ERA-I and P-ERA spring-summer precipitation values are the largest, they yield the most severe underestimations of summer and autumn discharge.

10

S2.2 Rhone

In the Rhone catchment, all simulated discharge cycles feature spring peaks not corresponding to the measurements (Fig. S3). ERA-I and P-ERA predict them for April, while WFDEI and PGF for May. Except for these peaks, the shape and timing of simulated discharge cycles are reproducedobserved fairly well. However, as mentioned above, the amplitudes of the simulated discharge cycles are overestimated.

P-ERA and ERA-I discharge cycles have similar shapes and timings. The difference between P-ERA and ERA-I discharge varies seasonally, reaching the maximum in winter and the minimum in summer, consistently with the conditional bias between their precipitation cycles. The local minima and maxima of all precipitation cycles are in phase with those of the reference datasets. However forcing cycles are above both benchmarks except for PGF in winter.

PGF and WFDEI minimise the discharge underestimation in summer and autumn. Despite the May peak, they provide the best approximations of the GRDC discharge cycle, consistently with the river discharge scores (Tables 2 and 3).

S2.3 Ebro

All simulations yield lower summer and autumn average discharge compared to GRDC (Fig. S4). Winter and spring discharge is overestimated by all simulations except for ERA-I, which is negatively biased (Table 2). At the other extreme, the P-ERA precipitation annual cycle is the largest for all months, resulting in the most positively biased discharge simulation. Furthermore, P-ERA produces the largest estimates of averaged winter and spring discharge. During winter and spring, WFDEI and PGF discharge cycles lie in the middle of the inter-simulation range, whose lower and upper bounds are well approximated by ERA-I and P-ERA. During summer and early autumn, WFDEI and PGF discharge values are very close to ERA-I and P-ERA, respectively.

ERA-I discharge is the lowest for nearly all months (except September), due to the combined effect of its forcing precipitation and simulated evapotranspiration: its average annual precipitation is smaller than P-ERA and WFDEI but larger than PGF, while its evapotranspiration is the second largest after P-ERA. The PGF and WFDEI precipitation cycles are very close to GPCC. The latter is almost parallel to the E-OBS cycle, with a difference that is comparable to the forcing spread.

Although the goodness-of-fit of simulated discharge annual cycles cannot be unambiguously ranked, we observe the following: all simulations reproduce the shapes of the annual cycles fairly well; ERA-I is the only simulation that does not severely overestimated the cycle amplitude; PGF and WFDEI yield the least biased simulations; while P-ERA simulates the most positively biased discharge.

S2.4 Po

Among all 35 catchments considered in this study, only the Po is characterised by a bimodal annual discharge pattern (Fig. S5). The spring and autumn GRDC discharge peaks match well with both E-OBS and GPCC precipitation annual cycles. The spring peak is well reproduced in timing by WFDEI and PGF, and in amplitude by P-ERA and PGF.

5   While all simulations feature a spring discharge peak, the autumn peak is not properly reproduced by any of them. The simulated annual cycles increase from August (P-ERA and PGF) or September (ERA-I and WFDEI), but all are below the average October GRDC discharge. As a consequence, all simulated annual discharges are negatively biased (Table 2). In contrast, forcing precipitation cycles are almost always above E-OBS and GPCC, except for a few ERA-I values. ERA-I precipitation cycle is the lowest among forcing datasets and the closest to the benchmarks. The other 3 forcing precipitation

10   cycles are close to each other, except for some significant differences in spring and autumn, and generally above ERA-I.
The apparent contradiction between precipitation overestimation and autumn discharge underestimation must be considered carefully, as both precipitation and discharge reference datasets are affected by uncertainties. Moreover, they are independent and there is no warranty about their reciprocal consistency. However, the inability of the model to reproduce the bimodal discharge annual cycle may indicate that the model does not simulate some important processes, e.g. reservoir regulation and

15   irrigation.
As for the Ebro, a clear ranking cannot be established among the simulated annual cycles of the Po catchment. Indeed shape, timing and bias issues affect all simulations.
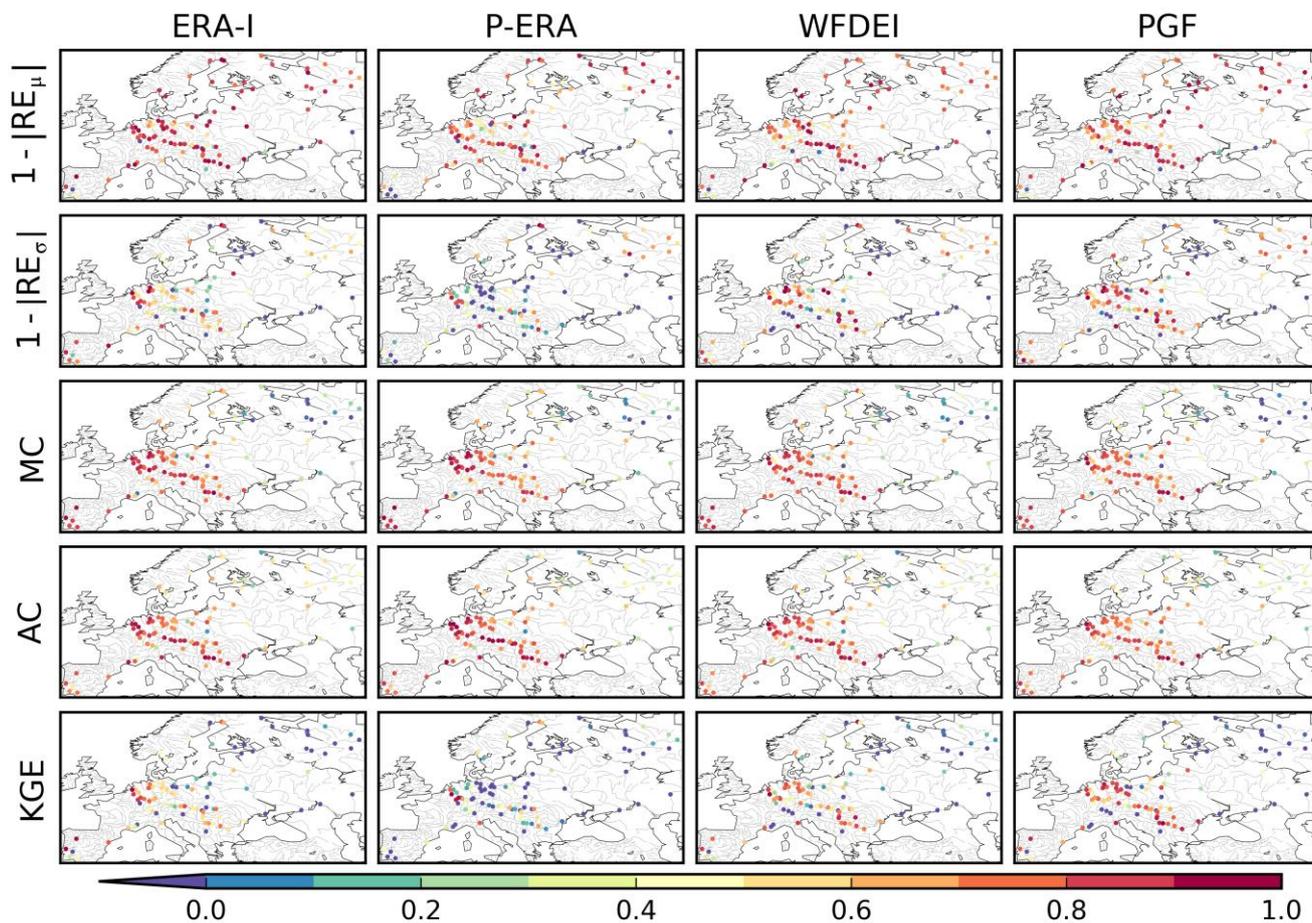
Figure S1. Maps of river discharge scores computed at all GRDC gauges.

5
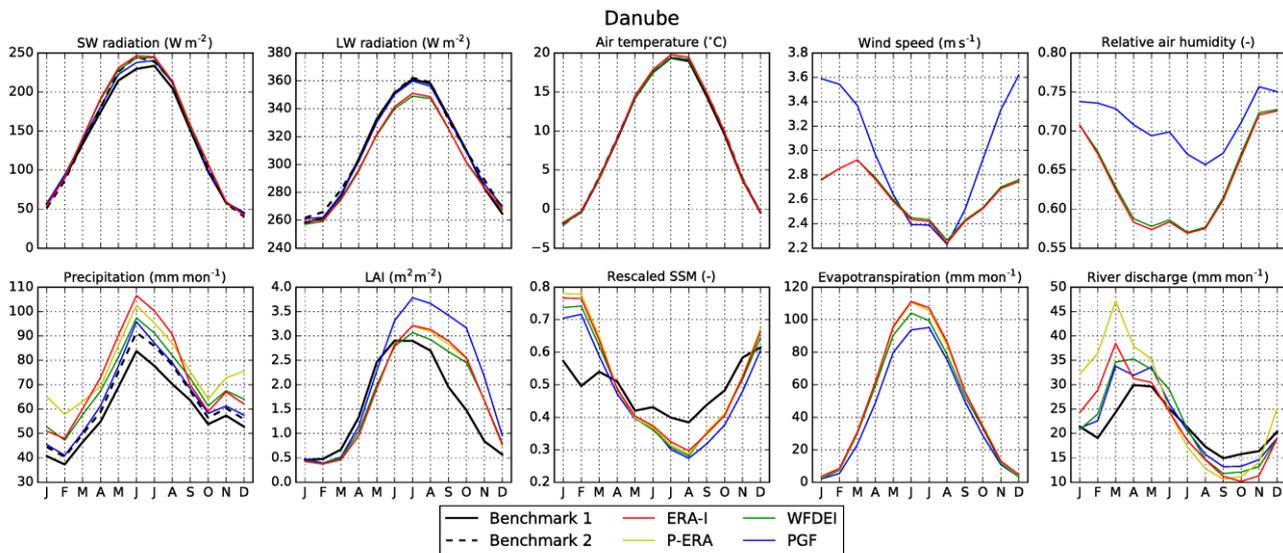
Figure S2. Danube river catchment – averaged annual cycles of forcing and simulated variables with corresponding reference datasets.

5
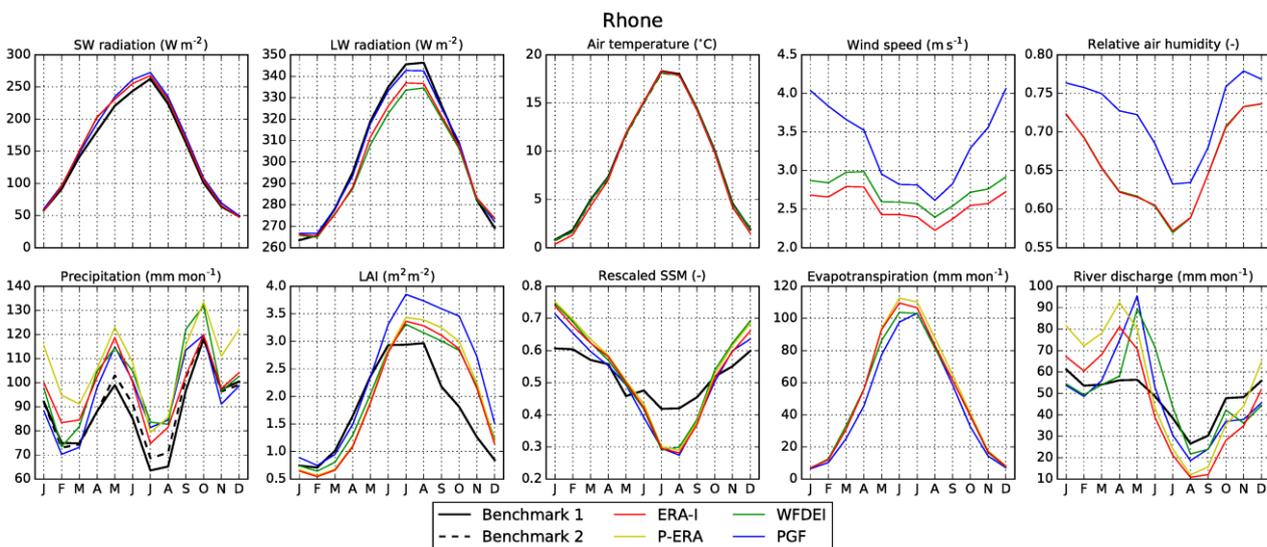
Figure S3. Rhone river catchment – averaged annual cycles of forcing and simulated variables with corresponding reference
10   datasets.
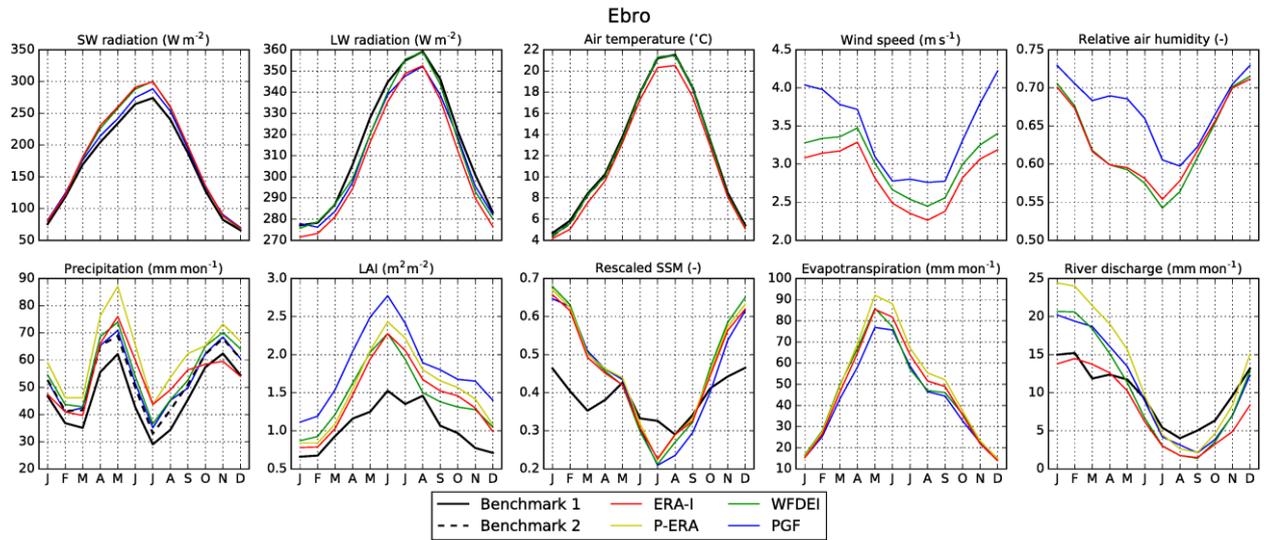
Ebro

Figure S4. Ebro river catchment – averaged annual cycles of forcing and simulated variables with corresponding reference datasets.

Po
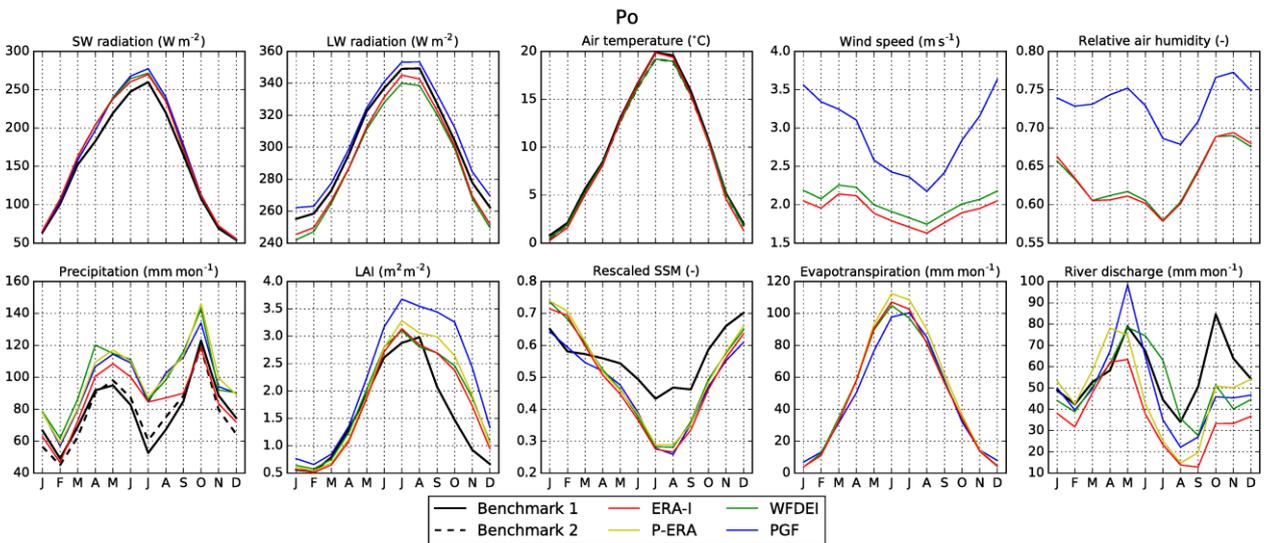
Figure S5. Po river catchment – averaged annual cycles of forcing and simulated variables with corresponding reference datasets.