Hydrology in a Changing World: Environmental and Human Dimensions
Proceedings of FRIEND-Water 2014, Montpellier, France, October 2014 (IAHS Publ. 363, 2014).

Selecting an optimal climatic dataset for integrated modelling of the Ebro hydrosystem

A. DEZETTER[1], J. FABRE[2], D. RUELLAND[2] & E. SERVAT[1]
[1] IRD, [2] CNRS–UMR HydroSciences Montpellier, Place E. Bataillon, 34395 Montpellier Cedex 5, France
alain.dezetter@ird.fr

Abstract This study aims at defining a method for selecting an optimal interpolated climatic dataset in the context of an integrated modelling of the Ebro (85 000 km^2, Spain) hydro-system. Two different sets of temperature and precipitation were chosen according to data availability criteria, and each set was interpolated on an 8 × 8 km grid covering the basin, with and without a monthly or annual altitudinal gradient applied by geographical area (Pyrenees, Cantabrian and Iberian ranges) over 1000 m a.m.s.l. Seven basins (464 to 2975 km^2) representing different flow regimes of the Ebro catchment were chosen to evaluate the performance of 24 different climatic datasets for modelling water resources. We used a conceptual model (GR4j) at a daily time step to test the sensitivity of hydrological modelling to the different sets of precipitation and temperature. A global score was attributed to each dataset according to the different hydrological modelling criteria in order to discriminate the best-performing dataset.

Key words climatic datasets; interpolation; rainfall–runoff modelling; Ebro basin; hydro-system modelling; selection method

INTRODUCTION

The Mediterranean basin is characterized by limited and unevenly distributed water resources. It was identified as one of the world's most vulnerable regions to climatic and anthropogenic changes (Alcamo *et al.*, 2007). According to Milano *et al.* (2012, 2013a), a 30–50% decline in freshwater resources is projected by the 2050 horizon over most of the Mediterranean basin, while anthropogenic pressures on water resources are expected to increase.

In this context the question of water resources' capacity to meet demand is essential. The GICC-REMedHE project (http://www.remedhe.org, *Identification et impacts du changement climatique sur la gestion intégrée des Ressources en Eau en Méditerranée : évaluation comparative Hérault-Ebre*), aims to develop a generic approach to represent and simulate Mediterranean hydro systems of different geographical scales and levels of complexity: the Hérault (2500 km^2, France) and the Ebro (85 000 km^2, Spain) catchments. In order to develop this approach's ability to simulate hydro-system evolutions under complex climatic and anthropogenic long-term scenarios, it must be tested over long past time periods (30–50 years). As a result a precise representation of water resources and their availability are essential in this integrated modelling context.

In this paper, we focus on the Ebro catchment's water resources. The Ebro is a highly regulated basin (over 200 dams) that hosts 60% of Spain's fruit production and 25% of its hydro-electric power production. The contrasting hydro-climatic conditions within the basin call for an accurate representation of climatic inputs in view of evaluating water resources through rainfall–runoff modelling.

This study aims at proposing a method to interpolate precipitation and temperature data on an 8 × 8 km grid over the Ebro catchment, taking into account climatic heterogeneities through altitudinal gradients and considering the objective of water resources evaluation. This method should lead to reference data-grids for precipitation and temperature over a nearly 30-year period.

STUDY AREA

The Ebro catchment is the largest basin in Spain. It drains an area of about 85 000 km^2 in the northeast of the country (Fig. 1). It is characterized by a Mediterranean valley, surrounded by mountains including the Pyrenees and Cantabrian range to the north, the Iberian massif to the south and the coastal Catalan chain to the east. Altitudes range from 4 to 3383 m a.m.s.l. according to the SRTM (http://srtm.csi.cgiar.org) digital elevation model (DEM). The main stream Ebro is a 910 km long river flowing from the Cantabrian Mountains to the Mediterranean Sea.

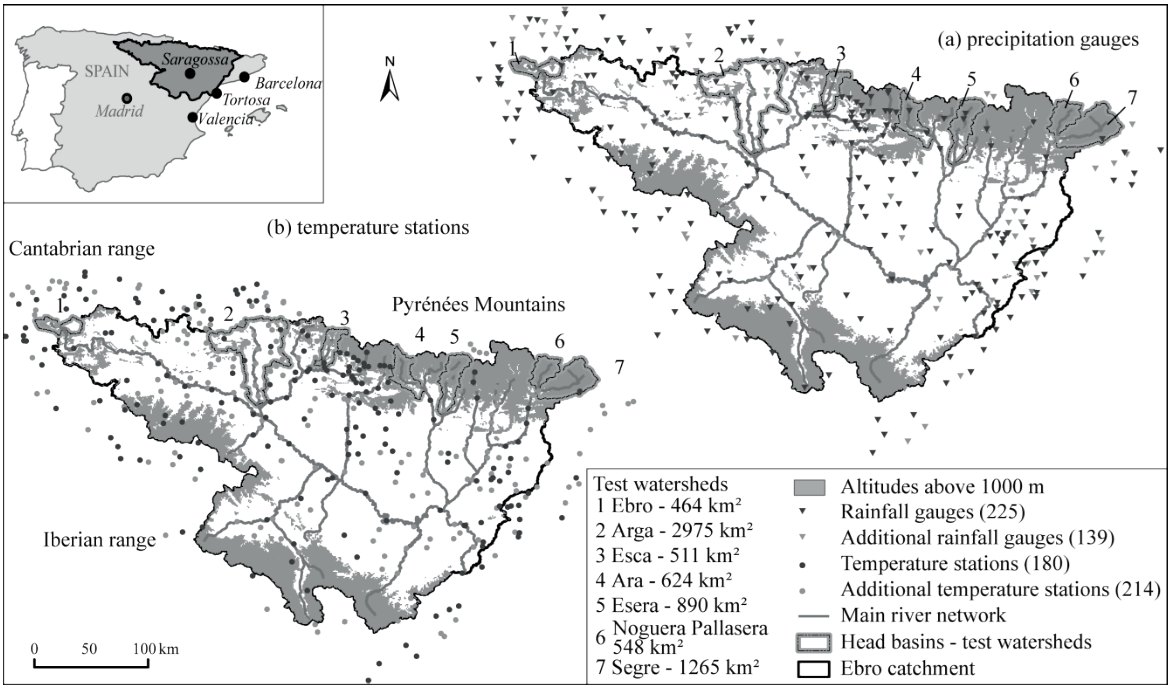

Fig. 1 Location of the Ebro basin, location of altitudes above 1000 m a.m.s.l. and test watersheds: (a) set of precipitation gauges, and (b) set of temperature stations.

The heterogeneous topography of the catchment as well as the climatic influence of both the Atlantic Ocean and the Mediterranean Sea lead to a complex spatial distribution of climate variables (Vicente-Serrano & López-Moreno, 2006; Milano *et al.*, 2013b). Mean annual precipitation and temperatures vary with altitude, ranging from more than 1400 mm and 8°C in the Pyrenees to less than 400 mm and 18°C in the Ebro valley, averaged over the 1969–2002 period. These climatic differences are reflected in the hydrological regimes. According to López-Moreno *et al.* (2010), Pyrenean sub-catchments contribute most to outlet runoff. They produced, on average, 56% of the outlet's river flows over the 1950–2005 period, although only covering 11% of the total area of the Ebro basin.

DATA AND METHODS

Data

Daily precipitation (P) and temperature (T) data were provided by the Spanish meteorological agency (Agencia Estatal de Meteorología, AEMET). Data from over 1000 measurement stations were available, however, the period covered by each station and the rate of gaps in the series were very variable. Precipitation and temperature stations were selected in order to have homogeneous spatial distribution of as much data as possible for the 1969–2002 period, using data with a minimum of gaps in the daily series. Two selections were made for both precipitation and temperature stations: stations presenting fewer than 30% and 50% daily gaps per decade or on the whole period, for precipitation and temperature data respectively. These criteria enabled 225 or 364 precipitation gauges (Fig. 1(a)) and 180 or 394 temperature stations (Fig. 1(b)) to be selected.

Altitudes were extracted from the DEM SRTM (V 4.1) and daily river flows were provided by the CEDEX (Centro de Estudios y Experimentación de Obras Publicas). Seven sub-catchments (representing 37% of Ebro's mean annual discharge at the outlet) with little or no regulation by dams and good data availability were chosen (Fig. 1). These basins all include altitudes over 1000 m a.m.s.l., which allowed the influence of altitudinal gradients to be tested in data interpolation.

METHODS

The method consists of two steps: (a) determining the precipitation and temperature altitudinal gradients in order to interpolate the daily grids; and (b) testing the sensitivity of water resources modelling to climatic inputs and rating results.

(a) determining the gradients and interpolating the daily grids Observed P and T gradients were calculated by geographical area (see Fig. 1(b)) at a monthly and annual time step using stations with altitudes over 1000 m a.m.s.l. Data were interpolated on an 8 × 8 km grid at a daily time step, using the inverse distance squared weighted method. The gradients were applied daily based on the altitude difference between the DEM resampled at an 8 km scale and the altitude interpolated from the stations presenting data.

(b) water resource modelling and result scoring The GR4j model (Perrin *et al.*, 2003) was calibrated over the seven selected catchments, with and without the snowmelt module developed by Ruelland *et al.* (2011). This daily conceptual model uses four parameters without the snowmelt module (this module relies on three additional parameters), as well as precipitation (P), evapotranspiration (PE) and temperature (T) input data. PE was calculated using a simple formula depending on mean T and extra-terrestrial radiation (Oudin *et al.*, 2005).

An automatic calibration of the model was carried out on each basin and for each tested dataset, over the 1971–2002 period with a 2-year warm-up (1969–1970). The model was calibrated using a multi-objective function (F_{agg}) proposed by Ruelland *et al.* (2012) and aggregating a daily Nash-Sutcliffe efficiency (NSE) index, a cumulative volume error and a mean annual volume error. Automatic calibration followed three steps. (1) A random draw of 3000 parameter combinations in the model parameter space (four or seven parameters) made it possible to identify the best parameter set that was then used as the starting point of a nonlinear Rosenbrock (1960) optimization of the production function parameters, followed by a second Rosenbrock optimization of the transfer function parameters; (2) 100 repetitions of the first step allowed the best performing combination of parameters to be identified; and finally, (3) a local refining of the parameters around the optimum, using Nelder and Mead's simplex method (1965), was used to optimize all the parameters based on the combination chosen at step (2).

Results were rated using a ranking of the datasets on each basin according to the F_{agg} values, the lowest value being attributed the best score. A global score was attributed to each dataset by adding up scores from all basins.

A second score was used: on each basin, each dataset was attributed a score calculated by averaging the difference in F_{agg} values between the given dataset and all other datasets. This mean value showed whether choosing one dataset over the others improved (negative value) or deteriorated (positive value) the water resources evaluation on each basin. Finally these values were averaged for each dataset over the seven basins.

RESULTS

Determining the gradients and interpolating the daily grids

For precipitation stations, only an annual gradient (2.14 mm/1000 m) could be determined in the Pyrenees. It was applied for grid cells with altitudes above 1000 m a.m.s.l. No gradient was found in the Cantabrian and Iberian ranges. The selection of this gradient and two sets of precipitation gauges (225 and 364) results in four precipitation grids, where the two selections of gauges were interpolated with or without the annual gradient.

For temperature stations, the Pyrenean range annual gradient is close to the average lapse rate of 6.5°C/1000 m, while the Cantabrian range gradient is lower (Fig. 2). The Iberian annual gradient is almost insignificant as confirmed by the monthly variations shown in Fig. 2. The mean monthly gradient of the Pyrenean range is higher in summer, unlike the Cantabrian one. As a result, six different grids of temperatures that associate the two selections of temperature stations (180 or 394, Fig. 1) and the three interpolation modes (without gradient, with annual gradient, with monthly gradient) were used.

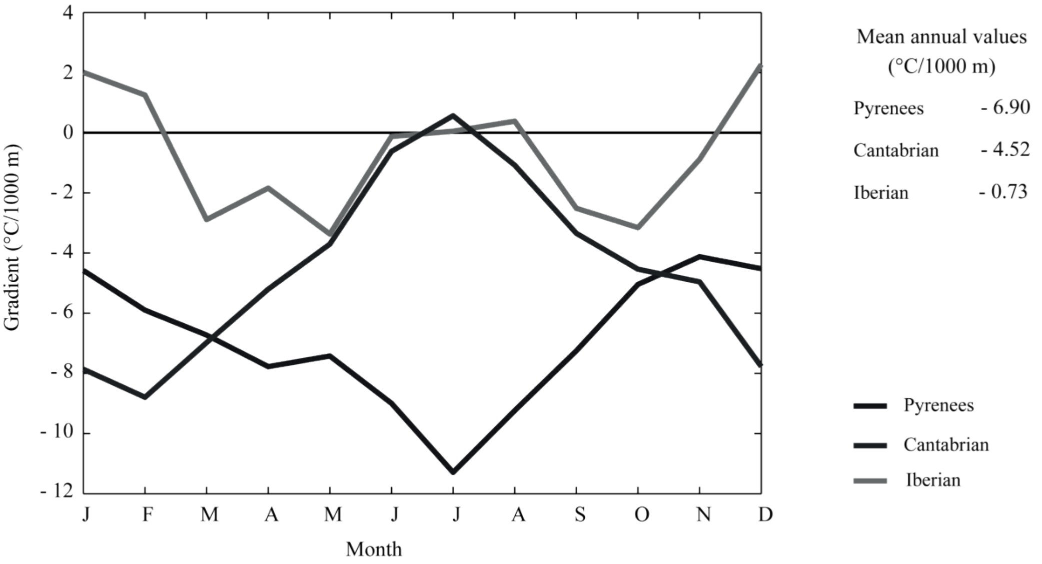

Fig. 2 Observed annual and monthly temperature gradients by geographical zone.

A total of 24 climatic datasets were then built by combining these four precipitation grids and six temperature grids.

Water resources modelling and result scoring

Calibration over the 1971–2002 period was thus carried out for the 24 datasets and for the seven selected basins with the GR4j model run with or without a snowmelt module. Results show that the snowmelt module is needed for all test basins, except for the Arga basin (Fig. 1) that has only 8.8% of area above altitudes of 1000 m a.m.s.l. Modelling reproduces water resources accurately on all basins, with volume errors always approaching zero and NSE values around 0.75 for the best dataset.

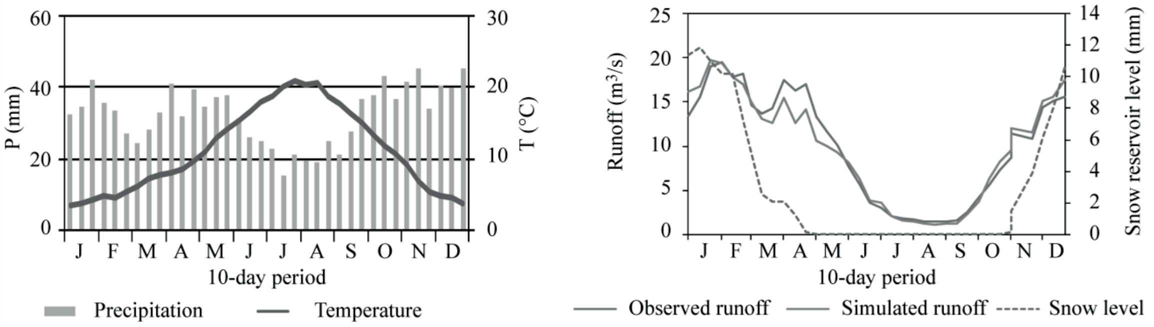

Fig. 3 Mean interannual 10-day values for the Esca basin over 1971–2002 with the best performing dataset (P364_wa_T394_wo).

Table 1 shows rankings for all basins considering the two ways of scoring. The table is sorted according to the second scoring method, the best performing dataset being in the first row (in bold). The last column indicates the dataset's ranking with the first scoring method (FSM). The best performing dataset comes out first with both methods of scoring. Datasets using the maximum number of stations (i.e. including significant gap periods) are often top ranked in comparison with datasets with fewer stations. In the 12 first rows (datasets that do not worsen the modelling compared to all other datasets), precipitation datasets based on 364 stations occur seven times, as well as temperature datasets based on 394 stations. Almost the same occurrence arises for datasets using altitudinal gradients (seven times over 12 with altitude for precipitation datasets and eight times over 12 for temperature datasets). The best performing dataset is the only one improving hydrological modelling for all basins. Considering all these criteria, two datasets can be considered as reference datasets for the Ebro basin: P364_wa_T394_wo and P364_wo_T394_wa_an (first and third row in Table 1).

Figure 3 gives an example of mean interannual 10-day values over the 1971–2002 period of the climate forcing and water resources modelling for the Esca basin with the best performing dataset, showing the good performance of the model.

Table 1 Ranking of all datasets over the seven test basins – datasets are named following this rule P*xxx*_w*y*_T*zzz*_w*y*_*t* where P*xxx* is the number of precipitation stations used for interpolation, w*y* where *y* can be *a* or *o* respectively for *w*ith *a*ltitude and *w*ith*o*ut altitude, T*zzz* is the number of temperature stations used for interpolation, *t* can be *an* or *mo* respectively for *an*nual gradient or *mo*nthly gradient for temperature interpolation. Dataset in bold is the best performing dataset according to this ranking. Negative values (*italic*) show improvements in modelling compared to all other datasets.

Dataset/Basin	1	2	3	4	5	6	7	Mean	FSM
P364_wa_T394_wo	***–11%***	***–1%***	***–7%***	***–1%***	***–3%***	***–14%***	***–14%***	***–7%***	**1**
P225_wa_T394_wo	3%	1%	5%	0%	*–8%*	*–20%*	*–14%*	*–5%*	6
P364_wo_T394_wa_an	*–13%*	*–1%*	*–7%*	0%	2%	6%	*–20%*	*–5%*	2
P225_wa_T394_wa_mo	10%	1%	7%	0%	*–7%*	*–20%*	*–14%*	*–3%*	12
P364_wo_T180_wa_an	*–8%*	0%	*–7%*	*–1%*	4%	9%	*–18%*	*–3%*	4
P225_wa_T180_wo	10%	1%	5%	*–1%*	*–6%*	*–16%*	*–14%*	*–3%*	8
P364_wa_T394_wa_mo	*–4%*	*–1%*	*–5%*	*–1%*	*–3%*	*–14%*	10%	*–2%*	3
P364_wa_T180_wo	*–6%*	0%	*–6%*	*–1%*	*–2%*	*–8%*	10%	*–2%*	7
P225_wo_T394_wa_an	2%	1%	6%	1%	*–1%*	*–1%*	*–18%*	*–2%*	14
P225_wo_T180_wa_an	6%	1%	6%	0%	2%	1%	*–16%*	0%	16
P364_wo_T394_wo	*–11%*	*–1%*	*–8%*	1%	8%	16%	*–5%*	0%	10
P364_wa_T180_wa_mo	4%	0%	*–4%*	*–1%*	*–1%*	*–9%*	12%	0%	11
P364_wo_T394_wa_mo	*–4%*	*–1%*	*–5%*	1%	8%	15%	*–5%*	1%	15
P225_wa_T180_wa_mo	16%	1%	8%	*–1%*	*–5%*	*–16%*	11%	2%	20
P225_wo_T394_wo	4%	1%	5%	2%	4%	6%	*–3%*	2%	19
P364_wo_T180_wo	*–6%*	0%	*–7%*	0%	10%	19%	3%	3%	17
P225_wo_T394_wa_mo	10%	1%	7%	2%	5%	6%	*–4%*	4%	22
P364_wa_T394_wa_an	*–12%*	*–1%*	*–6%*	*–1%*	*–5%*	12%	41%	4%	5
P364_wo_T180_wa_mo	4%	0%	*–4%*	0%	11%	19%	3%	5%	21
P225_wo_T180_wo	7%	1%	7%	1%	7%	10%	3%	5%	24
P364_wa_T180_wa_an	*–8%*	0%	*–4%*	*–2%*	*–4%*	13%	42%	5%	9
P225_wo_T180_wa_mo	17%	1%	8%	1%	7%	10%	2%	7%	23
P225_wa_T394_wa_an	2%	1%	6%	0%	–9%	5%	44%	7%	13
P225_wa_T180_wa_an	5%	1%	6%	*–1%*	*–7%*	7%	43%	8%	18

DISCUSSION AND CONCLUSION

Considering the context of integrated modelling of the Ebro hydro-system and the need to have a long-term representative climatic dataset of this large basin, a method has been proposed to determine which climatic dataset would be selected for the ongoing project.

Some limitations can be discussed. The first one concerns equifinality in model parameters, especially with the snowmelt module adding three parameters to the four existing parameters. This issue was addressed using a robust calibration method; however, compensations may persist. The fact that an altitudinal gradient is needed either for interpolating precipitation, or for interpolating temperature (rows 1 and 3 of Table 1) is a demonstration of these compensations. The snowmelt module has a threshold temperature parameter controlling snow melt. When no gradient is used, this threshold is higher than when an altitudinal gradient is used but parameters are able to compensate this interpolation shortcoming.

The second limitation concerns scoring methods that could be considered questionable or biased. In fact, considering the goal of the climatic grid obtained, this scoring method is well adapted. The interpolation method takes into account the best available information on stations and altitudes of the basin, and grids are tested on the more contributory basins. Gradients were applied on parts of the basin with altitudes over 1000 m a.m.s.l. and for other parts, no gradient was applied. Other uses of the grid in the integrated chain are not affected, especially in the central valley of Ebro where agricultural demands will be calculated with the chosen climatic dataset. In addition, the Noguera Pallasera and Segre basins (Fig. 1), where the density of precipitation and temperature stations is lower, have the highest improvement in water resources modelling when using gradients. Modelling of water resources in upper basins of the Ebro is satisfactory using the GR4j model with a snowmelt module. This snowmelt module is needed for these basins, except for the Ara basin which has only 8.8% of its area above 1000 m a.m.s.l. and does not receive much snow during winter.

Other investigations will be conducted in the project with the two datasets "P364_wa_T394_wo" and "P364_wo_T394_wa_an" to determine sensitivity of all components of the integrated modelling chain to the climatic interpolated grid. These two datasets have gradients either for precipitation or for temperature interpolations. Modelling of water resources has shown its ability to adapt its parameters to accurately simulate observed runoff. Other components, such as water demand estimation or the dam module, could be more sensitive to precipitation or temperature gradients. This would be another criterion to select the reference climatic dataset.

In conclusion, the method elaborated in this paper enabled the choice of two datasets for the 1969–2002 period over the Ebro catchment for integrated modelling of this large and complex hydro-system. Modelling of water resources in upper basins needs to take into account snowmelt. Ongoing work will allow further discrimination of the datasets.

Acknowledgements This work was carried out as part of the GICC REMedHE project funded by the French Ministry of Ecology, Sustainable Development and Energy for the period 2012–2015. The authors are grateful to the AEMET for having provided the necessary data for this study.

REFERENCES

Alcamo, J., Florke, M. & Marker, M., (2007). Future long-term changes in global water resources driven by socio-economic and climatic changes. *Hydrol. Sc. J.* 52(2), 247–275.

López-Moreno, J. I., Vicente-Serrano, S. M., Moran-Tejeda, E., Zabalza, J., Lorenzo-Lacruz, J. & García-Ruiz, J. M. (2010) Impact of climate evolution and land use changes on water yield in the Ebro basin. *Hydrol. Earth Syst. Sci.* Discuss. 7, 2651–2681.

Milano, M., Ruelland, D., Fernandez, S., Dezetter, A., Fabre, J. & Servat, E. (2012). Facing climatic and anthropogenic changes in the Mediterranean basin: what will be the medium-term impact on water stress? *CR Geoscience* 344, 432–440.

Milano, M., Ruelland, D., Fernandez, S., Dezetter, A., Fabre, J. Servat, E., Fritsch, J.-M., Ardoin-Bardin, S. & Thivet, G. (2013a) Current state of Mediterranean water resources and future trends under climatic and anthropogenic changes. *Hydrol. Sci. J.* 58(3), 498–518.

Milano, M., Ruelland, D., Dezetter, A., Fabre, J., Ardoin-Bardin, S., & Servat, E. (2013b). Modeling the current and future capacity of water resources to meet water demands over the Ebro basin. *J. Hydrol.* (in press). http://dx.doi.org/10.1016/j.jhydrol.2013.07.010.

Nelder, J. A. & Mead, R. (1965) A simplex method for function minimization. *Computer Journal* 7(4), 308–313.

Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F., Loumagne, C. (2005) Which potential evapotranspiration input for a lumped rainfall–runoff model? Part 2 – towards a simple and efficient potential evapotranspiration model for rainfall–runoff modelling. *J. Hydrol.* 303, 290–306.

Perrin, C., Michel, C. & Andreassian, V. (2003) Improvement of a parsimonious model for streamflow simulation. *J. Hydrol.* 279, 275–289.

Rosenbrock, H. H. (1960) An automatic method for finding the greatest or least value of a function. *Computer Journal* 3(3), 175–184.

Ruelland, D., Brisset, N., Jourde, H. & Oyarzun, R. (2011) Modelling the impact of climatic variability on the groundwater and surface flows from a mountainous catchment in the Chilean Andes. In: *Cold Regions Hydrology in a Changing Climate* (Proc. 25th IUGG General Assembly, Melbourne, Australia, 28 June–7 July 2011) (ed. by Daqing Yang *et al.*), 171–179. IAHS Publ. 346. IAHS Press, Wallingford, UK.

Ruelland, D., Ardoin-Bardin, S., Collet, L. & Roucou, P. (2012) Simulating future trends in hydrological regime of a large Sudano-Sahelian catchment under climate change. *J. Hydrol.* 424–425, 207–216.

Vicente-Serrano, S. & López-Moreno, J. I. (2006) Influence of atmospheric circulation at different spatial scales on winter drought variability through a semiarid climatic gradient in Nort-east Spain. *Int. J. Climatol.* 26(11), 1427–1453.