Development of a hybrid model to interpolate monthly precipitation maps incorporating the orographic influence

J. Álvarez-Rodríguez (a), M.C. Llasat (b, c), T. Estrela (d, e)

(a) Tagus River Basin Authority. Spanish Ministry of Energy, Environment and Climate Change. Avda. De Portugal, 81. Madrid 28011. Spain

(b) Department of Applied Physics, Universitat de Barcelona, C/ Martí i Franqués, 1, 08028 Barcelona, Spain

c) Water Research Institute (IDRA), University of Barcelona

(d) Júcar River Basin Authority, Ministry of Food and Fishing, Agriculture and Environment, Av/ Blasco Ibañez 48, 46010 Valencia, Spain

(e) Instituto de Ingeniería del Agua y Medio Ambiente (IIAMA) de la Universitat Politècnica de València, Spain.

Corresponding author at: Tagus River Basin Authority. Spanish Ministry for the Ecological Transition. Avda. De Portugal, 81. Madrid 28011. Spain.

Tel.: +34 91 453 96 43; fax: +34 91 470 03 04

E-mail address: javier.alvarez@chtajo.es (J. Álvarez-Rodríguez)

Abstract

This paper proposes an interpolation model for monthly rainfall in large areas of complex orography. It has been implemented in the Iberian Peninsula (continental territories of Spain and Portugal), Balearic, and Canary Islands covering a territory of almost 600.000 km^2. To do this a dataset that comprises a total number of 11,822 monthly precipitation series has been created (11,042 provided by the Spanish Meteorological Agency and 780 provided by the National Water Resources Information System of the Portuguese Water Institute). The dataset covers the period from October 1940 until September 2005. The interpolation model has been based on the assumption of two different components on monthly precipitation. The first component reflects local and seasonal characteristics and 24 different mean monthly precipitation maps (12) and standard deviations maps (12) compose it. It considers the varying influence of physiographic variables such as altitude and orientation. The second precipitation component reflects the synoptic pattern that dominated each month of the series and it is composed by series of anomalies of monthly precipitation (780). Anomalies have been interpolated by means of ordinary kriging once local spatial continuity was assumed. Gridded maps of each variable have been developed at 200 m resolution following a hybrid methodology that implements two different interpolation techniques. The first technique applies a regression analysis to derive maps depending on altitude and orientation; the second one is a weighting technique to consider the non-linearity of the precipitation/altitude dependence. Cross validation has been applied to estimate the goodness of both techniques. Results show an average annual precipitation of 655 mm/year. Although this figure is only 4% less than the estimate of MAGRAMA (2004), regional and local

differences are highlighted when the spatial distribution is considered. The model constitutes a comprehensive implementation considering the availability of historical records and the need of avoiding slow calculations in large territories.

1 Introduction and objectives

The analysis and validation of interpolation procedures of precipitation is a topic widely discussed in the fields of meteorology and hydrology (Daly et al., 2017; Singh et al., 1995; World Climate Programme, 1985; Linsley et al., 1949). Basic data are precipitation records of rain gauges, particularly when the studies are focused on historical periods prior to the development of remote observation techniques (radar and satellite). Due to the scarcity of records in areas where the variability of precipitation is greater (Lloyd, 2005), precipitation estimation is carried out considering the influence of physiographic factors and the spatial continuity of precipitation, combining statistical and experimental methodologies (Hanson, 1982), as well as physically based models (Barstad et al., 2007; Rotunno and Ferretti, 2001). Linear or multivariate regression models are used to construct statistical relationships between precipitation and some physiographic variables such as altitude, orientation, slope of the terrain, distance to water masses or altitude of nearby mountainous areas. These factors are directly related to the triggering effect and a forced uplift when wind direction and terrain's slope interact. Besides, the influence of orography is also reflected in the shield effect and in the driving effect of humid air masses through a complex topography (Bookhagen and Burbank, 2006; Barros et al., 2004; Dhar and Nandargui, 2004; Marquínez et al., 2003; Hay et al., 1998).

The methods of interpolation have been classified between deterministic or stochastic, although the analysis of some reveals conceptual similarities. The stochastic approach shows a formal definition to deal properly with uncertainties of record measurement or those derived from the complexity of the physical processes involved in precipitation generation mechanisms. However, variables such as precipitation are not stationary and depend on a high number of local non-stationary factors. The elementary predictive variable in the interpolation schemes is distance to available records. Interpolation methods use it not only explicitly, but also through the selection of records and the formulation of measures of spatial continuity. Location allows the definition of altitude, orientation, slope, etc. to be used as predictive variables.

In addition to the study of physiographic variables influencing precipitation, interpolation models explicitly incorporate the evaluation of the spatial continuity by means of covariances, polynomial structures, splines, variational approach, quadratic function and adjustment criteria such as error and variance minimization (Tobin et al. 2011; Naoum and Tsanis, 2004a and 2004b; Goovaerts, 2000; Martínez-Cob, 1996; Weber and Englund, 1994 and 1992; Tabios III and Salas, 1985; Creutin and Obled, 1982; Gambolati and Volpi, 1979). Although there are a large number of interpolation models, the question of the optimal or the best model cannot be answered straightforwardly. Gómez-Hernández et. al. (2001) concluded that complex models formally capable of integrating different types of relationships and models of continuity in a rigorous manner such as kriging or the variational approach (Mitas and Mitasova, 1988), do not guarantee obtaining better results than those derived from simpler models. A typical example is the Thiessen methodology

(Thiessen, 1911), that filters out redundancies based exclusively on the extrapolation of each record to the closest area (Falivene et al., 2010; Isaaks and Srivastava, 1989). However, the goodness of an interpolation model depends largely on the spatial variability of the precipitation event considered and on the density and representativeness of the ground stations network. It is to say, it depends both on the absence of records in places where the variability of precipitation is greater, as occurs in the mountains and the coast, but also on the redundancy of data recorded at close locations. Furthermore, it depends on the temporal step of the study, considering that the complexity of the precipitation variability increases the shorter the time interval is, and the random component becomes predominant. Particularly, the lack of data in mountainous areas does not make it advisable to use techniques whose parameterization is sensitive to the lack of information.

In spite of this, most studies recommend the use of altitude as the basic variable for interpolation at regional and seasonal scales. This is the case for procedures implemented in the Precipitation-elevation Regressions on Independent Slopes Model (PRISM) to estimate fields of precipitation across conterminous North America (Daly et al., 2017, 2008 and 1994).

Precipitation-elevation regressions were also used in Spain in combination with an Inverse Distance Weighting (IDW) algorithm to create the monthly precipitation maps that were used as input to the distributed hydrological model SIMPA with the objective of analyzing water resources distribution in Spain (MAGRAMA, 2004). To reflect orographic influence and the underestimation of precipitation given, a collation of pseudo precipitation records was then added to original records. Pseudo precipitation records were estimated by linear regression analyzed in certain

Spanish regions (Estrela et al., 1999, Álvarez-Rodríguez et al., 2017). Regions of approximately 5,000 km^2 were delimitated considering windward and leeward location. A criterion to control the accuracy of interpolated precipitation was obtained from the comparison of recorded runoff volume and the precipitation excess. But uncertainties were revealed when considering the quality of flow data and the calculus of base flow, abstractions and direct runoff. Moreover, the procedure followed in MAGRAMA (2004) was considered inadequate and tedious to update the water resources assessment and therefore, the updating of the pseudo precipitation data.

Rainfall-runoff models have been used to estimate natural water resources (unaltered) across the Spanish territory (Álvarez-Rodríguez et al., 2016; MAGRAMA, 2004). On the Iberian Peninsula, moist air masses from the Atlantic Ocean constitute the most important source of precipitation, while the spatial distribution of precipitation is a function of orography and direction of air flow. The influence of the Mediterranean Sea in precipitation occurrence is also important as reflected in the regional change of the seasonal precipitation pattern to maxima occurring in autumn and spring.

Álvarez-Rodríguez et al. (2017) described some basis to improve spatial estimates of rainfall for the Iberian Peninsula and Spanish Islands. They concluded that precipitation over this territory depends on its complex orographic structure and predominant weather types. Altitude and orientation are the main physiographic factors that would help to estimate precipitation. In Spain, precipitation tends to be positively correlated with altitude although this relationship varies depending on seasonality and location. Annual precipitation lapse rates (PLR) were found to range from 0.3 to 1.2 mm/m, reaching 1.5 mm/m in the Northern Iberian Peninsula and diminish at higher altitudes (Álvarez-Rodríguez et al., 2017). This would justify the use of non-linear functions in precipitation-altitude regression analysis as it will be shown in this paper. In coastal areas, large precipitation increments or decrements are found where small differences in altitude are given. Additionally, a source of uncertainty is identified considering that precipitation is mostly recorded at low elevations.

This paper proposes a hybrid model of interpolation at a regional scale that can be used to derive high resolution fields of precipitation over territories with complex orography. The interpolation model assumes two different components on monthly precipitation. The first component reflects local and seasonal characteristics. It is composed by 24 different monthly precipitation maps of means (12) and standard deviations (12). It considers the varying influence of physiographic variables such as altitude and orientation. The second precipitation component reflects the synoptic pattern and it is composed by normalized anomalies derived from monthly precipitation records and monthly means and standard deviations. The model constitutes a comprehensive implementation considering the availability of historical records and the need of avoiding slow calculations in large territories. This model has been applied to estimate monthly precipitation maps of 200 m resolution for the Iberian Peninsula, Balearic and Canary Islands, from October 1940 to September 2005. After the description of the database and data sources, the paper firstly describes the procedure used for the estimation of the monthly precipitation patterns and secondly, the interpolation of the anomalies of the precipitation records. The analysis carried out to validate these procedures is also shown. To conclude, the achievements of the hybrid interpolation model are remarked.

2 Data Sources

2.1 Recorded ground series of rainfall

The databases of ground recorded precipitation were provided by the Spanish Meteorological Office (AEMET) and the National Water Resources Information System of the Portuguese Water Institute (SNIRH-INAG). Spanish data are supplied by AEMET though its Virtual Office at https://sede.aemet.gob.es/. Portuguese data are available at https://snirh.apambiente.pt/. The whole database of monthly precipitation comprised 11,042 ground series from AEMET and 780 ground series from SNIRH-INAG. Although some series comprise records from the 19th century until the hydrological year 2004/05, the selected period is 1940/41-2004/05.

Existing gaps in recorded rainfall series were filled with regression-based data. Basis of the completion model as well as a description of available data may be found in Álvarez-Rodríguez et al. (2017).

2.2 Location, elevation data and derived models

Most of Spanish and Portuguese territories are a part of the Iberian Peninsula (almost 582,000 km^2), which is in southwestern Europe and surrounded by the Atlantic Ocean and the Mediterranean Sea. This research encompasses the Iberian Peninsula and the Balearic Islands in the Mediterranean Sea (5,000 km^2) and the Canary Islands (7,500 km^2) in the Atlantic Ocean, which are influenced by a tropical climate.

A Digital Elevation Model (DEM) has been composed joining Spanish and a Portuguese DEM to derive its main physiographic features as described in Álvarez Rodríguez et al. (2017). Figure 1 shows the Digital Aspect Model (DAM, cell angle at which terrain slope faces, counterclockwise from East) obtained considering relative elevation surrounding each cell of a DEM. This is done by means of the algorithm *r.slope.aspect* implemented in the GRASS-GIS (GRASS Development Team, 2012; Neteler and Mitasova, 2004).

Figure 1. Main Spanish mountain systems and hydrographic catchments are shown over a composition of Spanish and Portugal 200 m resolution DAM. Based on the UTM zone 30 Geographical coordinates the Canary Islands are displaced 500,000 m East and 750,000 m North to encompass the whole geographical territory in a workable layout.

3 The Hybrid Model for Interpolation

3.1 Rationale

The following 5 points are some preliminary requirements adopted for the development of an interpolation model to estimate monthly precipitation maps for the territory with a 200 m resolution:

1. The number of records to be interpolated varies from month to month;
2. The selection of records to be interpolated should consider both the scarcity of records in mountainous areas and the redundancies of records in lower altitudes;
3. Elevation and orientation are the predictive variables and their influences in precipitation vary throughout the territory;
4. The interpolation model should be capable of working with different humid air masses entering the territory and their different interactions with orography;

5. Finally, the time for calculation should be reduced enough considering the need of deriving a whole set of 780 monthly interpolated maps of precipitation from October 1940 to September 2005.

In accordance with these requirements, a hybrid interpolation model based on the decomposition of temporal components used in synthetic series completion and generation procedures has been proposed (Álvarez-Rodríguez et al., 2017; Salas et al., 1980; Fiering and Jackson, 1971). It has been named "hybrid model" because two different interpolation models were implemented for two precipitation components.

Figure 2. Flow chart of methodology

Figure 2 shows a flowchart of the methodology applied. After the compilation of records and completion of gaps in series of precipitation (Álvarez-Rodríguez et al., 2017), monthly statistics of precipitation are estimated. The first component of precipitation is composed by the monthly means and the monthly standard deviations. Being the statistics that represent monthly centrality and variability, it is considered that they represent the local influence on precipitation. The monthly step accounts for seasonality.

The second component of monthly precipitation is represented by the anomalies resulting from normalizing each monthly record of precipitation once monthly means and standard deviations are known. The anomalies vary in time and would be associated with the dominant synoptic circulation pattern each month. The following sections describe in detail the algebra of each component. Regression analysis is applied to derive monthly maps of means and standard deviations, while ordinary kriging after an automated parameterization is applied on anomalies.

3.2 Monthly Components of Centrality and Variability

3.2.1 Estimation of Local Patterns of Precipitation

Local patterns of precipitation were represented by monthly mean and standard deviation maps. Considering seasonal variability reflected in a monthly step, 24 different maps have been obtained by interpolation of monthly means (12 maps) and monthly standard deviations (12 maps) derived from recorded series of precipitation completed previously. Since orographic influence is variable, monthly means and standard deviations were interpolated by means of regression analysis. Altitude was used as a predictor in regression analysis. A regression equation was implemented in each cell of the model. Samples were selected considering the orientation of the place where each rain gauge station is located and distance from the center of a cell to nearby rain gauge stations. Then, given the scarcity of records at higher altitudes, a weighted regression equation was implemented to estimate precipitation to prioritize nearby records close to a cell.

Statistics of recorded monthly rainfall series were calculated for the period ranging between the hydrological years 1970/71 and 1999/00, which is the 30-year period of maximum data availability (Álvarez-Rodríguez et al., 2017). The selection of a unique period would assure homogeneity.

Monthly means and standard deviations were interpolated by a moving regression equation based on altitude but using the orientation of the terrain as a criterion to select the values of the sample to estimate each cell value.

The statistics obtained are georeferenced by means of the coordinates of each rain gauge station. Then a selection of statistics is made for each cell based on distance and orientation. Particularly, those rain gauge stations located over cells whose orientation (DAM of 200 m resolution) is included in the 180º semicircular sector formed by the orientation angle of the estimation cell and a semi-amplitude of ±90º are selected. It has been verified that semi-amplitude of less than 45° reduces excessively the number of records to formulate each regression equation; and larger semi-amplitudes, that is to say, between 45º and 90º, do not cause significant differences to the 90º finally chosen. If a cell's slope is less than 1%, it is considered that the orientation is not meaningful and rain gauge stations were selected depending only on the distance. The maximum search distance from the center of each cell is 100 km, or even larger till a minimum of 12 stations is found. The maximum number of stations for each sample is 18.

Then, a cell precipitation-altitude regression equation is fitted according to a moving weighted regression interpolation model (Lloyd, 2005; Naoum and Tsanis, 2004b; Daly et al., 1994). Each cell-regression equation is fitted by the minimum least squares criteria, independently of the equation fitted in nearby cells.

A simple linear regression equation between altitude and precipitation would involve the extrapolation of PLR estimated at medium and low altitudes where precipitation is mostly recorded. To improve estimations, logarithmic transformations have been used to reduce or extend the scale of the transformed variable.

Four laws have been formulated to be applied considering the more suitable variable to transform (precipitation or altitude) and the positive or negative correlation of precipitation and altitude.

1. Logarithmic transformation of altitude (Eq. 1). It has the property of extending the scale of the variable altitude in its lower levels and of reducing it in medium to high elevations. Therefore, when the altitude-precipitation correlation is

positive, this transformation imposes a convex curvature, which is in accordance with simplified theoretical approaches that describe a decrease in PLRs with altitude due to depletion of available humidity. This transformation is also applied in coastal areas where a negative correlation and a high variability of precipitation with respect to altitude happens. The relationship between altitude and precipitation is then given by Eq. (1):

$$P(X,Y) = a \cdot log[Z(X,Y)] + b \qquad \text{Eq. (1)}$$

where *Z(X,Y)* is the predictive variable in a cell of geographic coordinates *X* and *Y*, *P(X,Y)* is the recorded precipitation in that particular cell, *a* and *b* the parameters of the simple regression equation fitted by minimum least squares.

Then, the criterion to choose this case is that the altitude-precipitation correlation is positive and the average altitude of the sample is lower than the altitude of the cell. That is because it is considered that there are more records at low levels to estimate rain at higher levels. Moreover, this transformation is also applied when the correlation is negative and the average altitude of the sample is higher than that of the cell to be estimated because it is considered that there are more records at higher levels.

2. Logarithmic transformation of precipitation (Eq. 2). This transformation weakens the decrease in precipitation when the altitude-precipitation correlation is negative avoiding the extrapolation of negative PLRs from the coast to the inner territories. This typically occurs in coastal areas. It is also applied with positive PLRs where it is necessary to soften the reduction of rainfall. The relationship between altitude and precipitation is then given by Eq. 2:

$$log[P(X,Y)] = a \cdot Z(X,Y) + b \qquad \text{Eq. (2)}$$

Then, the criterion to choose this case is that the altitude-precipitation correlation is negative and the altitude of the cell is higher than the averaged elevations of the sample. Likewise, this transformation is applied if positive correlation and cell's altitude is lower than the averaged altitudes of the sample. It should be emphasized that the effect of the logarithmic transformation on precipitation is less significant, not only because the sensitivity of the results is lower with reduced precipitation, but also because in areas of low altitude, the density of the precipitation network is generally higher.

Being *z* the predictive variable altitude (*Z (X,Y)*) or its transformed (*log (Z (X,Y))*) in a cell of coordinates X and Y, *p* the variable precipitation (*P (X, Y)*) or its transformed (*log (P (X, Y))*), *i* the indicative sub-index of each statistic of a sample of size *N* (i = 1..N) and w_i the weight given to each statistic, the parameters *a* and *b* of the regression equation are obtained according to Eq. (3).

$$p = a \cdot z + b \qquad a = \frac{\sum_{i=1}^{N} w_i \cdot z_i \cdot p_i - \sum_{i=1}^{N} w_i \cdot z_i \cdot \sum_{i=1}^{N} w_i \cdot p_i}{\sum_{i=1}^{N} w_i \cdot z_i^2 - \left(\sum_{i=1}^{N} w_i \cdot z_i\right)^2}$$

$$b = \sum_{i=1}^{N} w_i \cdot p_i - \frac{\sum_{i=1}^{N} w_i \cdot z_i \cdot p_i - \sum_{i=1}^{N} w_i \cdot z_i \cdot \sum_{i=1}^{N} w_i \cdot p_i}{\sum_{i=1}^{N} w_i \cdot z_i^2 - \left(\sum_{i=1}^{N} w_i \cdot p_i\right)^2} \qquad \text{Eq. (3)}$$

The weight w_i assigned to station *i* is calculated with an inverse distance function of exponent *h* (Eq. 4). *h* takes the value of 2 after verifying that no significant differences are obtained between the results obtained with the frequent values, 1, 2 or 3. The distance d_j from i to j rain gauge station is calculated from the center of the coordinate cell (*X,Y,Z*) to each one of the *N* data selected (X_j, Y_j, Z_j).

$$w_i(X,Y) = \frac{\frac{1}{d_i^h(X,Y,Z)}}{\sum_j^N \frac{1}{d_j^h(X,Y,Z)}} \qquad d_j = \sqrt{\left(X_j - X\right)^2 + \left(Y_j - Y\right)^2 + \left(Z_j - Z\right)^2} \quad \text{Eq. (4)}$$

Considering the interpolation in the Iberian Peninsula, Balearic and Canary Islands, a number of about 15,000,000 cells and, consequently, regression equations were fitted per month. Figure 3 shows 4 mean monthly precipitation maps representative of the 4 seasons of a year. They were obtained from the monthly means of 30 years of precipitation records between the hydrological years 1970/71 and 1999/00.

Figure 3. Monthly mean precipitation maps of November (a), February (b), May (c) and August (d) considering the 30 years period from 1970/71 until 1999/00

Monthly mean and standard deviation maps may be interpolated following the methodology shown previously. But once interpolated means are calculated, maps of standard deviations may benefit both from the high correlation coefficients achieved between the monthly means and monthly standard deviations and from the softened spatial variability across the territory shown by their ratio, the monthly coefficient of variation, CV (Álvarez-Rodríguez et al., 2017). The softened spatial variability is a useful property to interpolate the 12 monthly CVs if assuming a local stationarity and implementing an ordinary kriging model (OK) based on an omnidirectional semivariogram (Isaaks and Srivastava, 1989). Figure 4 shows the monthly standard deviation maps obtained as a product of mean monthly precipitation maps by the monthly coefficient of variation estimated by OK.

Figure 4. Monthly Coefficient of Variation (CV) (a) and Standard Deviation (SD) (b) Maps of November

3.2.2 Validation of Mean Monthly Maps

A topic for discussion is the validation procedure followed to determine the goodness of the precipitation maps obtained. A basic criterion is the comparison with previous estimations. However, precedent estimations are also influenced by several sources of errors. The present methodology improves the method based solely on distances to nearest records that was applied in the MAGRAMA report (2004) as interpolation procedure. MAGRAMA (2004) was the starting point of this present work, aimed to develop a new model not only dependent on distances. Likewise, the Digital Climatic Atlas of the Iberian Peninsula published by Ninyerola et al. (2007 and 2005) and the Iberian Climatic Atlas published by AEMET (2011) were not available in a digital format. However, the visual comparison with the AEMET (2011) most recent estimation allowed concluding the agreement between the distributions of the monthly means of precipitation obtained.

Cross validation is a technique used to estimate the error of interpolation. A measurement of error is calculated from the comparison of each record against the value resulting from the interpolation using the rest of the records (Falivene et al., 2010; Isaaks and Srivastava, 1989). Figure 5 shows two scatterplots of mean monthly precipitation recorded in December and that estimated by the moving weighted regression interpolation procedure described in this paper, once the logarithmic transformations and the weighting technique have been applied. The scatterplots of the rest of the 11 months are similar, although quantities of precipitation vary. The first scatterplot (left) represents the dispersion of the complete sample of records in the Iberian Peninsula. The second one (right) shows the dispersion of a sample corresponding to stations located at an altitude of more than 1,600 masl (Figure 5).

Figure 5. Scatterplots of recorded and interpolated monthly precipitation (December) considering a linear regression and a weighted linear regression on transformed precipitation. The whole dataset in the Iberian Peninsula (a); records over a 1,600 m high (b)

Figure 5 shows that both the linear regression method and the transformed-weighted method underestimate monthly precipitation at higher locations, particularly over 300 mm of precipitation. However, this bias is lower at higher elevations when the transformed-weighted method is applied. Table 1 shows the mean relative errors (MRE) obtained for the Iberian Peninsula when the linear regression (LR) and the regression with logarithmic transformation and weighting (WR) are applied. The MRE is calculated based on the relative error (RE) of the series *i*, where *i = 1..N* where *N* is the total number of series (observatories) in the sample (Eq. 5).

$$RE_i = \frac{P_i^{interpolated} - P_i^{recorded}}{P_i^{recorded}}\% \qquad MRE = \sum_{i=1}^{N} \frac{RE_i}{N} \qquad \text{Eq. (5)}$$

Table 1. Monthly MRE (%) obtained for the Iberian dataset considering Linear Regression (LR) estimation and the Logarithmic Transformation and Weighted Regression (WR)

Based on the above, the logarithmic transformation and data weighting reduces the bias at high levels, in spite of the uncertainties due ultimately to the scarcity of information at the highest levels, whatever the chosen procedure is. The improvement obtained in areas of higher altitudes is considered to be related with the management of the PLR variability depending on altitude. The weighting technique applied gives more weight to nearest data and correct the higher PLR estimated at lower altitudes. So, this conclusion validates the use of the transformation and weighting techniques.

3.3 Monthly Anomalies of Recorded Rainfall

3.3.1 Definition and Estimation

The moving weighted regression interpolation procedure could also be applied in a monthly step from October 1940 to September 2005. Then, a total number of 780 monthly precipitation maps would have been obtained. But the computation time was considered too long. The hybrid model proposed in this paper only uses the moving weighted regression interpolation model to estimate 12 maps of monthly mean patterns and another 12 of standard deviations. Then it is proposed to implement a second model to interpolate the anomalies derived from each monthly precipitation record and the calculated statistics. Considering the applicability to large sets of maps, the reduction of the computational effort is a basic criterion when selecting an interpolation procedure.

As previously defined, monthly anomalies would represent the variability given by synoptic circulation patterns in a particular month of a year with respect to local variability characterized by monthly means and standard deviations. Monthly anomalies are calculated using the standardization formula (Eq. 6). Given a recorded series of precipitation and being μ_i and σ_i the mean and standard deviation at month i, the anomaly, $r_{i,j}$, of precipitation for the i month and j year, $P_{i,j}$, is given by Eq. 6.

$$r_{i.j} = \frac{P_{i,j} - \mu_i}{\sigma_i} \quad \text{Eq. (6)}$$

Then monthly anomalies from October 1940 to September 2005 were calculated for each rain gauge. Figure 6 shows the histogram of the complete set of anomalies of the Iberian Peninsula in November 1984. They are supposed to reflect a synoptic pattern being dominant in a particular month of a year. A similar histogram may be obtained for each month of the period considered. Generally speaking, the histograms show a central body of values with normal appearance and symmetry around the central value, but there are also cases with a positive bias as a consequence of the autumnal precipitation maxima in the Eastern areas of the Peninsula (Figure 6). Some other histograms show negative extremes derived from the transformation of precipitation values close to zero and low monthly deviations. This is usually the case in the summer.

Figure 6. Histograms of Precipitation Anomalies for November 1984 (a), February 1985 (b), May 1985 (c) and August 1985 (d)

Kriging and the analysis of the spatial continuity of data is used to interpolate maps of anomalies. They have a structural component of continuity that would be represented by means of an omnidirectional semivariogram. If monthly sample of anomalies show asymmetry and bias, then a Box-Cox transformation is applied to facilitate the interpolation and to reduce the sensitivity to the extremes. The well-known Box-Cox transformation (Eq. 7) depends on a parameter λ fitted to minimize the coefficient of asymmetry of a sample.

$$\lambda \neq 0 \Rightarrow y = \frac{x^{\lambda}-1}{\lambda} \lambda = 0 \Rightarrow y = ln(x) \qquad \text{Eq. (7)}$$

3.3.2 Interpolation of Anomalies

The geostatistical analysis of monthly anomalies was carried out using the statistical software R and the *gstat* package (Gräler et al., 2016, R Development Core Team, 2008, Pebesma, 2004). This software implements an automatically fitted semivariogram model using ordinary least squares criteria. Then, a set of monthly semivariograms is obtained for the period 1940/41-2004/05 in each of the 3 regions considered, Iberian Peninsula, Balearic and Canary Islands. The chosen semivariogram function is the exponential one. Parameters representing the spatial continuity are the nugget effect, the sill and the range (Figure 7).

Figure 7. Semivariogram of Iberian Peninsula anomalies of November 1984 fitted to an exponential one

Most semivariograms behave in the same way as the one shown in Figure 7. Nevertheless, some others show greater variability and oscillations. Table 2 shows the median of each of the 3 parameters (nugget, sill and range) of the exponential semivariograms fitted from October 1970 to September 2000. Sill and range values seem to fit higher values during the rainy season that, in the Mediterranean area correspond to spring and autumn, while in the Atlantic it extends from autumn to spring.

Table 2. Median of semivariogram parameter values found for the collation of anomalies obtained from October 1970 to September 2000

Ordinary kriging (OK) was used to interpolate anomalies taking into account that this model may operate with local stationarity. It also weights data to diminish the influence of redundancies (Isaaks and Srivastava, 1989). Finally, OK shows a conceptual equivalence with other deterministic models such as the variational approach by means of regularized spline with tension (RST) (Mitas and Mitasova, 1988). The next section evaluates the OK benefits in respect of the simpler but much faster IDW as well as the similarities given by a RST approach.

3.3.3 Interpolation Efficiency

The goodness of the interpolation methods applied on anomalies has been evaluated through the loss of efficiency obtained when the available data is reduced. Thus, a percentage of rain gauge stations (i.e., their series of anomalies) was randomly selected and removed from the original sample. Then, the available set of monthly maps is interpolated and an efficiency coefficient map is obtained. The efficiency coefficient is then associated to the interpolation model used. Eq. 8 describes the formula used to obtain the efficiency coefficient in each cell.

$$CE = \frac{\sum_{i=1}^{n}(r_i - m_r)^2 - \sum_{i=1}^{n}\left(s_i - r_i\right)^2}{\sum_{i=1}^{n}(r_i - m_r)^2} \qquad \text{Eq. (8)}$$

where s_i are the mean monthly maps of anomalies for each *i* year (from 1 to *n*) derived from the use of an interpolation model. Taking into account that 3 different interpolation models are used (IDW, RST and OK), 3 different sets of maps are estimated. The percentages of reduction from the complete set of rain gauge stations are 60%, 40% and 20%. That is to say that the 3 interpolation models are applied to 3 different sets that are equivalent to the use of 40%, 60% and 80% in respect of the complete set of series. r_i is the mean monthly map of anomalies for each year *i* interpolated by means of IDW, RST and OK, but for the whole set of series (i.e., a 100% of availability); m_r is the mean map of r_i.

Figure 8. Efficiency considering the interpolation method and a reduction in available records

Figure 8 shows the averaged efficiency coefficient dependent on the interpolation model (OK, RST and IDW) and on the availability from the complete sample of series. The faster loss of efficiency of the IDW is highlighted in respect of OK and RST models. Thus, improvements in efficiency are linked to modeling the spatial continuity as done in OK and RST models.

3.4 Hybrid Interpolated Monthly Precipitation Maps

Figure 9 shows a sequence of monthly rainfall maps interpolated during the hydrological year 1984/85. These maps have been obtained by combining the monthly maps of means and standard deviations, which would represent the local anomalies, and the precipitation anomalies related to synoptic atmospheric circulation. The "hybrid" model is finally composed by the use of the model of moving weighted regression on transformed precipitation (presented in 3.2.1) and by the OK to interpolate the precipitation anomalies (presented in 3.3.2).

Figure 9. Monthly precipitation maps interpolated by means of the hybrid model combining a Moving Weighted Regression on transformed Precipitation for mean and standard deviation and Ordinary Kriging for anomalies. Maps from the hydrological year 1984/85: November 1984 (a), February 1985 (b), May 1985 (c) and August 1985 (d)

4 Results and discussion

A mean annual precipitation map is derived from the monthly set of estimates (Figure 10). Thus, the average annual precipitation is 655 mm/year. Although this figure is only 4% less than the estimated precipitation map of MAGRAMA (2004), regional and local differences are highlighted when the spatial distribution is considered.

Figure 10. Mean annual precipitation maps (1940/41-1995/96) obtained by MAGRAMA (2004) (a) and by the implementation of the hybrid method (b)

The spatial distribution of precipitation maps in MAGRAMA (2004) is then attenuated when compared to results obtained by the hybrid model where the orographic structure is clearly remarked (Figure 10). Additionally, the isolation of certain data is reflected in the IDW methodology followed by MAGRAMA (2004) by means of rounded artifacts of interpolated precipitation while the help of orographic influence and the modeling of the spatial continuity clearly improve the results obtained with the hybrid model (Álvarez-Rodríguez, 2011). Furthermore, the spatial comparison of the precipitation map obtained in MAGRAMA (2004) and the one presented in this paper highlights several regional/local differences in precipitation amounts. Mostly, in areas (see Figure 1) such as the upper Ebro River Basin and its left margin (Pyrenees) as well as in the Cantabrian region of the Iberian Peninsula (Álvarez-Rodríguez, 2011). Finally, the hybrid model has the advantage of managing precipitation records in a systematic way avoiding the time and expertise needed in MAGRAMA (2004).

A topic for discussion is constituted by the effect of resampling in estimated precipitation. Greater resolution was resampled to 200 m to derive DEM in most of the territory. The altitude of the rain gauge was used for the regression analysis, but each cell altitude was used to estimate the precipitation of every cell. But a lot of variability exists in mountainous areas that would influence the representation of cell altitude and subsequently on estimated precipitation in every cell.

The set of monthly maps was implemented in a distributed hydrological model to estimate water resources in natural regime in Spain. This fact implied the need to reparametrize the hydrological model, opening the possibility of including physiographic factors such as soil textures or slopes in the calibration of the maximum soil storage capacity (Álvarez-Rodríguez et al., 2016). The need for a parameterization emphasizes the importance of the spatial distribution of precipitation and how the uncertainty is transferred to parameters of a hydrological model.

5 Conclusions

A hybrid model to improve the estimation of monthly precipitation distribution in great areas of complex orography has been proposed. It combines two interpolation models applied to each one of the two main different components distinguished in the precipitation.

Firstly, monthly means and standard deviations were interpolated by a moving regression equation based on altitude but using the orientation of the terrain as a criterion to select the values of the sample to estimate each cell value. The regression also incorporates the use of transformation functions (logarithms of precipitation or altitude) and weights as a function of the distance to prioritize nearby information avoiding the overestimation at higher altitudes using records of lower altitudes. Monthly maps of standard deviations can be either obtained by inferring regression equations based on altitude and orientation or by the product of maps of monthly variation coefficients by the already estimated maps of monthly means. Due to the high correlation between means and standard deviations, its ratio (coefficient of variation) does not show the spatial variability shown by the monthly means and can be assumed locally stationary, which facilitates its estimation in large territories through interpolation procedures such as ordinary kriging (OK). This first component provides information about the seasonal variability of precipitation at local scale.

The second component of precipitation is constituted by the anomalies, that are mainly related with synoptic situations that affect at regional scale. Its bias is corrected to reduce the asymmetry of each sample and then interpolated in a monthly step using an OK. A comparison between averaged efficiency coefficients derived from OK, variational approach (RST) and inverse distance algorithm (IDW) revealed how the implementation of a continuity structure in an interpolation model benefits the results. It means that methods working with a spatial continuity structure (OK and RST) are adequate to represent the precipitation in a complex terrain, obtaining accurate estimations even when the loss of observatories reached a 60%. Then, structures embedded in usual interpolation methodologies as OK and RST may replace a high percentage of redundancies existing in a meteorological network. In spite of this it is therefore necessary to insist on the need to improve the availability of precipitation records at higher altitudes in order to reduce the uncertainty of precipitation estimation.

The hybrid model presented in this paper has the advantage of reducing the computational time. An advantage of the linear regression method is its conceptual simplicity, while accounting for the non-linear relationship between precipitation and altitude. However, when it must be repeatedly applied to large territories for a great number of precipitation maps needed to subsequently force a hydrological model, the long time for calculation make its use makes the method unsuitable. Therefore, the hybrid approach limited its use to the estimation of the 24 maps of the monthly means and standard deviations. Next, anomalies associated to regional components are interpolated by means of OK after parameterizing the monthly semivariograms. As seen, OK and RST account for spatial continuity, which is variable from month to month and can be applied to a variable number of records without a significant loss of information about precipitation performance.

The main advantage of the proposed methodology relies on that it has been composed considering advantages of different procedures in order to represent precipitation both over large territories and complex terrain. Regression analysis considers precipitation/altitude relationships following usual procedures reviewed to implement the non-linearity in a straightforward way. Anomaly interpolation takes the advantage of automatic parameterization and methodologies capable of implementing the spatial continuity.

6 Acknowledgements

The Spanish Meteorological Agency (AEMET), the Portuguese Water Institute and the people maintaining HIDRO database of the Center of Hydrographic Studies of CEDEX as well as the Spanish Water Directorate for promoting the Water Resources studies in Spain for which this work was developed. This work has been partially supported by the Spanish Project HOPE (CGL2014-52571-R) of the Ministry of Economy, Industry and Competitiveness. The authors would also like to thank Maria Serneguet Belda, from the Mediterranean Network of Basin Organizations, for her constructive review of the English writing.

7 References

AEMET. 2011. Atlas Climático Ibérico. (Iberian Climate Atlas) VV.AA. Agencia Estatal de Meteorología. Ministerio de Medio Ambiente. ISBN: 978-84-7837-079-5. URL: http://www.aemet.es/documentos/es/conocermas/publicaciones/Atlas-climatologico/Atlas.pdf Last Access: 14/02/2018

Álvarez-Rodríguez J, Llasat MC and Estrela T. 2017. Analysis of geographic and orographic influence in Spanish monthly precipitation. Int. J. Climatol. doi:10.1002/joc.5007

Álvarez-Rodríguez J, Barranco Sanz LM, García Bravo N, Potenciano de las Heras Á, Villaverde Valero JJ. 2016. La Evaluación de Recursos Hídricos en España (Water Resources Assessment in Spain). (In Spanish), July/2016, 380 p. Centre for Hydrographic Studies of CEDEX ISBN/EAN: 9788477905783

Álvarez-Rodríguez J. 2011. Estimación de la distribución espacial de la precipitación en zonas montañosas mediante métodos geoestadísticos (Analysis of spatial distribution of precipitation in mountainous areas by means of geostatistical analysis). PhD thesis. Polytechnic University of Madrid, Higher Technical School of Civil Engineering

Barros AP, Kim G, Williams E and Nesbitt SW. 2004. Probing Orographic Controls in the Himalayas During the Monsoon Using Satellite Imagery. Nat. Hazards Earth Syst. Sci., 4, 29-51

Barstad I, Grabowski W and Smolarkiewicz P. 2007. Characteristics of large-scale orographic precipitation: evaluation of linear model in idealized problems. J. Hydrol., 340, 78-90

Bookhagen B and Burbank DW. 2006. Topography, relief and TRMM-derived rainfall variations along the Himalaya, Geophys. Res. Lett., 33

Creutin JD and Obled C. 1982. Objective Analyses and Mapping Techniques for Rainfall Fields: An Objective Comparison. Water Resour. Res., 18(2), 413-431

Dhar ON and Nandargi S. 2004. Rainfall distribution over the Arunachal Pradesh Himalayas. Weather, 59, 155-157

Daly C, Slater ME, Roberti JA, Laseter SH and Swift LW. 2017. High-resolution precipitation mapping in a mountainous watershed: ground truth for evaluating uncertainty in a national precipitation dataset. Int. J. Climatol. doi:10.1002/joc.4986

Daly C, Halbleib M, Smith JI, Gibson WP, Doggett MK, Taylor GH, Curtis J and Pasteris PP. 2008. Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. Int. J. Climatol., 28, 2031-2064

Daly C, Neilson RP and Phillips DL. 1994. A Statistical Topographic Model for Mapping Climatological Precipitation over Mountainous Terrain. J. Appl. Meteor., 33, 140-158

Estrela T., Cabezas F. and Estrada, F. 1999. La evaluación de los recursos hídricos en el Libro Blanco del Agua en España (Water Resources Assessment in the Water in Spain Book). Revista de Ingeniería del Agua, 6(2), 125–138, 1999

Falivene O, Cabrera L, Tolosana-Delgado R and Sáez A. 2010. Interpolation algorithm ranking using cross-validation and the role of smoothing effect. A coal zone example. Computers and Geosciences, 36 (4), 512-519

Fiering MB and Jackson BB. 1971. Synthetic Stream Flows, American Geophysical Union, Washington D.C., 98 p.

Gambolati G and Volpi G. 1979. A conceptual Deterministic Analysis of the Kriging Technique in Hydrology. Water Resour. Res., 15(3), 625–629

Gómez-Hernández J, Cassiraga E, Guardiola-Albert C, Álvarez-Rodríguez J. 2001. Incorporating Information from a Digital Elevation Model for Improving the Areal Estimation of Rainfall. In GeoENV III: Geostatistics for Environmental Applications. Monestiez, P., Allard, D., and Froidevaux, R., (ed.). Kluwer Academic Publishers, Dordrecht, 67-78

Goovaerts P. 2000. Geostatistical Approaches for Incorporating Elevation into the Spatial Interpolation of Rainfall. J. Hydrol., 228 (1-2), 113-129

Gräler B, Pebesma E and Heuvelink G. 2016. Spatio-Temporal Interpolation using gstat. The R Journal 8(1), 204-218

GRASS Development Team. 2012. Geographic Resources Analysis Support System (GRASS) Software. Open Source Geospatial Foundation Project. http://grass.osgeo.org

Hanson CL. 1982. Distribution and Stochastic Generation of Annual and Monthly Precipitation on a Mountainous Watershed in Southwest Idaho. JAWRA Journal of the American Water Resources Association, 18, 875-883

Hay LE, Viger R and McCabe G. 1998. Precipitation Interpolation in Mountainous Regions Using Multiple Linear Regression: Hydrology, Water resources, and Ecology in Headwaters, Proceedings of the HeadWater'98 Conference, Kovar K, Tappeiner U, Peters NE and Craig RG. IAHS Publication (248), 33-38

Isaaks EH and Srivastava RM. 1989. An Introduction to Applied Statistics. Oxford University Press

Linsley RK, Kohler MA and Paulhus JLH. 1949. Applied Hydrology. McGraw Hill

Lloyd CD. 2005. Assessing the effect of integrating elevation data into the estimation of monthly precipitation in Great Britain. J. Hydrol., 308, 128-150

MAGRAMA. 2004. Water in Spain. Ministry of Agriculture, Food and Environment. Technical Secretariat-General. Madrid. Spain

Marquínez J, Lastra J, García P. 2003. Estimation Models for Precipitation in Mountainous Regions: the Use of GIS and Multivariate Analysis. J. Hydrol. 270, 1-11

Martínez-Cob A. 1996. Multivariate Geostatistical analysis of Evapotranspiration and Precipitation in Mountainous Terrain. J. Hydrol., 174 (1-2), 19-35

Mitas L and Mitasova H. 1988. General variational approach to the approximation problem, Computers and Mathematics with Applications, 16, 983-992

Naoum S and Tsanis, IK. 2004a. Ranking Spatial Interpolation Techniques using a GIS-based DSS. Global Nest: the Int. J. 6 (1), 2004, 1-20

Naoum S and Tsanis IK. 2004b. Orographic precipitation modeling with multiple linear regression. J. Hydrol. Eng., 9 (2), 79-102

Neteler M, Mitasova H. 2004. Open Source GIS: A GRASS GIS Approach. 2nd edition. Kluwer Academic Publishers/Springer, Boston. 424 p.

Ninyerola M, Pons X and Roure JM. 2005. Atlas Climático Digital de la Península Ibérica. Metodología y aplicaciones en bioclimatología y geobotánica. Universidad Autónoma de Barcelona, Bellaterra

Ninyerola M, Pons X and Roure JM. 2007. Monthly precipitation mapping of the Iberian Peninsula using spatial interpolation tools implemented in a Geographic Information System. Theor. Appl. Climatol., 89, 195-209

Pebesma EJ. 2004. Multivariable geostatistics in S: the gstat package. Computers & Geosciences, 30, 683-691

R Development Core Team. 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org

Rotunno R and Ferretti. R. 2001. Mechanisms of intense Alpine rainfall. J. Atmos. Sci., 58, 1732-1749

Salas JD, Delleur JW, Yevjevich V, Lane WL. 1980. Applied Modeling of Hydrologic Time Series. Water Resources Publications. Fort Collins Colorado, U.S.A., 484 p.

Singh P, Ramasastri KS and Naresh K. 1995. Topographical Influence on Precipitation Distribution in Different Ranges of Western Himalayas. Nordic Hydrology, 26 (4/5), 259-284

Tabios III GQ and Salas JD. 1985. A Comparative Analysis of Techniques for Spatial Interpolation of Precipitation. JAWRA Journal of the American Water Resources Association, 21, 365-380

Thiessen AH. 1911. Precipitation averages for large areas. Mon. Wea. Rev., 39, 1082-1089

Tobin C, Nicotina L, Parlange MB, Berne A and Rinaldo A. 2011. Improved interpolation of meteorological forcings for hydrologic applications in a Swiss Alpine region. J. Hydrol., 401 (1-2), 77-89

Weber DD and Englund EJ. 1992. Evaluation and comparison of spatial interpolators. Mathematical Geology 24, 381-391

Weber DD and Englund EJ. 1994. Evaluation and comparison of spatial interpolators II. Mathematical Geology 26, 589-603

World Climate Programme. 1985. World Meteorological Organization. Review of Requirements for Area-Averaged Precipitation Data, Surface-Based and Space-Based Estimation Techniques, Space and Time Sampling, Accurancy and Error; Data Exchange. Boulder Colorado, EE.UU., 17-19

WMO. 1994. Guide to hydrological practices. WMO-168. Data acquisition and processing, analysis, forecasting and other applications. 5th edition, 1994. World Meteorological Organization. Geneva. ISBN: 92-63-30168-9