Title: Monthly precipitation mapping of the Iberian Peninsula using spatial interpolation tools implemented in a Geographic Information System

Authors:

Miquel Ninyerola (Department of Animal Biology, Vegetal Biology and Ecology. Autonomous University of Barcelona)

Xavier Pons (Department of Geography and Center for Ecological Research and Forestry Applications. Autonomous University of Barcelona)

Joan M. Roure (Department of Animal Biology, Vegetal Biology and Ecology. Autonomous University of Barcelona)

Number of figures: 1 flow chart, 3 maps and 6 tables

Corresponding author email: miquel.ninyerola@uab.es

Summary

In this study, spatial interpolation techniques have been applied to develop an objective climatic cartography of precipitation in the Iberian Peninsula (583,551 km^2). The resulting maps have a 200m spatial resolution and a monthly temporal resolution. Multiple regression, combined with a residual correction method, has been used to interpolate the observed data collected from the meteorological stations. This method is attractive as it takes into account geographic information (independent variables) to interpolate the climatic data (dependent variable). Several models have been developed using different independent variables, applying several interpolation techniques and grouping the observed data into different subsets (drainage basin models) or into a single set (global model). Each map is provided with its associated accuracy, which is obtained through a simple regression between independent observed data and predicted values. This validation has shown that the most accurate results are obtained when using the global model with multiple regression mixed with the splines interpolation of the residuals. In this optimum case, the average R^2 (mean of all the months) is 0.85. The entire process has been implemented in a GIS (Geographic Information System) which has greatly facilitated the filtering, querying, mapping and distributing of the final cartography.

1. Introduction

Precipitation is a climatic element that is essential for life but can also be a destructive agent. For this reason, precipitation has been extensively studied using different approaches. In this study, the focus has been on average precipitation mapping using GIS (Geographic

Information System) techniques. GIS allows spatial analysis of the territory as well as improving management. Climatic maps are therefore an essential source of information to include in these systems. In general, precipitation maps are needed by researchers working in disciplines related with Earth sciences (climatology, hydrography, forestry sciences, agronomic sciences, vegetation sciences, etc) as well as by land planning and management experts.

The study of precipitation on a global scale has attracted the interest of researchers and there have been a number of efforts to map worldwide precipitation. There are many examples, from the more classic approaches such as the Times Survey Atlas of the World (Bartholomew 1922) to more recent initiatives that apply GIS techniques such as the Digital Atlas of the World Water Balance (Maidment et al. 1998). Although these studies provided general mapping, in some applications more local maps were needed. Consequently, Lennon and Turner (1995) developed an objective cartography for mean daily temperature in Great Britain. Although this paper is concerned with precipitation, this is a very interesting point of reference due to its applied interpolation methodology. Regarding precipitation research, Thompson et al. (1998) developed an atlas that includes maps of mean air temperature and total precipitation for North and Central America and Atkinson and Lloyd (1998) mapped precipitation in Switzerland.

In our case, the entire Iberian Peninsula has been mapped because of its geographic interest as a unit. This area has a complex climate (see section 2) that makes it scientifically interesting as a study area. Moreover, there have been few attempts to map the monthly precipitation of this region from a quantitative point of view. Examples of local mapping such as the Climatic Atlas of Extremadura (in the west of the Iberian Peninsula) by Felicísimo et al. (2001) and the Digital Climatic Atlas of Catalonia (in the northeast of the Iberian Peninsula) by Ninyerola et al. (2000) and Ninyerola et al. (2001b), are currently available. In addition, Guijarro (1986) mapped the mean temperature and precipitation of the Balearic Islands and Vicente-Serrano et al. (2003) developed annual maps of temperature and precipitation with the aim of comparing different interpolators in the Ebro basin. An alternative approach to developing a cartography was proposed by Sánchez et al. (1999), but only for Spain (not the entire Peninsula) and without independent data validation of the results. A similar case can be found in Portugal with the precipitation maps of the Atlas do Ambiente (Instituto do Ambiente 2002).

Regardless of whether or not they were implemented on a GIS, all these studies use quantitative techniques. Nevertheless, there have also been non-numerical approaches such as the Climatic Atlas of Catalonia (Clavero et al. 1996) and the Pluviometric Atlas of Catalonia (Febrer 1930). Another example, for the whole of Spain, is the Atlas of Spain (Font Tullot 1983), produced using traditional tools (based mainly on the knowledge of the researcher) and including maps where the climatologic fields are represented by isolines (i.e. in a non-explicitly continuous spatial variation).

On the other hand, the size of the Iberian Peninsula (583,551 km^2) permits the production of detailed maps (one point for every 200m) taking into account currently available data and GIS tools. However, it is not only a matter of having cartography in a digital format, but also of having all the data on an environment with spatial analysis capabilities, an easier and more consistent integration of the various steps involved in the process, etc. Among these processes, data filtering and selection, query, modelling, mapping and printing, as well as widespread publication on the Internet and updating should be emphasized. In short, GIS techniques allow one to develop mapping processes and to build the final cartography while facilitating the integration of this information with other kinds of spatial data.

This study presents a precipitation climatic cartography for the whole of the Iberian Peninsula and extends previous studies, which were limited to Spain or other smaller regions. Moreover, a more robust methodology is applied which in turn allows independent map validation. The method used is based on spatial interpolation and a GIS implementation. The spatial interpolation used is multiple regression with residual correction. This method is useful as it uses geographic information to interpolate the precipitation data from rain gauge stations. Subsequently, the multiple regression residuals (i.e. the proportion of non-explained variance) are spatially interpolated. Furthermore, different models and other spatial interpolation techniques have been tested (see section 3). Geostatistical techniques such as kriging and its derivatives were not tested due to the high computing time required (because of the sophistication of these methods and the number of tests needed to find the best variogram fit) to develop an extensive and accurate cartography, as proposed here. Moreover, in our case some previous tests have shown that these techniques failed to improve the results (Ninyerola et al., 2000). It should also be borne in mind that our methodology provides precipitation maps with an estimate of error based on an independent test. This point is relevant because knowledge of the map accuracy is always a useful piece of metadata to take into account when using the map.

This climatic precipitation cartography can be considered a subset of the Digital Climatic Atlas of the Iberian Peninsula (ACDPI) that we are developing. The ACDPI is defined as a "set of objective digital climatic maps of mean air temperature (minimum, mean and maximum), precipitation and solar radiation for the whole of the Iberian Peninsula, with a monthly and annual temporal resolution and a spatial resolution of 200 m^2 (Ninyerola et al., 2005). This atlas is linked to the Digital Climatic Atlas of Catalonia (Ninyerola et al. 2000) mentioned above, but in the present paper the methodology has been improved and new solutions introduced for the climatic modelling of this larger area (583,551 km^2 compared with 32,116 km^2 for Catalonia) while maintaining a spatial resolution of about 200 m. This study presents a set of 13 monthly and annual precipitation maps that will be published on the Internet, together with the rest of the atlas, as was the case for Catalonia (Ninyerola et al. 2001b).

2. Objectives

This study has three main objectives. The first is related with mapping purposes: to obtain monthly and annual climatic mapping of precipitation for the whole of the Iberian Peninsula. The other two objectives are derived from the techniques applied to create the cartography. The second objective is to investigate which geographic factors play an important role in precipitation modelling in this area; while the third is to investigate suitable numerical methods to obtain these maps, paying particular attention to the following possibilities: global versus basin models, buffered versus unbuffered models, role of distance from the sea and different interpolation methods.

3. Study area

The Iberian Peninsula is in the south-west of Europe, between longitude coordinates 9.5 W and 3.32 E (~1000 km), and between latitude coordinates 36 N and 43.8 N (~850 km). Official terrestrial cartography is in Universal Transverse Mercator (UTM) projection in Spain and Gauss-Krüger in Portugal. All the cartography has been converted into the UTM projection using rigorous geodetic methods. Therefore, the area comprises UTM zones 29, 30 and 31 but all the data have been reprojected into the zone 30 reference system because this is the central zone of the Iberian Peninsula.

3.1. Some climatic considerations about the Iberian Peninsula

Situated in the temperate zone, the Iberian Peninsula is at the intersection of several influences. The African anticyclone allows Saharan influences (subtropical air masses) while westerlies contributes with Atlantic influences (Atlantic sea air masses) as Capel (2000) stated, without forgetting Mediterranean influences (sea warming) in the eastern area. As García de Pedraza and Reija (1994) have pointed out, the peninsula is in the southern limit of the polar front and in an aerological contact area with the sub-Saharan influence. From the point of view of precipitation mapping this represents a serious complication, but at the same time it makes the spatial interpolation performance more attractive because of the need to find an acceptable solution to this complex problem.

The complexity of the climate is not only due to the location of the peninsula, but is also related to its orography. Mean altitude, obtained from the digital elevation model (DEM), is 640 m, but 15% of the area is at altitudes of over 1000m. Further information on the climate of the Iberian Peninsula can be found in Linés (1970), Font Tullot (2000) and Capel (2000).

4. Spatial interpolation methodologies

4.1. Multiple regression with residual correction

Although different spatial interpolators have been used, the main purpose is to investigate the role of multiple regression analysis with the residual correction method. According to Burrough and McDonnell (1998) this method of interpolation is scientifically interesting because, in addition to the interpolation process, it gives information about the relationship between the geographic reality of the land and climate.

This interpolation methodology is a combination of statistical (multiple regression) and spatial interpolation (splines and inverse distance weighting) tools. Unlike the classic approach of interpolating precipitation values themselves, this method interpolates the residuals obtained from a multiple regression analysis, taking advantage of the predictive nature of the regression. Only the unexplained variation derived from multiple regression has been included in the spatial interpolation process.

The aim is to perform several multiple regression analyses (one for each month) with the climatic variable (precipitation) as the dependent one and the geographic variables (altitude, distance from the sea, etc) as the independent ones. Once the multiple regression analysis (backward stepwise method) has been carried out and the multiple regression coefficients have been obtained, the map algebra tools are applied. This means reproducing the equation of the regression fit using the raster matrices of the independent variables. The outcome of this procedure is a raster matrix map for each month.

The maps obtained after multiple regression have the inherent error of the regression fit. In other words, they explain what the geographic variables used are able to tell us based on the model. The residuals of the regression fit at each meteorological station reflect unexplained variation, i.e. other factors, usually more or less local, but also of any kind not considered by the model. In addition, the residuals include all kinds of data uncertainties (data handling errors, etc). If these residuals are spatially interpolated to obtain a surface, the map obtained from the regression model can be corrected. This has two advantages: on the one hand, this correction allows us to have an exact interpolator at the location of the meteorological stations and, on the other, it will greatly improve the results for variables that are more difficult to model using only geographic variables, such as precipitation.

There are two steps prior to the application of the regression model. The first is to obtain and filter out the data from the meteorological stations. This climatic data will be the dependent variable. In section 4, there are some comments and information about these stations. The second step is to select the independent variables and to produce the corresponding raster maps. These maps will contain the main geographic factors modelling the climate that are currently available on our GIS. In this case, the independent variables used to model the precipitation are altitude, latitude, distance from the sea, terrain curvature and solar radiation. Different types of distance from the sea (linear, logarithmic and quadratic) have been tested in order to find the most suitable function. In section 5, the meaning of each variable in this model and the process followed to obtain each one of them are explained in depth. These maps perform two roles: to get data (such as distance from the sea or solar radiation) for each meteorological station which is not usually provided by meteorological agencies, and to be the basis for obtaining the final climatic cartography using map algebra. Figure 1 shows a flow chart of the whole methodological process. More details about this methodology can be found in Ninyerola et al. (2000). Predictive statistical models relating climate and geographical variability have also been reported by Basist et al. (1994), Daly et al. (1994), Egido et al. (1985), Goodale et al. (1998), Hernández et al. (1975), Hay et al. (1998) and Perry and Hollis (2005).

4.2. Splines and inverse distance weighting

Two interpolators (splines and inverse distance weighting) that do not take into account geographic information have been tested with the aim of comparing them with the multiple regression with residual correction method. These methods have also been used to interpolate the residuals of the regression method but, in this case, they have been employed to interpolate the precipitation values obtained from the meteorological stations. In the case of distance weighting (Burrough and McDonnell 1998), a square exponent has been used as established in previous tests. In the case of splines (Mitasova and Mitas 1993), a tension=400 and smooth=0 have been used as these parameters provided the best fit for the validation set.

References to other experiences with splines can be found in Hutchinson (1995) and Wahba and Wendelberger (1980), while Tomczak (1998) investigates inverse distance weighting behavior. Vicente-Serrano et al. (2003) compare different interpolators including those mentioned above but they found suitable results with geostatistical interpolators, although their study covers a relatively small area (Ebro basin).

4.3. Testing the models

In order to know the robustness of each method, a validation test with two, randomly-chosen, subsets has been used. During the research, we made several tests by running the model with different fit/validation sets derived from different random selections, without finding any significant differences. The first subset is made up of 60% of the meteorological stations used to build the model, and the second subset comprises the remaining 40% of the stations that have been used to independently test the model.

The final maps incorporate all the meteorological stations. This will allow us to obtain better (or, in the worst case, the same) results, although the minimum accuracy will be indicated by the test using 40% of the independent stations.

4.4. Summarizing the tested models

To make it easier to read, these are the different models that have been tested:

- RG_LI_IDW: multiple regression (using linear distances from the sea) and residuals interpolated with inverse distance weighting.
- RG_LG_IDW: multiple regression (using logarithmic distances from the sea) and residuals interpolated with inverse distance weighting.
- RG_QU_IDW: multiple regression (using quadratic distances from the sea) and residuals interpolated with inverse distance weighting.
- RG_LI_SP: multiple regression (using linear distances from the sea) and residuals interpolated with splines.
- RG_LG_SP: multiple regression (using logarithmic distances from the sea) and residuals interpolated with splines.
- RG_QU_SP: multiple regression (using quadratic distances from the sea) and residuals interpolated with splines.
- IDW: inverse distance weighting, using the precipitation values observed at the meteorological stations.

- SP: splines, using the precipitation values observed at the meteorological stations.

5. Climatologic data

The meteorological stations were bought from the National Institute of Meteorology of Spain (INM) and also obtained from the literature in the case of Portugal. The series used correspond to the period 1950-99. A filtered set of an average of 2928 rain gauge stations (depending on the month) with monthly data and 1999 stations with annual precipitation data have been used. We have not considered rain gauge inherent uncertainty because an analysis taking this factor into account is beyond the scope of this research.

These stations have been selected by compromising between series length (temporal stability) and density (spatial coverage) from the original 7293 rain gauge stations that were at our disposal. The optimal length of the series (20 years) has been identified through statistical tests. Other studies of statistical prediction (Egido et al. 1985) use 8 to 20 years for precipitation. If the 30-year periods recommended by the WMO (1989) are applied, only 1208 rain gauge stations would be usable nowadays. Moreover, if these stations are plotted, it is observed that there are large areas with no coverage and that the test results are worse than decreasing the series length to maintain a denser coverage. In this study, a 199 km^2 average density of stations has been used.

Similarly, if these 20 years are restricted to the same period, the results are worse than using series which are as long as possible within the whole 1950-99 period. This is not the conventional way to proceed, and of course would not be adequate for detailed station or area climatic characterization. However, our concern is the mapping of a large area and it is important to reach an optimum spatial-temporal compromise. In other words, series stability does not seem to be as critical as spatial representability, at least in the Iberian Peninsula area. This compromise is also present in the study by Sánchez et al. (1999).

It is important to note, as mentioned in section 2, that the Portuguese meteorological stations used in this report are five times less dense than the Spanish ones. Unfortunately, the most complete set of data available for us has been the report by Tormo et al. (1992).

5.1. Single large or several small?

One important question when modelling a relatively large area is "Should we treat the area as a whole?" or "Should we split the area up?". In the latter case, this means grouping different meteorological stations together to produce different sub-models. The idea is to identify which areas provide the best fits for the model.

Two different approaches have been tested:

- using the entire peninsula as a unit (*global model*)

- using drainage basins as a unit of spatial climatic homogeneity (*basin model*)

The number of stations (without filter and filtered) and the density for each one of the drainage basins and for the entire peninsula are shown in Table 1.

5.2. Global model

Using the whole of the Iberian Peninsula as a unit is problematical for two reasons. The first is the enormous climatic variability within the area. Indeed, common sense suggests that the solution would be to divide up this large area in the interests of greater spatial climatic homogeneity. Consequently, the *global model* has been developed to compare it with the *basin model* that should provide a priori better results. Surprisingly, as we will see later (section 6), our results show that the *global model* works better than most of the *basin models*.

The second problem is related to the size of the digital files, which causes storage and computing-time problems. However, considering the rapid evolution of storage systems (hard-drives, CD-ROM, etc.), this problem should not be over-emphasized. On the other hand, computing time seems far from being solved, at least when there are a lot of stations involved in the process. Batch files and macros have been used to automate and develop the main computations that have allowed us to reduce computing-time problems.

5.3. Drainage basin models

The first idea was to use the main drainage basins as a starting point. In other words, to begin to work with relatively heterogeneous areas and then, depending on the results, split the basins up to obtain more homogenous areas (sub-basins) or join them together (searching for a more general situation). Regarding splitting, for example, one possibility was to divide up the Ebro basin based on the influence of the Iberian mountains. An approach to obtain non-stationary models can be found in the geographically-weighted regression methodology applied by Brunsdon et al. (2001).

The climatic model has been developed for each one of the drainage basins used to divide up the entire Iberian Peninsula (Figure 2). A detailed description of these basins can be found in SMN (1968).

The list of the drainage basins, classified according to the sea into which they drain, is as follows:

Atlantic basins

- **Septentrional basin**: northern slopes draining into the Atlantic Ocean (in the west) as well as into the Bay of Biscay (in the north).
- **Duero basin**: slopes flowing into the River Duero.
- **Tajo basin**: slopes flowing into the River Tagus.
- **Guadiana basin**: slopes flowing into the River Guadiana.
- **Guadalquivir basin**: slopes flowing into the River Guadalquivir.

Mediterranean basins

- **Eastern Pyrenees basin**: slopes flowing into the northeast Mediterranean Sea.
- **Ebro basin**: slopes flowing into the River Ebro.

- **Llevant basin**: slopes flowing into the eastern-central Mediterranean Sea.
- **Segura basin**: slopes flowing into the River Segura and slopes flowing into the southeast Mediterranean Sea.
- **Meridional basin**: slopes flowing into the south Mediterranean Sea.

The code for each meteorological station that is provided by the INM includes information about the main drainage basin where they are situated. It has been easy, therefore, using GIS database tools, to separate the subsets of the meteorological stations needed to build the sub-models. Figure 2 shows the definitive plot of the stations used.

5.4. Buffer models

There is another problem to solve when selecting the stations for each basin. Is it suitable only to use stations inside the basin or is it better also to use surrounding stations outside the basin? Both situations have been checked:

- Using only the meteorological stations inside the basin.
- Using the stations inside the basin, but adding the ones that are outside, but near, the basin. In other words, a surrounding buffer area has been introduced. This buffer has been chosen empirically with the aim of obtaining a number of peripheral meteorological stations within 20 km of the basin limits. However, it could be interesting in a future study to test different buffer distances to identify which distances provide the best climatic interpolation fit.

One distance map (produced using GIS tools) has been computed for each one of the basins of all the meteorological stations. This information has allowed us to select the stations that were inside the buffer and then perform the analysis and test it. In section 6, we will see in which cases this buffer has improved the model's fit.

It is important to bear in mind that all the methods described in section 3 have been developed and validated by using a buffer (only in the case of the basins) and without using a buffer (in the case of the basins and in the global case).

6. Geographic data

The independent variables have been chosen according to the most cited geoclimatic factors in the literature (Lacoste and Salanon 1973) and the possibilities of obtaining the data.

Altitude and latitude are consensus factors regarding their effect on climate (Schermerhorn 1967). Moreover, Font Tullot (2000) refers to the latitudinal gradient of precipitation in the Iberian Peninsula. Solar radiation is another factor with a clear relation to climate, but one that is normally not extensively used because it is often not available. Solar radiation has topographic information (slope and aspect) that can influence cloud formation or wind circulation (Font Tullot 2000).

The case of continentality is different. While also being an important factor for climate modelling, it is usually substituted, in interpolation studies, by the X position (Thompson et al. 1998; Sánchez et al. 1999). While the Y position makes physical sense since it matches the latitude, the X position, based on the reference meridian, has no physical meaning and it is better, if possible, to use a model which makes more geographic sense. Sea distance or distance from an important geographic event such as main general circulation lines are interesting possibilities.

Logically, there are other geographic variables that could be used as climatic factors. Egido et al. (1985) and Hernández et al. (1975) use two interesting variables. On the one hand, the second derivative (laplacian operator) of the altitude is included in the precipitation models. This shows if a point on the ground is in a convex or concave location (local orography). On the other hand, the first authors also use distances to the main general circulation fronts. In fact, this is a sophistication of the latitude as these fronts travel in an easterly direction. Lorente (1946), who also includes sea currents, corroborates this idea. Bellot (1978) cites the morphology of the coast and sea currents as important factors. One example can be found in the Basque Country (Euskadi), where the Gulf Stream allows a higher temperature than expected for its latitude (García de Pedraza and Reija 1994).

At present we have not added other variables or sophistications, but they could be explored in the future. Our aim is to keep the model as simple as possible and only to introduce new variables when their contribution is clearly important.

Among these other variables are topographic aspect and slope. However, these are implicit in the solar radiation model, and in preliminary tests we have observed that they do not introduce additional variations of the model as previously published (Hutchinson 1995).

Finally, the geographic variables selected are: altitude, latitude, continentality, terrain curvature and solar radiation. It is important to note that latitude and solar radiation have been introduced independently. This is because the solar radiation model is computed using the central point of the geographic area (Iberian Peninsula) and therefore does not include the latitude factor.

6.1. Altitude

Altitude has been obtained by the interpolation of contour lines digitised from the 1:200000 topographic maps (SGE 1967-71) using the IsoMDE module of the MiraMon software (Pons 2002). These contour lines are spaced 100 m in height and the interpolation gives an altitude point every 200 m x 200 m.

There are different variations on the usage of this climatic factor. For example, Thompson et al. (1998) use a smoothed altitude and Lennon and Turner (1995) use different parameters based on the orography as the altitude with different resolutions, which allows them to obtain descriptive statistics to be used as variables (maximum altitude, standard deviation, etc). In this study, the central altitude of the DEM cells has been used.

In the studies cited in section 1, we find resolutions of 5 km for Great Britain (*ca*. 225,000 cells), 25 km for North America (*ca*. 15,800 cells) and 1 km for the Spanish Peninsula (*ca*. 505,000 cells). White and Smith (1982) also worked in Great Britain with a 10 km resolution (*ca*. 56,250 cells). On the other hand, if we consider the year of publication of these reports, it is easy to understand the limitation of the amount of cells to be processed regarding the capabilities of computers in the final decades of the 20th century.

6.2. Latitude

This raster matrix is easily obtained using typical GIS tools.

6.3. Continentality

Continentality is a complex factor that depends on the strength of the sea wind influence. For this reason sea distance and orographic barriers must be taken into account. However, in previous studies (Ninyerola, 2001a), we attempted to model continentality? using a sigmoidal distance function that takes into account the effect of the mountains that run parallel to the Catalan coast. The best test results were obtained when this sigmoidal function was similar to a straight line. It was clearly a problem of the number and spatial situation of the meteorological stations. This may be solved when more stations get enough length on their series.

In this case, continentality has been modelled as the coastal distance according to three mathematical functions: linear, logarithmic and quadratic. Their predictive behavior is described in the results section. The linear function has the advantage of simplicity for testing sophisticated models, as explained above. In the same way, the logarithmic function attempts to take into account the barrier effect of coastal mountain ranges (two close points near the coast will have more different values than two close points far from the coast). In Hargy (1997) there is an example of logarithmic sea distance use applied to temperature interpolation. The quadratic distance has the advantage of exaggerating the value for points far from the coast.

The Iberian Peninsula coast has been split into three sections (see Figure 2): the Mediterranean coast (from the north of Catalonia to Gibraltar), the Atlantic coast (from Gibraltar to the north of Galicia – Cape Estaca de Bares) and the Cantabrian coast (from Cape Estaca de Bares to the French border). This division of the coast in order to compute the distance makes physical sense due to the different climatic behavior of the different coastal areas (Bellot 1978).

6.4. Terrain curvature

This variable is derived from the DEM, since curvature is the second derivative of the altitude. Therefore, it provides an index of greater or lesser convexity and concavity classifying the area into deep-valley cells, plain cells and top-hill cells with different strengths. For more information on this, see the MiraMon GIS manual (Pons 2002).

6.5. Solar radiation

It is important to notice that solar radiation is obtained using a different methodology from the other climatic variables. In fact, it is a physical computational model based on relief and the position of the Sun. The model integrates the Sun's position throughout the day, the local incident angles accounting for slope and aspect, topographic shadowing effects (cast-shadows), astronomic parameters, etc. For more information, see Pons (1996) and the MiraMon GIS manual (Pons 2002).

Unlike the case of Catalonia (Ninyerola et al. 2000), this model has not been corrected with ground data (meteorological stations recording solar radiation) because in the case of the whole of the Iberian Peninsula the network of stations recording solar radiation is still poor.

7. Results and discussion

7.1. Single large: the global model wins

The *global model* (using all the average 2928 filtered stations from the Iberian Peninsula) has provided better mean results than the *basin model* (taking the stations of each drainage basin separately). It should be borne in mind that “mean results” in this case refer to the mean of the determination coefficients computed from the 40% of independent test stations for all the months in the case of the *global model* and to the mean of all the months from all the basins in the case of the *basin model*. While there are some situations (some months in some basins) with better values than the *global model,* the mean results show clearly that the *single large* option is better.

Table 2 shows that the best mean result of the *global model* is R^2=0.84 (RG_QU_SP) while the best mean result for the *buffered basin model* is only R^2=0.70 (RG_LI_SP). Moreover, if we look at the same table, the *global model* is also better than the *basin model* for the other interpolation methods (t-test differences are significant at p<001). Table 3 shows monthly results of RG_QU_SP model in terms of R^2 as well as RMS error. Figure 3 and Figure 4 show, as an example, the residual map and the corrected map for annual precipitation, respectively.

Therefore, although the original idea was to build the Iberian Peninsula precipitation surfaces through the combination (mosaicking) of all the main drainage basin maps, the best solution is to apply the global model cartography. Although some month-basin-method combinations have provided better results than the global model, the mean results show an interesting scientific rule: the general model is better than local models.

However, from the point of view of mapping, two strategies can be distinguished:

- General strategy: global models are often more scientifically interesting and less prone to artefacts. Moreover, in this case the results allow the use of only one map for each month (*global model*). This situation reduces the complexity of map manipulation (e.g. mosaicking of basins). Therefore, with the exception of a few cases, the majority of times the *global model* will be used, even if the objective is to work in a smaller area of the Iberian Peninsula.

- Puzzle strategy: it is possible to refine and improve the results by using the highest result (for month and interpolation methods) regardless of whether it comes from the global fit or a specific *basin model*. This strategy would provide a map built from different local maps. This strategy is not recommended when the aim is to work with the entire peninsula, especially because the results of the *global model* are already suitable, as we will see below. The only reason for using *basin models* is when it is necessary to work locally and the results of the basin in question are clearly better than the *global model*. For example, the Ebro basin during January provides better R^2 results than the Iberian model in that area (0.90 versus 0.81, $p<0.001$).

7.2. Comparing the interpolation methods

When analysing all the interpolation methods used, some general patterns emerge:

- In the case of the global model, regression methods show better results than interpolators that do not use geographic information (IDW and SP): RG_QU_SP is more reliable than IDW (0.84 versus 0.81, $p<0.05$). In the case of the basin models, the regression (RG_LI_SP) mean results are slightly greater than non-geographical interpolators (SP), but the differences are not significant (0.67 versus 0.65, $p<0.33$).

- Logarithmic distance provides worse results when compared to previous publications (Ninyerola 2001a). When modelling precipitation it is more suitable to use quadratic or linear distances from the sea. However, the differences between using one continentality model or another are not significant (0.83 versus 0.84, $p<0.46$).

- SP fits the precipitation surface better than the IDW when interpolating the observed values (0.68 versus 0.63, $p<0.05$) as well as when interpolating the regression residuals (RG_QU_SP and =0.69 and RG_QU_IDW =0.65, p<0.05). There is, however, one main exception in the case of the global model. It is possible that splines application in a such large area could be improved but, for the moment, this would involve a time-consuming process that discourages different tests and validations.

7.3. Residual correction improvement

There is a big difference between the validations made before and after applying the residuals correction. For example, in the case of the global model (RG_QU_SP) the results are, respectively, 0.51 versus 0.84 (Table 2). While regression analysis explains part of the variability, residual interpolation increases map accuracy. This behavior is different from the data obtained for the temperature models (Ninyerola 2000) where the regression itself explains the greater part of the variability, which in fact underlines the idea that precipitation is more difficult to model than temperature.

7.4. Significance of the independent variables

On the basis of the acceptance of the RG_QU_SP of the *global model* as a suitable general solution, some general comments can be made about the independent variables.

There are four significant variables throughout the year: altitude, curvature and quadratic distances to the Atlantic and Cantabrian coasts. The other variables computed have different patterns for which the distances are difficult to interpret. Latitude is not relevant during winter and summer and solar radiation is only significant during spring and summer.

Regarding the sign of the regression coefficients, predicted values have usually been found. Therefore, and to summarize, precipitation is higher as:

- Altitude and latitude increase.
- Solar radiation decreases. As atmospheric variations have not been taken into account, this is not a cloudiness effect. However, the areas that are more cast-

shadowed tend to have higher precipitation. This variable is only significant during spring and summer.

- Terrain curvature is concave. This means that valley floors receive more precipitation than top sites.

- Quadratic distance to the Mediterranean coast decreases during summer and increases in the other months.

- Quadratic distance to the Atlantic coast decreases during winter and some spring months and increases in the other months. This variable has an unpredictable pattern.

In Table 4, beta coefficients for different annual, global and basin regression models are presented.

7.5. Basin models

In Table 6, it can be seen that the most successfully predicted basins are the Ebro (R^2=0.86) and the Duero (R^2=0.80) while the least successfully predicted are the Septentrional (R^2=0.59) and the Eastern Pyrenees (R^2=0.65). Note that the Ebro and Duero basins are the only models that provide results with non-significant differences for the global model (R^2=0.84).

Buffering usually provides slightly improved results, but differences with non-buffered models are not significant. In Table 5, both situations are compared in detail.

8. Conclusions and future trends

A set of monthly precipitation maps has been obtained as well as an annual map for the Iberian Peninsula that are probably the most accurate and meaningful currently available taking into account refined spatial resolution (200m), the number of meteorological stations used (2928), the combination of numerical methods (spatial interpolation methods, statistical analysis, and GIS tools) and the independent validation, which provides a known quality index for each map. The whole process has been fully implemented in a GIS and will be published on the Internet. Additionally, the results are interesting as they provide information about the relationship between geographic variables and precipitation. Furthermore, the results show that, generally speaking, the global model performed better than the local basin models.

In future studies, statistical interpolation (multiple regression or polynomial fit) will allow us to investigate new independent variables that could help to explain climatic spatial variability. Thus, variables related with altitude and distance from the sea (orographic barriers and morphology of the coast, for example) can be modelled with GIS tools. In addition, there are other disciplines (remote sensing) and approaches (diagnostic models) that could be used to produce interesting variables. In the first case, there is the example of Creech and McNab (2002) who used the Normalized Difference Vegetation Index (NDVI) to produce a spatial trend for precipitation mapping purposes. In the second case, there are reports like that of Thompson et al. (1997) that investigate the behavior of precipitation in relation to orography using physical models.

References

Atkinson PM, Lloyd CD (1998) Mapping precipitation in Switzerland with ordinary and indicator kriging. J Geogr Inf Decis Anal 2:72-86

Bartholomew JG (1922) The Times Survey Atlas of the World. The Times, 122 pp

Basist A, Bell GD, Meentemeyer V (1994) Statistical relationships between topography and precipitation patterns. J Climate 7:1305-1315

Bellot F (1978) El tapiz vegetal de la península ibérica. Madrid: Blume, 423 pp

Brunsdon C, McClatchey J, Unwin DJ (2001) Spatial variations in the average rainfall-altitude relationship in Great Britain: an approach using geographically weighted regression. Int J Climatol 21:455-466

Burrough PA, McDonnell RA (1998) Principles of Geographical Information Systems. New York: Oxford University Press, 333 pp

Capel JJ (2000) El clima de la península ibérica. Barcelona: Ariel, 281 pp

Clavero P, Martín-Vide J, Raso JM (1996) Atles climàtic de Catalunya. Barcelona: Institut Cartogràfic de Catalunya and Departament de Medi Ambient, 42 pp

Creech TG, McNab AL (2002) Using NDVI and elevation to improve precipitation mapping. In 13th Conference on Applied Meteorology, 13-16 May 2002. American Meteorological Society.

Daly C, Neilson RP, Phillips DL (1994) A statistical-topographic model for mapping climatological precipitation over mountainous terrain. J Appl Climatol 33:140-158

Egido A, De Pablo F, Egido M, Garmendía J (1985) La precipitación en la cuenca del Duero como función de los factores geográficos y topográficos. Revista de Geofísica 41:183-190

Febrer J (1930) Atlas pluviomètric de Catalunya. Barcelona: Institució Patxot, 518 pp

Felicísimo AM, Morán R, Sánchez JM, Pérez D (2001) Elaboración del Atlas climático de Extremadura mediante un sistema de información geográfica. GeoFocus 1:17-23

Font Tullot I (1983) Atlas climático de España. Madrid: Instituto Nacional de Meteorología, 43 pp

Font Tullot I (2000) Climatología de España y Portugal. Salamanca: Universidad de Salamanca, 422 pp

García de Pedraza L, Reija A (1994) Tiempo y clima en España (Meteorología de las autonomías). Madrid: Dossat, 409 pp

Goodale CL, Aber JD, Ollinger SV (1998) Mapping monthly precipitation, temperature, and solar radiation for Ireland with polynomial regression and a digital elevation model. Climate Res 10:35-49

Guijarro JA (1986) Contribución a la Bioclimatología de Baleares. Ph.D. dissertation summary, 27 pp. Universitat de les Illes Balears [Available from Universitat de les Illes Balears Valldemossa Road 7.5. km 07122 Palma de Mallorca Spain]

Hargy VT (1997) Objectively mapping accumulated temperature for Ireland. Int J Climatol 17:909-927

Hay L, Viger R, McCabe G (1998) Precipitation interpolation in mountainous regions using multiple linear regression in Hydrology Water Resources and Ecology in Headwaters. Kovar K, Tappeiner U, Peters NE, Craig RG eds. International Association of Hydrological Sciences: 33-39

Hernández JA, Garmendía J, Hernández E, Sánchez JF (1975) Importancia de la laplaciana de la altitud en las precipitaciones. Revista de Geofísica 35:140-147

Hutchinson MF (1995) Interpolating mean rainfall using thin plate smoothing splines. Int J Geogr Inf Syst 9:385-403

Instituto do Ambiente (cited 2002) Atlas do Ambiente. Lisboa: Direcção General do Ambiente de Portugal [Available online at http://atlas.isegi.unl.Pt/website/atlas/din /viewer.htm.]

Lacoste A, Salanon R (1999) Éléments de biogéographie et d'écologie Paris: Nathan, 318 pp

Lennon JJ, Turner JR (1995) Predicting the spatial distribution of climate temperature in Great Britain. J Anim Ecol 64:370-392

Linés A (1970) The climate of the Iberian Peninsulav in World survey of climatology climates of the Northern and Western Europe. Wallen CC ed. Amsterdam: Elsevier, 253 pp

Lorente JM (1946) Climas Españoles. Revista de Geofísica, 18:204-231

Maidment DR, Reed SM, Akmansoy S, Asante K, McKinney DC, Olivera F, Ye Z (cited 1998) Digital Atlas of the World Water Balance. Center for Research in Water Resources (University of Texas) [Available online at http://www.crwr.utexas.edu/gis/gishyd98/atlas/Atlas.htm.]

Mitasova H, Mitas L (1993) Interpolation by Regularized Spline with Tension: I. Theory and Implementation. Math Geol 25:641-655

Ninyerola M, Pons X, Roure JM (2000) A methodological approach of climatological modelling of temperature and precipitation through GIS techniques. Int J Climatol 20:1823-1841

Ninyerola M (2001a) Modelització climática mitjançant eines SIG i la seva aplicació a l'anàlisi quantitativa de la distribució d'especies vegetals a l'Espanya peninsular. Ph.D. dissertation. Autonomous University of Barcelona, 202 pp [Available from Universitat Autónoma de Barcelona Campus Bellaterra 08193 Barcelona Spain. Available online at http://www.tdx.cbuc.es/TDX-0618101-111736/index.html.]

Ninyerola M, Pons X, Roure JM (cited 2001b) Digital Climatic Atlas of Catalonia on the Internet. [Available online at http://www.uab.es/atles-climatic.]

Ninyerola M, Pons X, Roure JM (2005) Atlas Climático Digital de la Península Ibérica. Metodología y aplicaciones en bioclimatología y geobotánica. ISBN 932860-8-7. Bellaterra: Autonomous University of Bellaterra, 44 pp

Sánchez O, Sánchez F, Carretero MP (1999) Modelos y cartografía de estimaciones climáticas termopluviométricas para la España peninsular. Madrid: Ministerio de Agricultura Pesca y Alimentación. Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria, 192 pp

Perry M, Hollis D (2005) The generation of monthly gridded datasets for a range of climatic variables over the UK. Int J Climatol 25:1041-1054

Pons X (1996) Estimación de la Radiación Solar a partir de modelos digitales de elevaciones. Propuesta metodológica. Proc VII Coloquio de Geografía Cuantitativa Sistemas de Información Geográfica y Teledetección, Vitoria-Gasteiz, Spain, 87-97 [Available from http://magno.uab.es/atles-climatic/rad_pot.pdf]

Pons X (2002) MiraMon. Geographical information system and remote sensing software. Version 4.4. Centre de Recerca Ecològica i Aplicacions Forestals CREAF. ISBN: 84-931323-4-9

Schermerhorn VP (1967) Relations between topography and annual precipitation in western Oregon and Washington. Water Resour Res 3:707-711

SGE (1971) Mapa militar de España. Serie 1:200000. Servicio Geográfico del Ejército, 105 pp

SMN (1968) Situación geográfica e indicativos de las estaciones pluviométricas españolas. Servicio Meteorológico Nacional, 11 pp

Solé-Sabarís L, Font i Quer P, Llopis N, Masachs V (1952) Geografía Física in Geografía de España y Portugal. Terán M ed. Barcelona: Montaner i Simón

Thompson CS, Sinclair MR, Gray WR (1997) Estimating long-term annual precipitation in a mountainous region from a diagnostic model. Int J Climatol 17:997-1007

Thompson RS, Anderson H, Barlein PJ (1998) Atlas of relations between climatic parameters and distributions of important trees and shrubs in North America. Introduction and Conifers. Denver: U.S. Geological Survey, 269 pp

Tomczak M (1998) Spatial interpolation and its uncertainty using automated anisotropic Inverse Distance Weighing (IDW) - Cross-validation/Jackknife approach. J Geogr Inf Decis Anal 2:18-30

Tormo R, Ruiz T, Devesa JA (1992) Aportación a la bioclimatología de Portugal. Anales del Jardín Botánico de Madrid 49:245-264

Vicente-Serrano SM, Saz-Sánchez MA, Cuadrat JM (2003) Comparative analysis of interpolation methods in the middle Ebro Valley (Spain) application to annual precipitation and temperature. Climate Res 24:161-180

White ES, Smith, RI (1982) Climatological maps of Great Britain. Institute of Terrestrial Ecology, 40 pp

Wahba G, Wendelberger J (1980) Some new mathematical methods for variational objective analysis using splines and cross validation. Mon WeaRev, 108:1122-1143

World Meteorological Organization (1989) Calculation of monthly and annual 30-year standard normals. Geneva: WMO-TD/No. 341, 11 pp

FIGURE CAPTIONS

Figure 1. Methodology flow chart. White numbers on black background indicate the corresponding text sections.

Figure 2. This figure provides graphic information about three aspects: a) map of the main drainage basins used as a basis to assign the meteorological stations to the different basin models; b) plot of the rain gauge network finally used in the interpolation process; c) coastal division used to model continentality: Mediterranean coast (light grey line), Atlantic coast (black line) and Cantabrian coast (dark grey line).

Figure 3. Residual map of annual precipitation. The dots represent the meteorological stations used in the interpolation process. The regression residuals have been computed at each meteorological station and interpolated to the entire territory through splines. Dark grey represents areas where the regression fit predicts higher values than the observed ones. Light grey represents the opposite.

Figure 4. Annual precipitation map of the Iberian Peninsula. This surface has been obtained through spatial interpolation (multiple regression with residual correction) from 1999 meteorological stations data in the case of the global model (RG_QU_SP). The spatial resolution is 200 m and the associated error is R^2=0.84. A zoomed image has been hill shaded for greater clarity.

Table 1. Number and density of the rain gauge stations before and after filtering*

	Before filtering		After filtering	
Drainage basins	N. stations	Density (km^2)	N. stations	Density (km^2)
Eastern Pyrenees	453	36.3	152	107.9
Septentrional	1141	47.2	296	181.7
Duero	1026	76.8	562	140.3
Tajo	554	100.8	191	292.5
Guadiana	774	77.5	300	200.1
Guadalquivir	1097	57.6	325	194.6
Meridional	384	46.8	136	131.7
Segura	315	60.0	157	120.5
Llevant	755	56.8	269	159.2
Ebro	859	99.6	437	195.7
Iberian Peninsula	7405	78.8	2566	227.4

* meteorological stations with 20 or more monthly (or annual) values have been used in the interpolation process.

Table 2. Independent test mean results comparison between the Global and the Basin models for the different interpolation methods. In the former case the mean R^2 of all months is shown. In the later case the mean R^2 of all months and all basins is shown. Greater R^2 are expressed in bold. R^2 obtained from validations before the residual correction is specified between brackets. Differences between Global and Basins model (t-test) are all significant at $p<0.05$.

Interpolation method	Global model	Buffered basins model
RG_LI_IDW	0.81	0.66
RG_LI_SP	0.84	**0.70**
RG_QU_IDW	0.81	0.65
RG_QU_SP	(0.51) **0.84**	0.69
RG_LG_IDW	0.80	0.65
RG_LG_SP	0.83	0.69
IDW	0.81	0.63
SP	0.80	0.68

RG=multiple regression, IDW=inverse distance weighting, SP=splines, LI=lineal distance to the coast, QU=quadratic distance to the coast, LG= logarithmic distance to the coast.

Table 3. Independent mean test results: R^2 and RMS (Root Mean Squared error) from the RG_QU_SP model.

Month	R^2	RMS (in mm)
January	0.83	18.7
February	0.83	16.0
March	0.72	17.5
April	0.83	12.5
May	0.84	11.1
June	0.88	7.6
July	0.92	5.9
August	0.90	7.9
September	0.86	9.6
October	0.80	14.8
November	0.83	17.0
December	0.84	20.5
Mean	0.84	13.3
Annual	0.84	137.8

Table 4. Beta coefficients from different annual regression models (basin and global). The standard error of beta coefficients are shown between brackets. In all cases the significance values are lower than p<0.001. Empty cells indicate non-significant beta coefficients.

	ALT	LAT	QUDIME	QUDIAT	QUDICA	SOLRAD	CURV
Global (PI)	0.14 (0.02)	-1.34 (0.08)	0.34 (0.03)		1.18 (0.08)		
Eastern Pyrenees	0.86 (0.10)				0.77 (0.09)		0.32 (0.09)
Septentrional		-1.00 (0.20)		0.40 (0.09)	1.13 (0.22)		
Duero	0.24 (0.04)	-2.38 (0.17)	1.25 (0.11)	1.12 (0.12)	2.90 (0.19)		
Tajo	0.38 (0.09)	0.28 (0.08)	0.41 (0.07)				0.20 (0.06)
Guadiana	0.95 (0.11)	-3.28 (0.44)		-0.93 (0.13)	3.28 (0.44)		
Guadalquivir	0.46 (0.08)	-2.73 (0.58)			3.08 (0.59)		
Meridional	0.50 (0.09)	-3.59 (0.73)			4.47 (0.74)		
Segura	0.51 (0.10)		1.18 (0.15)	0.64 (0.15)	0.72 (0.13)		0.18 (0.05)
Llevant	0.62 (0.12)	6.11 (1.09)	1.51 (0.23)	2.92 (0.42)	-3.97 (0.81)		
Ebro		-2.40 (0.19)	-1.50 (0.20)	-2.23 (0.27)	2.30 (0.25)		

ALT (altitude), LAT (latitude), QUDIME (quadratic distance to the Mediterranean Sea), QUDIAT (quadratic distance to the Atlantic Sea), QUDICA (quadratic distance to the Cantabric Sea), SOLRAD (potential solar radiation) and CURV (terrain curvature).

Table 5. Independent mean test (all basins and all months) results comparison between using only meteorological stations that are inside the basin (non-buffered model) or using also the meteorological stations that surround the basin (buffered model). The R^2 differences are not significant at $p<0.05$.

Interpolation method	Non-buffered model	Buffered model
RG_LI_IDW	0.64	0.66
RG_LI_SP	0.67	0.70
RG_QU_IDW	0.63	0.65
RG_QU_SP	0.67	0.69
RG_LG_IDW	0.63	0.65
RG_LG_SP	0.66	0.69
IDW	0.62	0.63
SP	0.65	0.68

Table 6. Mean test results for the different drainage basin models.

Drainage basin	Mean monthly R^2
Eastern Pyrenees	0.65 (RG_LI_SP)
Septentrional	0.59 (SP)
Duero	0.80 (RG_QU_SP)*
Tajo	0.66 (RG_QU_SP)
Guadiana	0.69 (RG_LI_SP)
Guadalquivir	0.70 (RG_LI_SP)
Meridional	0.73 (RG_LG_SP)*
Segura	0.75 (SP)
Llevant	0.68 (RG_QU_SP)
Ebro	0.86 (RG_LI_SP)

Best interpolation method between brackets
* Best results obtained with non-buffered models

Mailing adress for galley proofs:

Miquel Ninyerola

Department of Animal Biology, Vegetal Biology and Ecology (Botany Unit). Science Faculty (C Building). Bellaterra Campus. 08193 Cerdanyola del Vallès. Autonomous University of Barcelona. Spain.