INTERNATIONAL JOURNAL OF CLIMATOLOGY
Int. J. Climatol. **20**: 1823–1841 (2000)

A METHODOLOGICAL APPROACH OF CLIMATOLOGICAL MODELLING OF AIR TEMPERATURE AND PRECIPITATION THROUGH GIS TECHNIQUES

MIQUEL NINYEROLA[a,*], XAVIER PONS[b] and JOAN M. ROURE[a]

[a] *Department of Animal Biology, Vegetal Biology and Ecology (Unit of Botany), Autonomous University of Barcelona, Bellaterra 08193, Spain*

[b] *Department of Geography and Centre for Ecological Research and Forestry Applications (CREAF), Autonomous University of Barcelona, Bellaterra 08193, Spain*

Received 31 March 1999
Revised 24 March 2000
Accepted 24 March 2000

ABSTRACT

This study proposes an empirical methodology for modelling and mapping the air temperature (mean maximum, mean and mean minimum) and total precipitation, all of which are monthly and annual, using geographical information systems (GIS) techniques. The method can be seen as an alternative to classical interpolation techniques when spatial information is available. The geographical area used to develop and apply this model is Catalonia (32000 km^2, northeast Spain). We have developed a multiple regression analysis between these meteorological variables as the dependent ones, and some geographical variables (altitude (ALT), latitude (LAT), continentality (CON), solar radiation (RAD) and a cloudiness factor (CLO)) as the independent ones. Data for the dependent variables were obtained from meteorological stations, and data for the independent variables were elaborated from a 180 m resolution digital elevation model (DEM). Multiple regression coefficients (b_n) were used to build final maps, using digital layers for each independent variable, and applying basic GIS techniques. The results are very satisfactory in the case of mean air temperature and mean minimum air temperature, with coefficients of determination (R^2) between 0.79 and 0.97, depending on the month; in the case of mean maximum air temperature, R^2 ranges between 0.70 and 0.89, while in the case of precipitation, it ranges between 0.60 and 0.91. Copyright © 2000 Royal Meteorological Society.

KEY WORDS: Catalonia (northeast Spain); GIS; multiple regression analysis; interpolation; climatological modelling; mean maximum air temperature; mean air temperature; mean minimum air temperature; precipitation

1. INTRODUCTION

Many disciplines (biogeography, hydrology, forest management, agriculture, ecology and others) use spatial information about climatological data as a basis for understanding the processes they study. Though there are some useful local models (Azevedo *et al*., 1998; Blennow and Persson, 1998 etc.) to predict climatological variability, our aim is to develop a generalist model over a relatively large area. Nevertheless, available information is usually limited to the meteorological stations and, therefore, to discrete points in space. This makes it necessary to use interpolation techniques to map the corresponding meteorological variables. Traditionally, the method used to interpolate variables had been linear interpolation between stations and the drawing of isolines (contour lines) based on the researcher's knowledge of the area under study (Burrough and McDonnell, 1998). With the spread of computers, more automatic interpolation procedures were proposed and implemented. Interpolation procedures can be simple mathematical models (inverse distance weighting, trend surface analysis, Thiessen polygons etc.), or more complex models (geostatistical methods, such as kriging and thin plate splines). Nonetheless, in most

* Correspondence to: Faculty of Sciences (C building), Department of Animal Biology, Vegetal Biology and Ecology (Unit of Botany), Autonomous University of Barcelona (UAB), Bellaterra 08193 (Barcelona), Spain; e-mail: miquel.ninyerola@uab.es

cases, those models do not take into account geographical information, although there are more sophisticated methods that incorporate this kind of information, such as co-kriging and elevation-detrended kriging techniques (Phillips *et al.*, 1992).

There are more recent works on climatological interpolation, dealing with the factors that model it, such as the research of Bigg (1991), that uses kriging interpolation, adding the influence of some variables in the process, and the research of Hutchinson (1995), that uses aspect to improve splines interpolation. There are also studies that seek for statistical relationships between geographical variables (orography, latitude and continentality), or landscape variables and climatological variables (Berndtsson, 1989; Benzi *et al.*, 1997; Chessa and Delitala, 1997; Hargy, 1997; Llasat, 1997; Vogt *et al.*, 1997), as well as studies that use geographical information systems (GIS) to model these climatological variables (Gessler *et al.*, 1995; Hutchinson, 1995; Menz, 1997), but there are few works encompassing both knowledge areas. An example of this approach can be found in Blennow and Persson (1998) and in Blennow (1998), where GIS tools are used to obtain sky view factors from digital elevation models (DEM) and thematic maps, as well as multivariate statistics to correlate climatological variables with environmental factors. Chuvieco and Salas (1996) also use GIS and multiple regression, but they restrict themselves to modelling maximum air temperature during summer months, given that their aim is to obtain a fire danger index. There also exist works in other disciplines that use the regression modelling approach, as in the case of Narumalani *et al.* (1997).

Our main goal is to develop an objective mapping and interpolation method that takes into account geographical variables, and uses GIS techniques to obtain final maps of different climatological variables over a relatively large area. The idea of an objective mapping it is not new (Hamilton *et al.*, 1988; Chessa and Delitala, 1997; Hargy, 1997; Menz, 1997), but the use of GIS techniques, together with multiple regression analysis, can significantly improve the techniques for elaborating such an objective mapping. In fact, some recent works, such as the one from Sánchez Palomares *et al.* (1999) uses similar approaches, although not modelling minimum, nor maximum air temperatures, and only using altitude and X, Y location as independent variables.

We attempt to develop an empirical and statistical procedure to forecast the mean maximum air temperature, mean air temperature, mean minimum air temperature and total precipitation (at monthly and annual timescales) in Catalonia (northeast Spain). This procedure is empirical because it uses data obtained from meteorological stations for building and for validating the model. This procedure is also statistical because it is based on a multiple regression analysis and its corresponding validation. As dependent variables, we used air temperature or precipitation, whereas as independent variables, we used altitude (ALT), latitude (LAT), continentality (CON) and solar radiation (RAD) (in the case of air temperature) or a cloudiness factor (CLO) (in the case of precipitation). The procedure and the selected variables used in this study result, from our point of view, in a very simple, useful and realistic model. Obviously, in further studies, it would be possible to increase its complexity, if it is thought appropriate, to add new variables or change the way they contribute to explain the variability, but as will be seen, current results give very good outcomes.

This model aims to predict air temperature and precipitation, merely in a numerical and objective way. Thus, at least in the beginning, the intuition and knowledge of the researcher are lost (except when the variables are chosen). It is *a posteriori* when the model can be changed or improved by an expert.

The outcome of this work is 52 maps, classified as follows: 13 maps (monthly and annual) of the mean maximum air temperature, 13 of mean air temperature, 13 of mean minimum air temperature, and 13 of total precipitation. There also exists a corrector map for each one of the final maps.

2. SITE DESCRIPTION

Catalonia, located in the northeast of the Iberian Peninsula, is the geographical area where this work is centred. It is a 32000 km^2 area, located on the Mediterranean coast. It has high variability and very strong contrasts resulting from its relief (Bolòs, *et al.* 1983) and its geographical situation, receiving

Mediteranean, Atlantic (although attenuated by the Iberian range and the Pyrenees, Martín Vide (1987)), and even Saharan influences. Even though we can characterize it as having a typical Mediterranean climate, as described by the pluviothermic index of Emberger (1952), some parts of it are classified as semiarid, subhumid and humid (Debazac, 1983 and Piñol *et al*., 1991). More information on its climate can be found in Bolòs *et al*. (1983). Figure 1 shows the DEM used in this research.

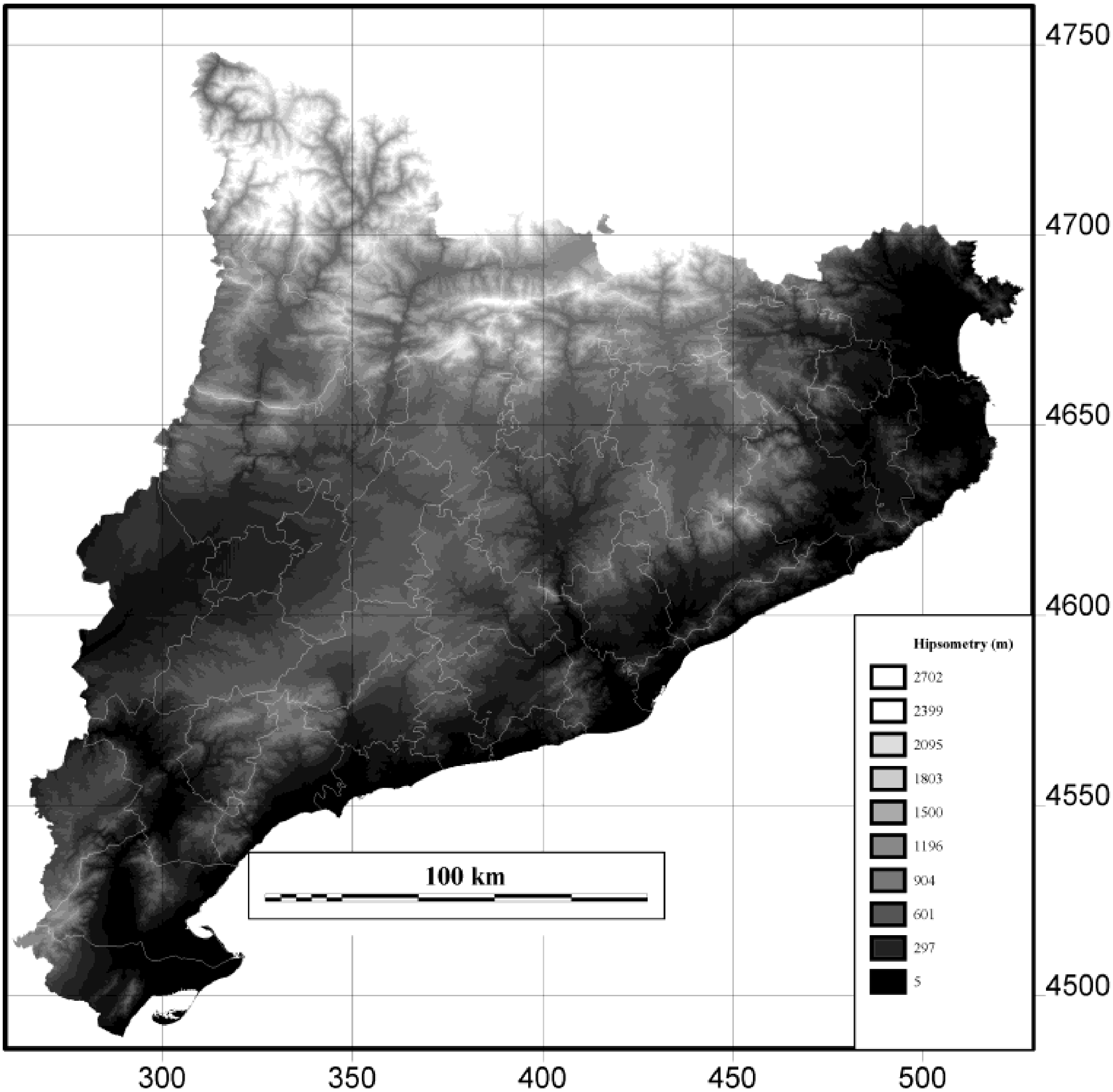

Figure 1. General view of the DEM of Catalonia, in universal transversal mercator (UTM) projection

3. MATERIAL AND METHODS

The density of meteorological stations in Catalonia is one station for every 64 km^2 but their locations are not ideal for our purposes (Figure 2). Indeed, a lot of stations are placed in flat areas, so that they are lacking in the steepest areas, even in the Pyrenees region. As a result, in the case of solar radiation, for example, all the stations are quite homogeneous and, as we will see later, they contribute less than expected in the multiple regression model.

3.1. Filtering data

Even though there are different criteria to filter climatological data, in this work we have placed emphasis on spatial patterns, always bearing in mind the length of the series. The World Meteorological Organization gives 30 years as the optimal length for a series (WMO, 1967), but it will depend on the variability in the area under study. In this case, for each month, we have filtered out the air temperature series within 15 years, and the precipitation series within 20 years. In the case of annual data, we have filtered with the same criteria, but only keeping meteorological stations with complete years. These figures have been obtained by trying to achieve a compromise between series length and density of stations

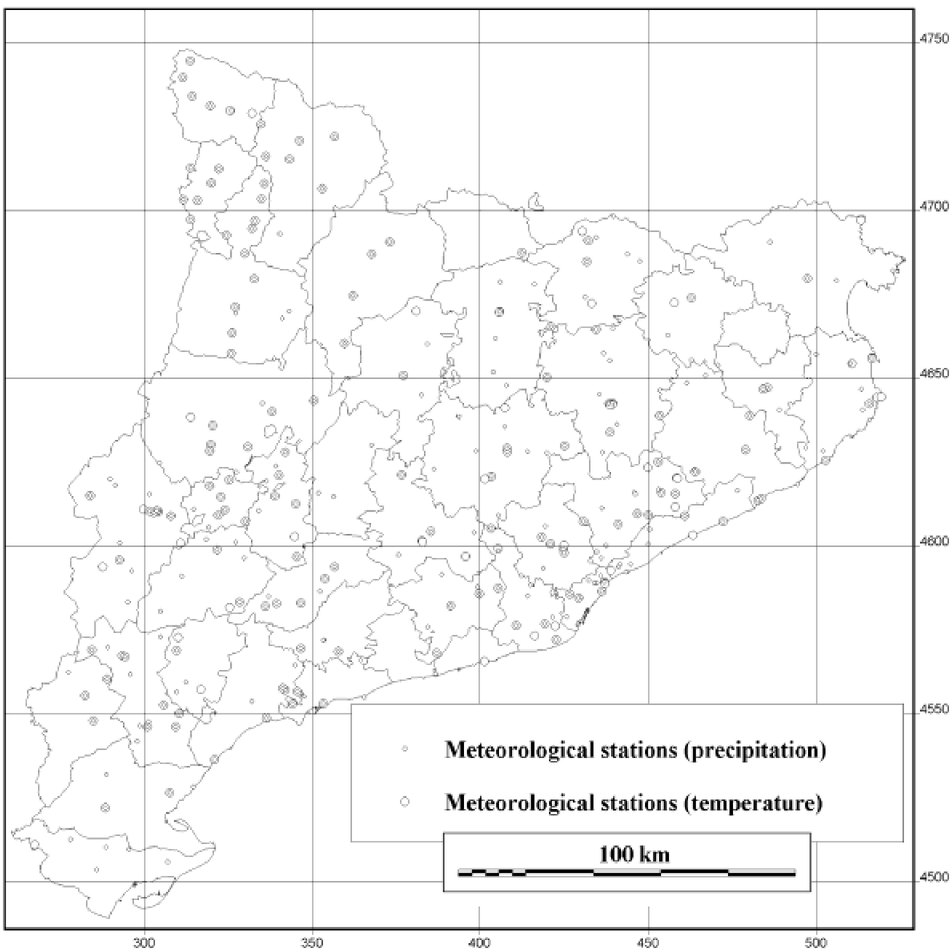

Figure 2. Geographical location of the filtered meteorological stations. Because of the fact that each month can have a different number of meteorological stations, we represent, for each case, precipitation and temperature, the months with a mean number of meteorological stations (April and February, respectively)

(Hubbard, 1994), particularly critical in the case of precipitation. Martín Vide (personal communication) helped us in the final choice of the series length. In any case, it is obvious that the process can be repeated with longer series in the next few years.

Finally, we have worked with 160 air temperature stations (from the 463 initial ones) and with 257 precipitation stations (from the 695 initial stations). Thus, the density of the filtered stations is one every 200 km^2 for air temperature, and one every 125 km^2 for precipitation. In the case of annual data, we have worked with 110 air temperature stations and with 166 precipitation stations.

All the stations were provided by the Instituto Nacional de Meteorología (INM) of Spain, and the series belong to the 1951–1999 period.

3.2. Choosing dependent variables

From INM data, we have calculated monthly and annual means of the series with complete years. For air temperature, we have chosen mean maximum air temperature, mean air temperature and mean minimum air temperature. For precipitation, we have chosen total precipitation, because the data of the days of precipitation, also available, gave us insufficient results in preliminary tests, probably because of the quality of the initial data and/or its unpredictable behaviour.

3.3. Choosing independent variables

In this case, the choice has been made following some of the most acknowledged factors cited in the literature that contribute to the climate. For air temperature, we have ALT, LAT, CON and RAD. For precipitation, we use the same variables, except RAD. This variable has been replaced by CLO. Although easting is used in some other works (Hargy, 1997), we have not used it because it does not seem to explain much variability in this case.

Although terrain aspect could be regarded as another independent variable, it is not interesting in our case, neither as solar illumination, nor as wind factor. Indeed, terrain aspect as an illumination factor is implicit in the RAD variable, and in a better way than merely accounting for aspect. Nevertheless, an aspect such as wind factor results in noise into our general model because the role of wind and orography can be different, depending on each local area; in fact, we have done some trials, particularly in precipitation, and aspect never helped to improve the model, nor was it significant.

Variable ALT is the nominal altitude of the stations, and it gives information on the variability caused by the relief.

LAT is, in fact, the cosine of the nominal latitude of the stations because, as a result of the curvature of the Earth, it is a better measure than the straight latitudinal value. LAT has not been included in the RAD model because, owing to Catalonia's limited latitudinal range, the RAD model takes the central point of the DEM as reference. Therefore, we have to include the LAT in the air temperature and precipitation models, because variation in LAT is not considered in the RAD variable. In this manner, LAT will get us close to the variability generated by the zonal atmospheric circulation.

CON has not been available from the nominal data of the meteorological stations, so it has been modelled using the linear distance to the sea with the distance module of Idrisi (Eastman, 1997). Other non-linear models have also been tested (sigmoid models); in these tests, distance is measured as the least effort in moving over a friction surface (cost module of the same software), accounting for the influence of abrupt differences in relief, especially in the first ALTs. Nevertheless, the best results are obtained with the linear model (when the sigmoid function becomes a straight line). That is probably a result of the distribution of the meteorological stations and the relief in Catalonia (only a few stations are directly faced to the sea). Other works (Hargy, 1997) use the logarithm of the distance to the coast as an independent variable, but, in our case, this has also worked out worse than using the linear distance. We plan to study this phenomenon in depth in the future, although, as Driscoll and Yee Fong (1992) state, it is very difficult to model this variable.

Finally, the RAD variable has been obtained from a potential radiation model proposed by Pons (1996), and implemented in the MiraMon software (Pons, 1998). This model is entirely computational, is based in a DEM and uses a constant optical density of the atmosphere ($\tau = 0.288$); this optical density is an empirical estimation, valid for an average clear forest atmosphere and, therefore, it does not account for clouds (Rothermel *et al.*, 1986; Dozier, 1989). The model takes into account the solar constant, Earth–Sun distance, solar geometry and incident angles at each location (slope/aspect), cast-shadowed and self-shadowed areas, and direct versus diffuse radiation. Moreover, this potential radiation model has been calibrated using data from meteorological stations; the calibrated model is used as an independent variable in the case of air temperature. The model also provides a CLO (by comparing potential and measured radiation) that is used for modelling precipitation. This factor expresses, in fact, the lack of clouds.

Note, therefore, that for developing the multiple regression analysis, we use the nominal values of the meteorological stations for the case of ALT (in meters) and LAT and, for the case of CON (in km) and RAD (in 10 kJ/m^2.day) (or CLO), we obtain the values from the models explained above. The obtention of these values from digital maps for each meteorological station have been done using the XY_DBF module of the MiraMon software; this module allows entering a set of X, Y coordinates, and extracting the value of each point placed in the corresponding location of the map.

3.4. Regression model: some considerations

We have used a multiple regression analysis using the *backward stepwise* method for choosing the independent variables included in the model. In the 'Results' section, we show the multiple coefficient of determination (R^2), as well as multiple regression coefficients (b_n). These coefficients were used, as explained below, to build the final maps. All the calculations have been done with $\alpha = 0.05$. p-values regarding the relationship between the dependent variable and the independent variables are not shown because, in all cases where they are significant, these values are very low (at least lower than 0.001). This method has been chosen for its simplicity, and because it is one of the most widely used techniques when the aim is to develop an empirical model (Lanzante, 1996).

3.5. Fitting and testing the model

We randomly put aside 40% of the filtered stations (test set), and we ran the multiple regression analysis for the 60% remaining stations (fitting set). Then we compared the values predicted by the model with the ones measured on the test stations. We will term R^2_{nc} referring to the coefficient of determination resulting from that comparison; the 'nc' subscript denotes that the R^2 corresponds to a model that is still 'non-corrected' in the sense we propose later in subsection 3.7.

This comparison, using 40% of independent stations, gives a measure of the reliability of the model, because it accounts for its ability for predicting climate values on stations not included in the fitting process. Nevertheless, as the final maps have been built using the whole set of stations (to preserve as much spatial variability as possible), the final reliability is at least equal or better than that obtained with the test set. Then we calculated the value of the dependent variable for the test stations using these coefficients. The fitting set served to adjust the model, by obtaining the coefficients for each significant independent variable. This model is refined by using its own residuals (correctors), as we will explain later (subsection 3.7).

3.6. Mapping the model

Once the model is considered good enough to be spatially representative, it can be applied by using the following expression (air temperature models):

$$y = b_0 + b_1(\mathrm{ALT}) + b_2(\mathrm{LAT}) + b_3(\mathrm{CON}) + b_4(\mathrm{RAD}) \tag{1}$$

where b_n are the multiple regression coefficients, adjusted for each month and for each dependent variable (mean air temperature etc.) from the data of the meteorological stations. In the case of the precipitation model, the equation is basically the same, but with the variable RAD replaced by the variable CLO. The annual maps have been computed from monthly maps, and not from multiple regression equations, although we also provide the results for the annual cases.

To apply Equation (1), we have prepared a set of map layers in raster format, i.e. as geographical matrices where each point, regularly spaced, represents a datum. Data in these matrices correspond to the independent variables of the model (ALT, LAT etc.), and are all obtained from the DEM, as explained before. Operating these raster maps has been performed using the module CALCIMG of the MiraMon software that uses map algebra techniques (Tomlin, 1990; Burrough and McDonnell, 1998, p. 184). That is, the module multiplies the cells of each raster map by the corresponding regression coefficient for each month, makes the addition of all the rasters and adds the interception value (b_0). These techniques allow us to obtain what we call the *potential maps*. The term *potential maps* has been chosen because the maps result from the multiple regression model; these maps could also be named 'predicted maps', but we prefer the term 'potential' because they have been obtained before applying the correctors we will explain in subsection 3.7; these 'potential maps' will be converted into 'real maps' through the correctors; the term 'real' will be used because at the points where a meteorological station is located, the real map will show the same values of the meteorological station.

The combination of regression models and map algebra techniques has been applied in other works, provided that the independent variables are available over all the study area. Usually, these models are inexact interpolators in a spatial sense (Burrough and McDonnell, 1998, p. 112). Nevertheless, as we explain in subsection 3.7, in our case, and owing to the use of corrector map layers, our method results in an exact interpolator. In fact, the method can be seen as an alternative to classical interpolation techniques when spatial information is available.

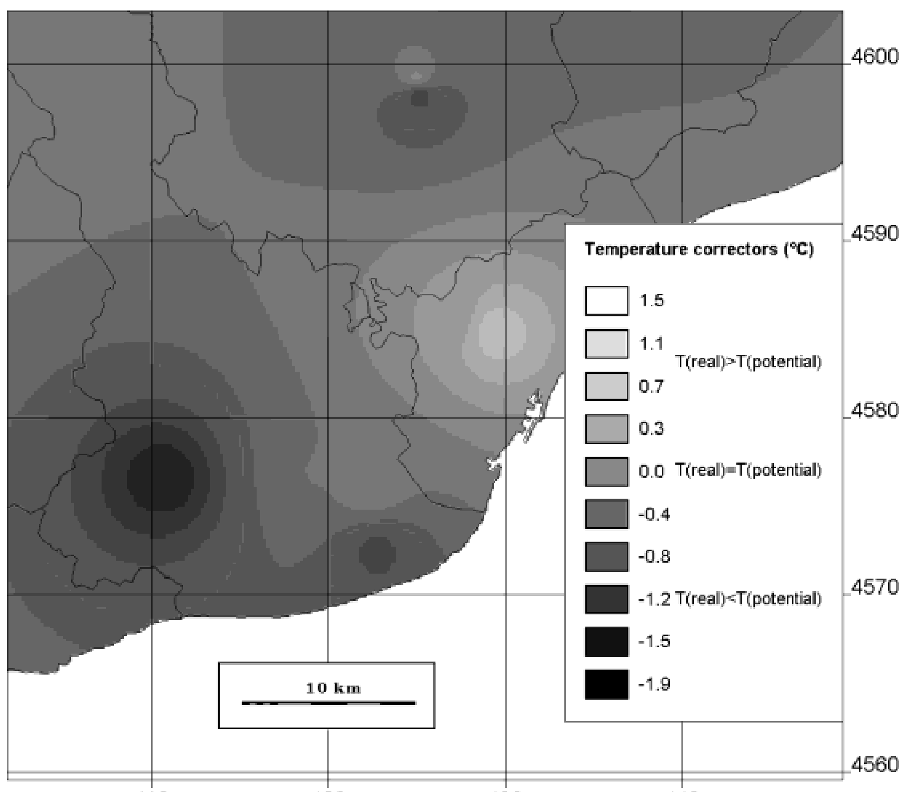

Figure 3. Zoom view of the anomaly digital map of mean annual temperature, showing in detail Barcelona city and its neighbourhood. Pale colours represent heat islands and dark colours the opposite

To build the final air temperature and precipitation maps, we start working with one raster map for each of the independent variables used in the multiple regression analysis. The ALT raster has been obtained from a DEM with a resolution of 180 × 180 m. The LAT and CON rasters have been built from this DEM, using the distance module of Idrisi (Eastman, 1997). In the first case, we compute the linear distance to the bottom row of the raster map because our study area is in the northern hemisphere, and the maps are in universal transverse mercator projection. In the second case, we also use the linear distance, but the target feature was the coast of Catalonia.

For RAD and CLO, as mentioned above, we used data from previous works. In the case of RAD, it is important to note that we used a model based on a 500 × 500 m DEM, instead of the 180 × 180 m one. We have done so because, in previous tests, we had observed better fittings using a 500 × 500 m DEM than using a DEM with higher resolution. The reason for this is probably that the relationship between RAD and air temperature is less correlated with detailed orographic factors than with the general location of the station, as the temperature at each point is buffered by the air itself.

3.7. Building the correctors to refine results

Each cell of the potential maps has the values of air temperature and precipitation predicted from the results of the regression model. However, if we compare predicted values (obtained from the model) with observed values (obtained from the stations), the differences give a good estimation of the residual error in the regression analysis for each geographical location where we have field data (meteorological

Table I. Results of the multiple regression analysis in the case of mean temperature

Month	Multiple regression coefficients (b)			R^2_{nc}	R^2_c
January	ALT = −0.003	CON = −0.026	RAD = ns	0.76	0.85
	LAT = ns	Interception = 8.281			
February	ALT = −0.005	CON = −0.016	RAD = ns	0.82	0.84
	LAT = ns	Interception = 9.613			
March	ALT = −0.005	CON = ns	RAD = ns	0.85	0.86
	LAT = 96.701	Interception = −60.663			
April	ALT = −0.006	CON = ns	RAD = ns	0.86	0.86
	LAT = 71.332	Interception = −39.506			
May	ALT = −0.006	CON = 0.01	RAD = ns	0.86	0.85
	LAT = 113.865	Interception = −67.875			
June	ALT = −0.006	CON = 0.015	RAD = ns	0.84	0.85
	LAT = 164.964	Interception = −102.223			
July	ALT = −0.006	CON = 0.017	RAD = ns	0.83	0.83
	LAT = 198.921	Interception = −124.387			
August	ALT = −0.006	CON = 0.013	RAD = ns	0.85	0.84
	LAT = 197.845	Interception = −123.621			
September	ALT = −0.005	CON = ns	RAD = ns	0.88	0.86
	LAT = 134.546	Interception = −78.99			
October	ALT = −0.005	CON = ns	RAD = ns	0.88	0.87
	LAT = 111.062	Interception = −66.291			
November	ALT = −0.004	CON = −0.021	RAD = ns	0.81	0.85
	LAT = ns	Interception = 12.069			
December	ALT = −0.003	CON = −0.026	RAD = ns	0.75	0.85
	LAT = ns	Interception = 8.909			
Annual	ALT = −0.005	CON = ns	RAD = ns	0.95	0.97
	LAT = 99.192	Interception = −58.551			

The multiple regression coefficients calculated with all the meteorological stations (b), the coefficient of determination of the non-corrected model (R^2_{nc}), and the coefficient of determination of the corrected model (R^2_c) are shown. R^2_{nc} and R^2_c figures are based on the independent test done using 40% of the stations from the model built upon 60% of the stations.
'ns' means that the variable is not significant.

stations). In this way, we have obtained, for each meteorological station, an error value that we call the *corrector*. We have chosen an additive corrector (Equation (2)), so that it would be unrelated with the magnitude of the values.

$$\text{Corrector} = \text{Temperature}_{\text{observed}} - \text{Temperature}_{\text{predicted}} \quad (2)$$

At this step, it would be interesting to use the correctors to refine the results, that is, to correct the potential maps into *real maps*. Nevertheless, as these correctors are irregularly distributed in space, it is necessary to interpolate them. If we interpolate the corrector values of each station and dependent variable over the whole area, we will obtain *corrector maps* that will estimate the error of the model at each geographical point. Because of the good results obtained by this methodology, we have used a typical interpolation technique, based on the inverse of the squared distance. Also, we have calculated the corrector maps using kriging, a more sophisticated interpolation technique (Phillips *et al.*, 1992; Ashraf *et*

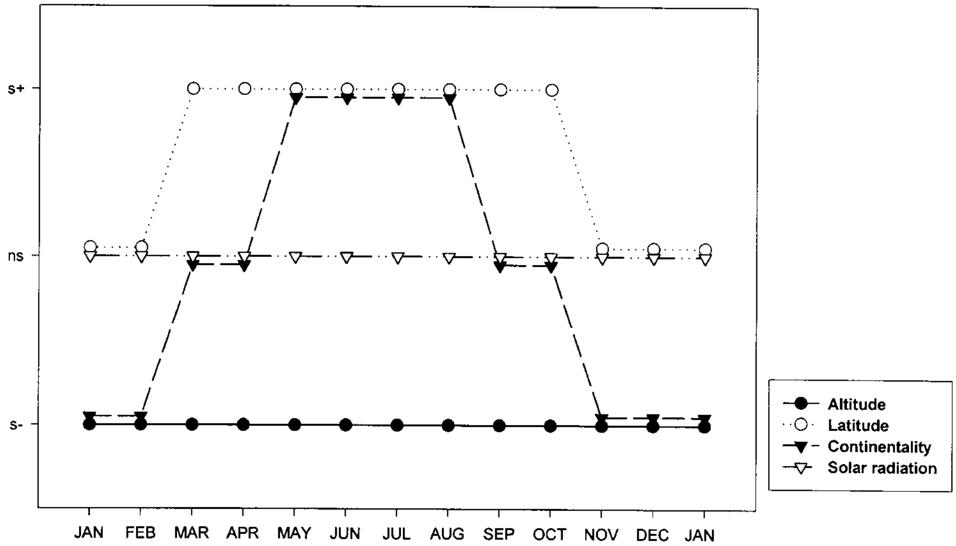

Figure 4. Statistical importance of the independent variables during the year in the case of mean temperature. 's + ' means that the variable is significant ($\alpha = 0.05$), and has a positive contribution, 's − ' means that the variable is significant ($\alpha = 0.05$), and has a negative contribution, and 'ns' means that it is not significant

Table II. Statistical parameters of the mean temperature (°C) for the 990 375 cells of the matrix inside Catalonia

Month	Mean	Standard deviation	Minimum value	Maximum value
January	4.6	2.8	−4.7	11.7
February	5.6	3.3	−6.7	11.7
March	8.2	3.2	−4.3	13.3
April	10.1	3.5	−4.2	15.4
May	14.0	3.4	−0.3	19.6
June	17.9	3.5	3.3	23.4
July	21.3	3.6	6.6	27.0
August	21.1	3.7	6.3	26.7
September	18.1	3.3	5.3	23.2
October	13.2	3.3	0.7	18.9
November	8.3	3.0	−2.5	14.7
Dec	5.3	2.8	−3.9	11.2
Annual	12.3	3.2	0.0	17.3

al., 1997) with the aim of comparing both techniques. The corrector maps of the final real maps (calculated with all the stations) have been built with the interpolator (inverse of the squared distance or kriging) that has shown better results for each month and dependent variable.

These corrector maps will not be uniform, but they will show maximum variability in the more unpredictable areas, and minimum variability in the predictable sites. In this sense, corrector maps can be seen as *anomaly maps*, of great interest to reveal the singularities of the climate at the local scale. The most unpredictable areas are usually correlated with rugged zones.

By adding the corrector maps to the potential maps (Equation (3)) we will obtain *real maps*. Obviously, in the cells that match meteorological stations, the observed values will be equal to the measured ones, which is also interesting for the use of our maps. In the rest of the cells, we will obtain predicted values modified by these correctors.

$$T_{\text{observed}} = T_{\text{predicted}} + \text{Corrector} \quad (3)$$

In Equation (3), we can see that the Corrector is (a) the difference between $T_{\text{observed}} - T_{\text{predicted}}$ at the meteorological stations, and (b) the interpolation among the correctors in the meteorological station at the other points of the map. We will term R_c^2 (corrected) to denote the co-efficient of determination of the corrected model that uses the coefficients of the fitting set and the anomaly maps (interpolation of the fitting residuals). The R_c^2 coefficient is obtained by reading in the corrected maps the values at the points of the 40% of the meteorological stations, and calculating the co-efficient of determination of the new fit resulting from comparing the data from the stations with the data from the corrected maps.

3.8. Considerations about the meaning of the anomaly maps

Anomaly maps express the climatic variability which is not explained by our model. Such residual variability can come from the initial data: the digital geographical information (positional error and value error) and the climatological information of the stations (positional errors, errors in the calibration of the measuring instruments, in the readings of the data obtained by the devices, and in the transcription of the information, as reported by Martín Vide (1987)). Residual variability could also come from our data

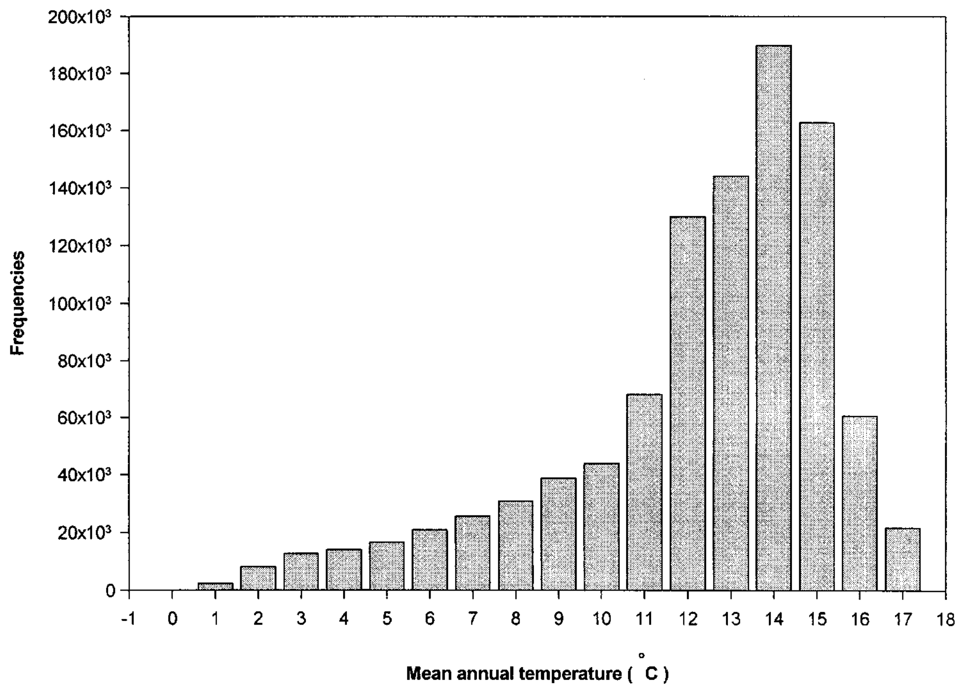

Figure 5. Histogram of the mean annual temperature. The high frequency values are caused by the matrix size (990375 cells)

manipulation, as well as from inherent factors of the model (omission of relevant parameters, non-linear behaviour of some variables and method of interpolation for correctors). Figure 3 represents an anomaly map zoom, from Barcelona and its neighbourhood, where the heat island (observed values higher than predicted values) of the city is very evident.

4. RESULTS AND DISCUSSION

4.1. Mean air temperature

We can see in Table I that all the co-efficients of determination of the non-corrected model (R^2_{nc}) range from 0.75 to 0.95. Good co-efficients of determination, using similar techniques, can also be found in Hargy (1997). In the corrected model, the coefficients (R^2_c) are slightly better (from 0.83 to 0.97). Therefore, the model is improved when it is corrected with the meteorological data but, as non-corrected model fittings are already quite good (are predictive), this improvement is not very important, and, even in 4 months, is slightly worse.

In all cases, the R^2_c using the inverse of squared distance as the interpolation technique for the correctors is always better than kriging (the mean R^2_c for all months in the former case is 0.85, while in the latter case, it is 0.83), so the real maps with all the stations were built always from inverse weighting distance correctors.

Finally, note that all the months have similar predictors, except the annual data that, by far, is the best predicted one.

Figure 4 shows the statistical significance of the independent variables during the year. ALT is significant in every month, and always has negative multiple regression coefficients (b). This means that

Table III. Results of the multiple regression analysis in the case of mean maximum temperature

Month	Multiple regression coefficients (b)			R^2_{nc}	R^2_c
January	ALT = −0.004	CON = −0.016	RAD = ns	0.67	0.76
	LAT = ns	Interception = 12.638			
February	ALT = −0.006	CON = ns	RAD = ns	0.78	0.78
	LAT = ns	Interception = 14.556			
March	ALT = −0.007	CON = 0.019	RAD = ns	0.73	0.76
	LAT = 105.296	Interception = −61.754			
April	ALT = −0.007	CON = 0.026	RAD = ns	0.78	0.81
	LAT = 148.063	Interception = −91.593			
May	ALT = −0.007	CON = 0.033	RAD = ns	0.71	0.77
	LAT = 165.157	Interception = −100.787			
June	ALT = −0.008	CON = 0.041	RAD = ns	0.69	0.77
	LAT = 228.865	Interception = −144.523			
July	ALT = −0.007	CON = 0.045	RAD = ns	0.65	0.74
	LAT = 254.219	Interception = −160.002			
August	ALT = −0.007	CON = 0.038	RAD = ns	0.68	0.75
	LAT = 217.356	Interception = −132.676			
September	ALT = −0.007	CON = 0.031	RAD = ns	0.74	0.78
	LAT = 195.904	Interception = −119.807			
October	ALT = −0.007	CON = 0.01	RAD = ns	0.78	0.78
	LAT = ns	Interception = 22.088			
November	ALT = −0.006	CON = ns	RAD = ns	0.74	0.78
	LAT = ns	Interception = 16.405			
December	ALT = −0.004	CON = −0.026	RAD = ns	0.61	0.70
	LAT = −121.55	Interception = 104.364			
Annual	ALT = −0.006	CON = 0.009	RAD = ns	0.87	0.89
	LAT = ns	Interception = 21.271			

Information as with Table I.

mean air temperature reduces with increasing ALT. CON is less important in the spring months, probably because the air temperature gradient between the interior and coast is smaller. It has a negative '*b*' during winter (when the interior is colder), and positive '*b*' during summer (when inner land is hotter). RAD is never significant, probably because of the lack of meteorological stations in the steepest areas, as

Table IV. Results of the multiple regression analysis in the case of mean minimum temperature

Month	Multiple regression coefficients (*b*)			R^2_{nc}	R^2_c
January	ALT = −0.002	CON = −0.036	RAD = ns	0.68	0.81
	LAT = ns	Interception = 3.916			
February	ALT = −0.003	CON = −0.032	RAD = ns	0.72	0.81
	LAT = ns	Interception = 4.681			
March	ALT = −0.004	CON = −0.028	RAD = ns	0.76	0.81
	LAT = ns	Interception = 6.627			
April	ALT = −0.004	CON = −0.019	RAD = ns	0.77	0.81
	LAT = ns	Interception = 8.392			
May	ALT = −0.004	CON = −0.018	RAD = ns	0.79	0.82
	LAT = ns	Interception = 12.004			
June	ALT = −0.004	CON = −0.012	RAD = ns	0.82	0.84
	LAT = 100.042	Interception = −59.157			
July	ALT = −0.004	CON = −0.012	RAD = ns	0.84	0.84
	LAT = 143.651	Interception = −88.795			
August	ALT = −0.004	CON = −0.013	RAD = ns	0.85	0.86
	LAT = 177.98	Interception = −114.301			
September	ALT = −0.004	CON = −0.018	RAD = ns	0.83	0.86
	LAT = 149.033	Interception = −95.294			
October	ALT = −0.004	CON = −0.032	RAD = ns	0.80	0.84
	LAT = ns	Interception = 12.236			
November	ALT = −0.003	CON = −0.036	RAD = ns	0.74	0.81
	LAT = ns	Interception = 7.605			
December	ALT = −0.002	CON = −0.035	RAD = ns	0.69	0.79
	LAT = ns	Interception = 4.699			
Annual	ALT = −0.004	CON = −0.027	RAD = ns	0.84	0.85
	LAT = ns	Interception = 10.755			

Information as with Table I.

Table V. Statistical parameters of the mean maximum temperature (°C) for the 990 375 cells of the matrix inside Catalonia

Month	Mean	Standard deviation	Minimum value	Maximum value
January	9.1	2.8	−1.3	16.2
February	10.8	3.3	−3.0	17.6
March	13.7	3.6	−2.2	19.4
April	15.9	3.8	−0.5	21.7
May	19.9	3.7	3.6	25.8
June	23.9	4.1	5.4	30.2
July	27.9	3.7	10.3	34.2
August	27.3	3.7	10.2	33.5
September	23.8	3.7	7.3	29.4
October	18.4	3.5	3.1	25.4
November	12.8	3.3	−0.5	19.8
December	9.6	2.8	−0.9	16.7
Annual	17.7	3.3	3.3	22.5

explained in section 3. LAT has positive '*b*' throughout the year, except in winter months. This means that mean air temperature reduces with increasing LAT (note that we use cosine of the LAT). We plan to explore these patterns in depth in the future. When examining these results, it should be noted that the independent variables are correlated in the case of Catalonia (i.e. a northern latitude and a longer distance to the sea usually correspond to a higher altitude).

Plate 1 shows the digital map of mean annual air temperature. Table II shows the descriptive statistics for this map. The extreme months are January ($T = 4.6$°C) and July ($T = 21.3$°C). These values are computed, as for the other dependent variables discussed later, from the whole set of 180 × 180 m cells of Catalonia (990 375), and not from the meteorological stations. As explained in the introduction, maps were computed for each month. Finally, Figure 5 shows the histogram of frequencies for the mean annual air temperature.

Table VI. Statistical parameters of the mean minimum temperature (°C) for the 990 375 cells of the matrix inside Catalonia

Month	Mean	Standard deviation	Minimum value	Maximum value
January	0.1	2.8	−8.2	9.7
February	0.6	3.1	−9.5	8.3
March	2.5	3.2	−8.5	9.9
April	4.5	3.1	−6.3	11.9
May	8.2	3.0	−2.5	15.1
June	12.0	3.2	0.7	18.7
July	14.9	3.4	3.2	21.8
August	15.0	3.5	3.0	22.4
September	12.3	3.5	0.4	19.9
October	7.9	3.4	−3.5	15.4
November	3.6	3.1	−5.9	11.6
December	1.0	2.8	−7.0	8.2
Annual	6.8	3.1	−3.4	14.1

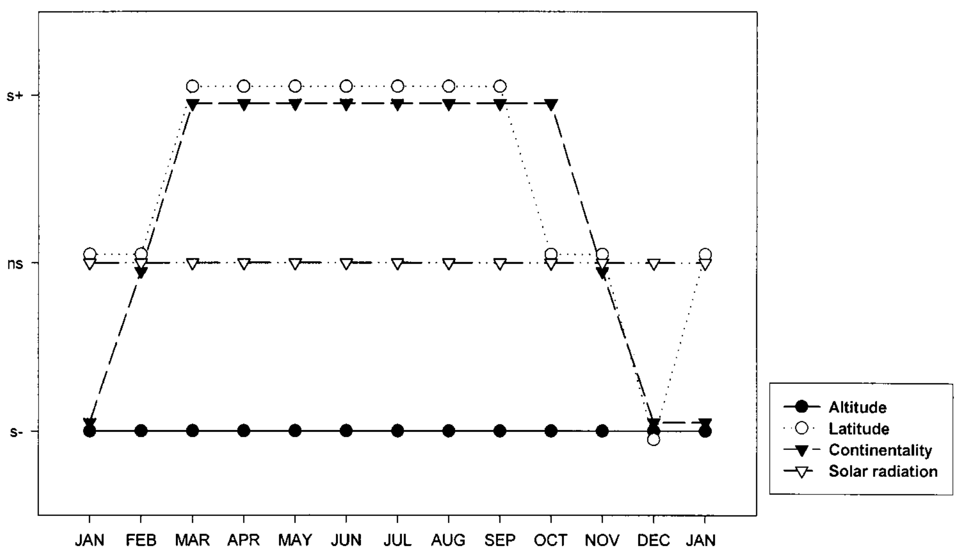

Figure 6. Statistical importance of the independent variables during the year in the case of mean maximum temperature. 's + ' means that the variable is significant ($\alpha = 0.05$), and has a positive contribution, 's − ' means that the variable is significant ($\alpha = 0.05$), and has a negative contribution, and 'ns' means that it is not significant

4.2. *Mean maximum air temperature and mean minimum air temperature*

The statistical parameters of the multiple regression analysis are shown in Tables III and IV and the descriptive statistics in Tables V and VI. We represented the statistical importance of the independent variables during the year in Figures 6 and 7, the mean annual maps in Plate 1, and the histogram of frequencies in Figures 8 and 9.

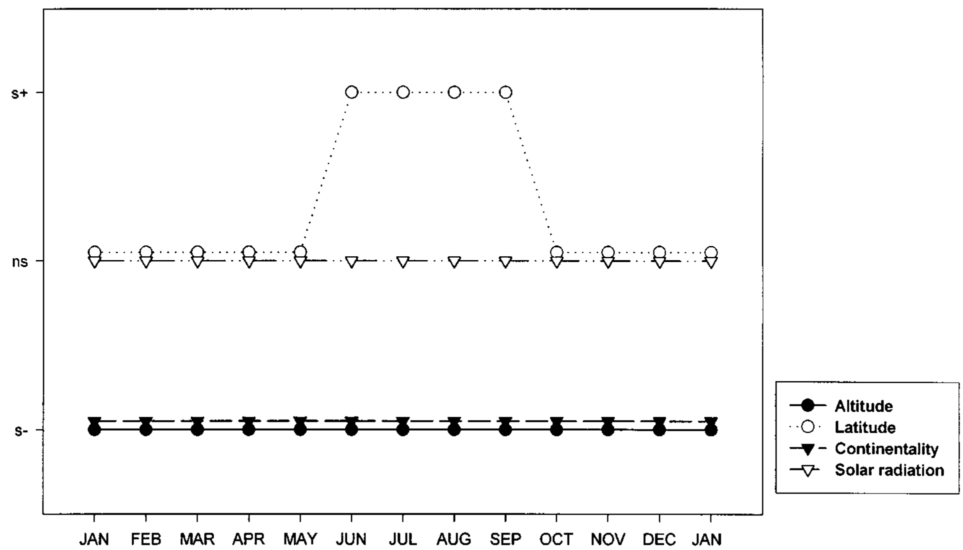

Figure 7. Statistical importance of the independent variables during the year in the case of mean minimum temperature. 's + ' means that the variable is significant ($\alpha = 0.05$), and has a positive contribution, 's − ' means that the variable is significant ($\alpha = 0.05$), and has a negative contribution, and 'ns' means that it is not significant

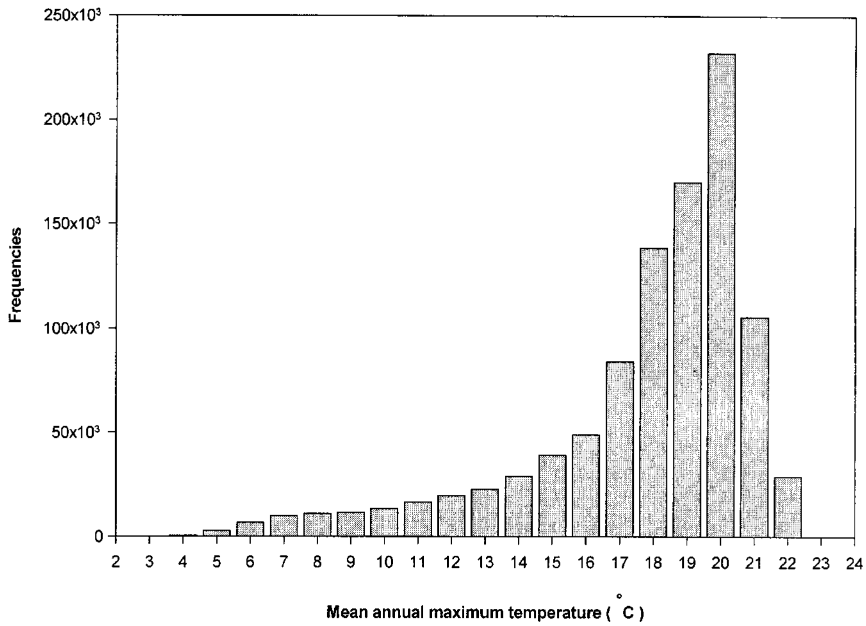

Figure 8. Histogram of the mean annual maximum temperature. The high frequency values are caused by the matrix size (990375 cells)

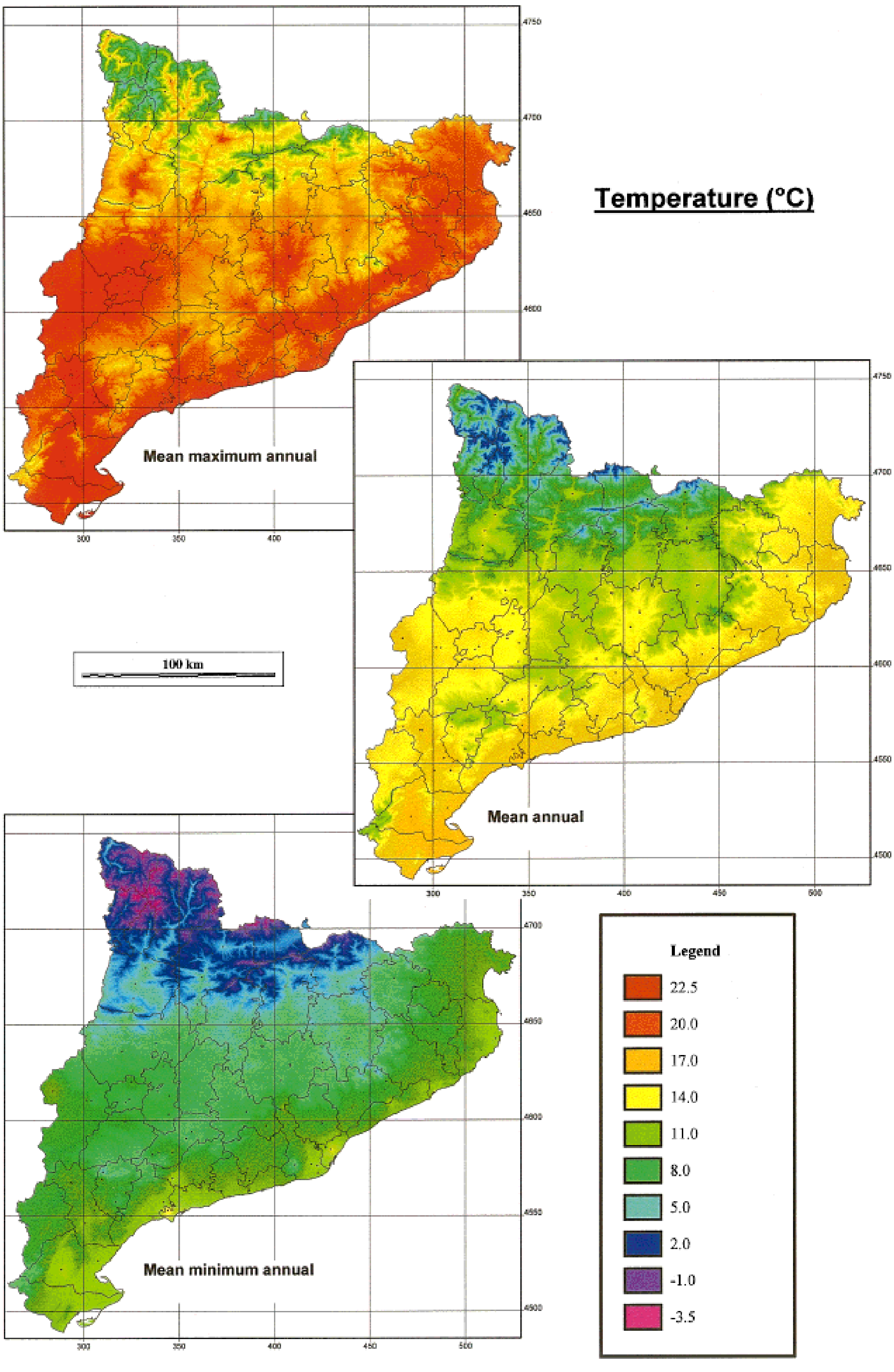

Plate 1. General view of the digital maps of mean annual maximum temperature, mean annual temperature and mean annual minimum temperature. For helping the comparison among these maps, the values of the legend range between the absolute minimum and the absolute maximum

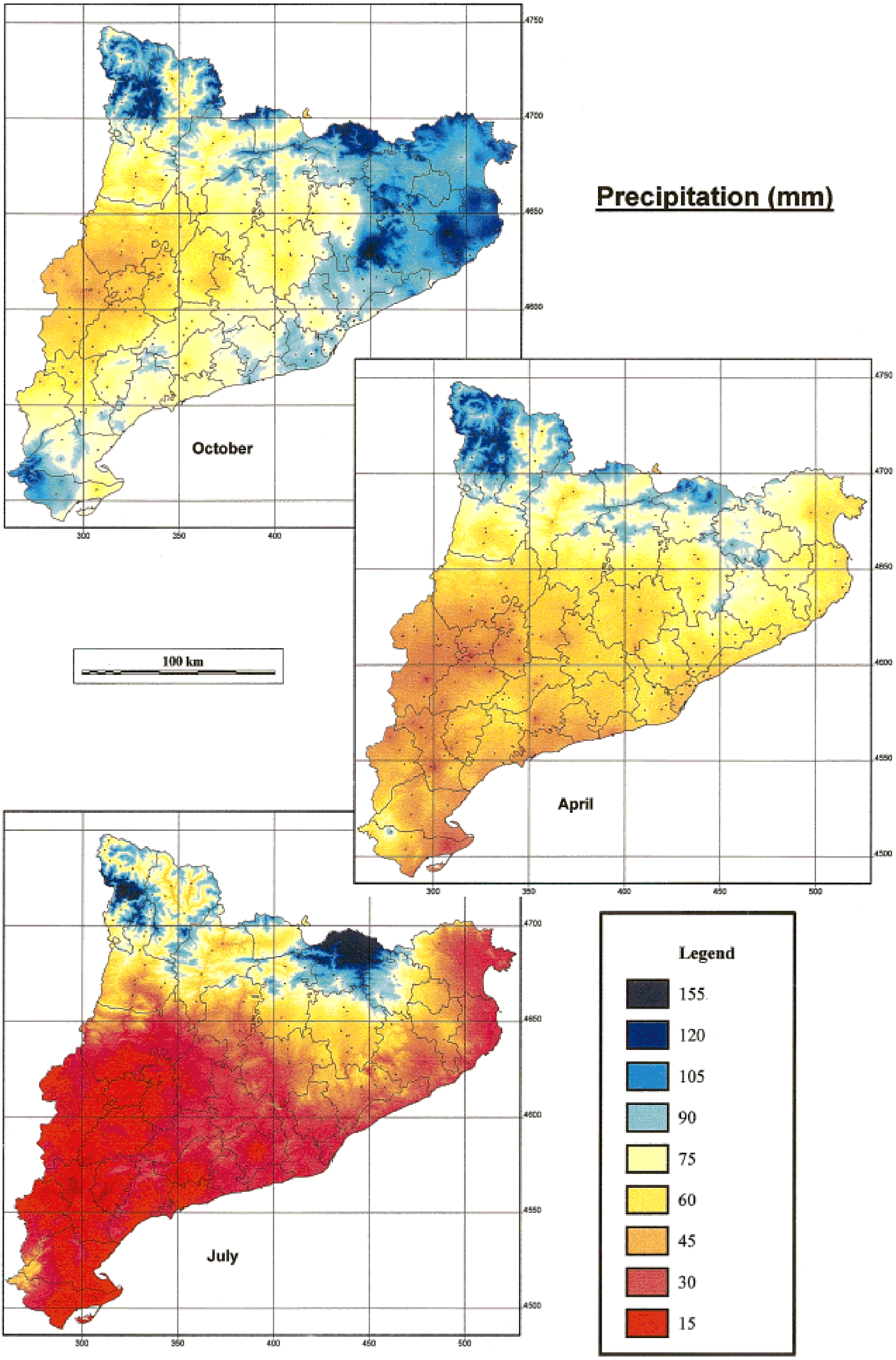

Plate 2. General view of the digital maps of total precipitation of two extreme months (July as the driest and October as the wettest) and one intermediate month (April). For helping the comparison among these maps, the values of the legend range between the absolute minimum and the absolute maximum

In the case of mean maximum air temperature, the extreme months are January ($T = 9.1$°C) and July ($T = 27.9$°C), and in the case of mean minimum air temperature, we also have January ($T = 0.1$°C) as the coldest month, and August ($T = 15.0$°C) as the hottest month.

With regard to the statistical importance of the independent variables, Chuvieco and Salas (1996) found similar results in the case of mean maximum air temperature during summer in the centre of Spain.

It is important to note that the statistical outcomes shown in Tables III and IV, though quite good, are almost always lower than those obtained in the case of mean air temperature. This is to be expected, as extreme values are more difficult to predict than mean values. The exceptions are the months of July and August, where the mean minimum temperature is predicted better than the mean temperature. Indeed, R^2_{nc} range from 0.61 to 0.87 in the case of mean maximum air temperature, and from 0.68 to 0.84 in the case of mean minimum air temperature, while in mean air temperature they ranged from 0.75 to 0.95. Nevertheless, these differences are smaller when correcting the model using the meteorological stations (R^2_{c}), which leads to values ranging from 0.70 to 0.89 for mean maximum air temperature, and from 0.79 to 0.86 for mean minimum air temperature. Mean minimum air temperature is always better predicted than mean maximum air temperature.

As in the case of mean air temperature, in all cases the R^2_{c} using the inverse of squared distance as the interpolation technique for the correctors is always better than kriging (the mean R^2_{c} for all months in the former case is 0.77 for maximums, and 0.83 for minimums, while in the latter case, it is 0.74 and 0.81, respectively), so the real maps with all the stations were built always from inverse weighting distance correctors.

Figures 6 and 7 show that the statistical significance of the variables during the year is similar, in general terms, to the observed in the case of mean air temperature. ALT always has a negative influence on temperature, LAT always has a positive influence, except during winter months, when it is not significant, and RAD is never significant. The CON pattern changes in the case of mean air minimum temperature, where it has a negative influence throughout the year, instead of the winter–summer pattern.

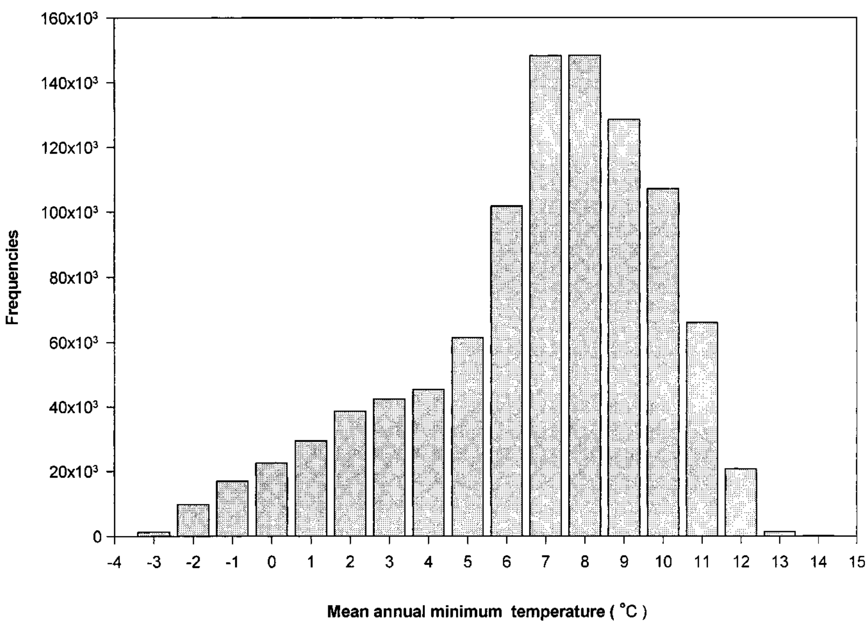

Figure 9. Histogram of the mean annual minimum temperature. The high frequency values are caused by the matrix size (990 375 cells)

Table VII. Results of the multiple regression analysis in the case of total precipitation

Month	Multiple regression coefficients (b)			R^2_{nc}	R^2_c
January	ALT = 0.016	CON = ns	CLO = −317.672	0.44	0.63
	LAT = ns	Interception = 265.28			
February	ALT = 0.011	CON = ns	CLO = −265.794	0.56	0.80*
	LAT = −565.76	Interception = 645.705			
March	ALT = 0.017	CON = ns	CLO = −325.495	0.43	0.60
	LAT = ns	Interception = 277.287			
April	ALT = 0.019	CON = ns	CLO = −219.769	0.66	0.73
	LAT = −1041.762	Interception = 985.266			
May	ALT = 0.035	CON = ns	CLO = −154.296	0.60	0.79*
	LAT = −1087.019	Interception = 978.331			
June	ALT = 0.037	CON = ns	CLO = −131.736	0.70	0.91*
	LAT = −1544.507	Interception = 1292.315			
July	ALT = 0.033	CON = −0.179	CLO = ns	0.75	0.91*
	LAT = −2991.219	Interception = 2262.055			
August	ALT = 0.032	CON = ns	CLO = −287.141	0.69	0.91*
	LAT = −1493.12	Interception = 1364.141			
September	ALT = 0.028	CON = −0.113	CLO = −255.993	0.32	0.64*
	LAT = ns	Interception = 251.808			
October	ALT = 0.025	CON = −0.241	CLO = −398.814	0.46	0.72
	LAT = ns	Interception = 366.988			
November	ALT = 0.027	CON = ns	CLO = −325.397	0.51	0.78*
	LAT = ns	Interception = 282.937			
December	ALT = 0.024	CON = ns	CLO = −310.632	0.46	0.69*
	LAT = ns	Interception = 266.975			
Annual	ALT = 0.302	CON = ns	CLO = −4141.815	0.66	0.86*
	LAT = ns	Interception = 3532.531			

Information as with Table I.
The asterisk (*) means that the corrector maps had been obtained through kriging instead of inverse of the squared distance.

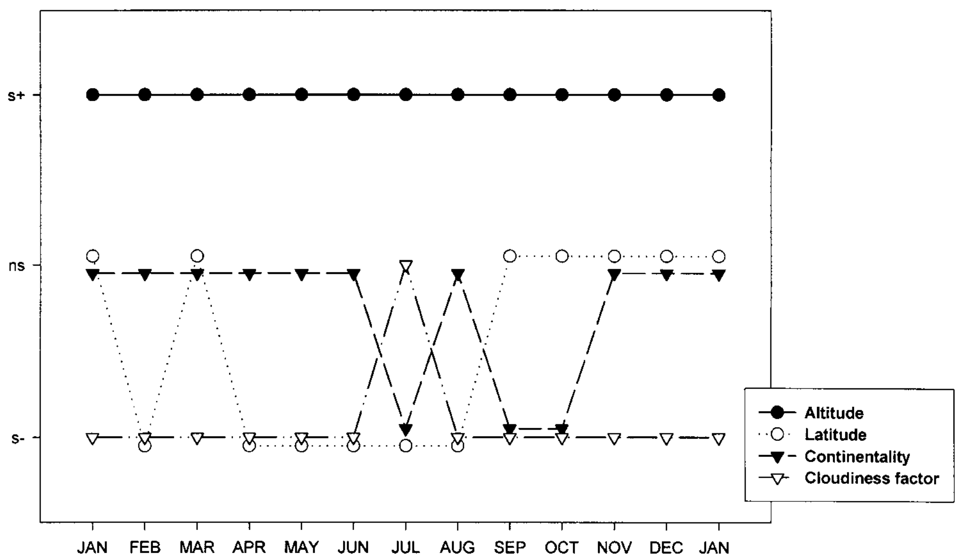

Figure 10. Statistical importance of the independent variables during the year in the case of total precipitation. 's + ' means that the variable is significant ($\alpha = 0.05$), and has a positive contribution, 's − ' means that the variable is significant ($\alpha = 0.05$), and has a negative contribution, and 'ns' means that it is not significant

4.3. Precipitation

Table VII shows that the coefficients of determination are lower than those observed in the case of air temperature, as is to be expected, because it is more difficult to model precipitation than mean air temperature when using only geographical data. The most unpredictable month is September ($R^2_{nc} = 0.32$), and also October, being one of the lowest ($R^2_{nc} = 0.46$). This could be a result of the kind of perturbations that take place in these months on the Mediterranean coast (Clavero *et al.*, 1996). Winter months are very

Table VIII. Statistical parameters of the total precipitation (mm) for the 990 375 cells of the matrix inside Catalonia

Month	Mean	Standard deviation	Minimum value	Maximum value
January	45.5	14.4	16.9	98.8
February*	38.6	14.0	10.8	89.3
March	52.7	14.6	19.1	113.3
April	62.0	18.1	27.6	129.0
May*	78.8	26.1	37.8	177.1
June*	69.1	29.9	24.2	178.5
July*	41.7	28.6	3.3	155.4
August*	62.4	30.3	10.5	176.7
September*	77.1	19.2	31.2	169.6
October	80.2	18.5	35.2	141.1
November*	64.2	20.8	23.4	156.6
December*	56.6	19.2	20.8	140.2
Annual*	727.1	223.3	335.0	1593.7

The asterisk (*) means that the corrector maps had been obtained with kriging instead of the inverse of the squared distance.

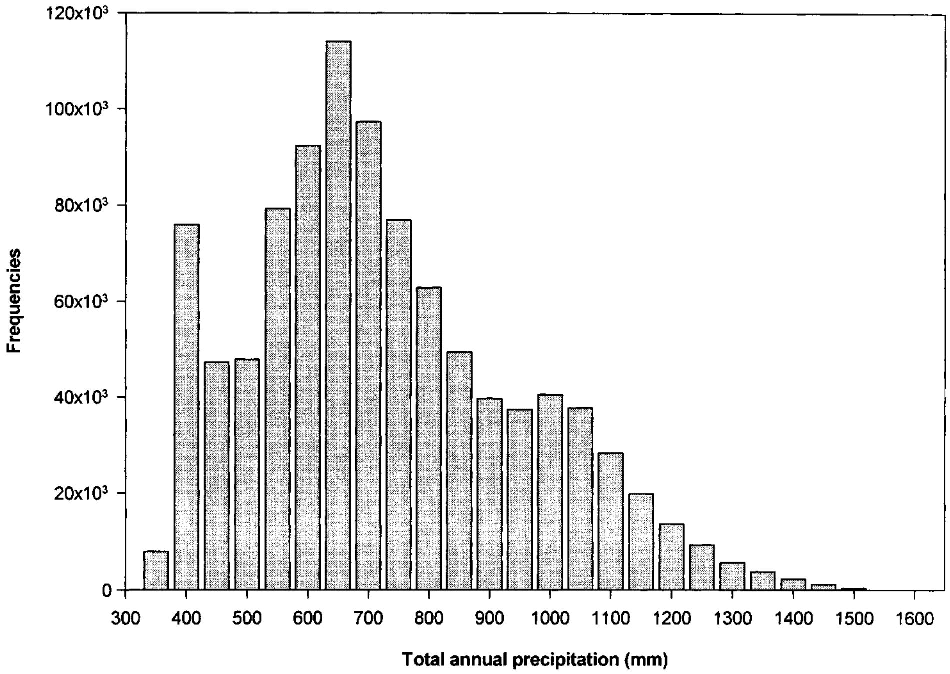

Figure 11. Histogram of the annual precipitation. The high frequency values are caused by the matrix size (990375 cells)

unpredictable too ($R^2_{nc} = 0.44$ in the case of January). The summer months are the ones with better predictions ($R^2_{nc} = 0.75$ in the case of July). Unlike in the air temperature case, we can see that the coefficients of the corrected model (R^2_c) are significantly higher than the coefficients of the non-corrected model, with R^2_c ranging from 0.60 (March) to 0.91 (June, July and August). Therefore, this model is dramatically improved once it is corrected with data from the meteorological stations.

Unlike in the temperature case, applying the kriging interpolation technique to the correctors gives better results than the inverse of squared distance interpolator in more cases (the mean of the R^2_c for all months is 0.76 in the former case and 0.74 in the latter case). Despite applying an interpolation technique to the correctors, the corrected model is always better than the non-corrected model.

Figure 10 shows, as in the air temperature case, the statistical importance of the variables used. In this case, ALT, as well as CLO, are significant during almost the entire year. CON is significant in autumn because the stations located near the sea receive more precipitation during the Mediterranean perturbations. LAT is significant during spring and summer, probably because of the presence of the Pyrenees in the north of the area, where summer rainfall is high. We can see that the model is strongly influenced by the ALT factor, as stated in other works (Bigg, 1991; Hutchinson, 1995). Such data can help to complement or improve the conclusions of other papers, such as work like that of Berndtsson (1989), which discusses the variability between different months and the influence of the topography and distance to the coast.

Plate 2 shows the digital maps of annual precipitation for three representative months, and Table VIII shows the descriptive statistics for this map. The months with lower precipitation are February (38.6 mm) and July (41.7 mm), while the months with higher precipitation are May (78.8 mm) and October (80.2 mm). Finally, Figure 11 shows the histogram of frequencies for annual precipitation.

5. CONCLUSIONS

Highly accurate and objective monthly and annual climatic maps of air temperature and precipitation can be made by following the proposed methodology that integrates statistical and GIS techniques. Even if only geographical data are used, good results are obtained, particularly for air temperature. When using the correctors from meteorological stations, very good results are obtained, with R^2_c ranging from 0.70 to 0.97 for air temperature, and from 0.60 to 0.91 for precipitation.

It is important to emphasize that the final results of this work are digital maps that can be automatically updated with new meteorological data, as well as easily queried in a GIS environment. To apply this methodology to other areas, only a DEM and the data from meteorological stations are needed.

Future trends of this research include introducing new independent variables (the influence of land use over air temperature, Vogt *et al.*, 1997), correcting the model with new factors (influence of the urban heat islands), using new interpolation techniques for building the corrector maps, and trying to improve the current variables, especially the CON factor. Also, through a DEM, we could obtain cold air drainage that can be useful to correct air temperature values. This is especially interesting in the case of mean minimum air temperature, because it has more complex relationships than mean maximum and mean air temperature (Blennow, 1998; Bolstad *et al.*, 1998). Other techniques, such as the support of remote sensing (Menz, 1997), can be useful to refine the knowledge of air temperature.

ACKNOWLEDGEMENTS

We want to acknowledge, especially, the advice of Professor Martín Vide (Barcelona University) on some climatological aspects. Raimon Salvador has helped us in reviewing the statistical methodology. Joan Masó helped us with his computer expertise. Also, we want to thank the Department of Environment of the Generalitat of Catalonia (DMA), Department of Agriculture of the Generalitat of Catalonia (DARP), and National Meteorological Institute (INM) of Spain for supplying the meteorological data to us. Jordi Porteros was of great help in starting this project.

REFERENCES

Ashraf M, Loftis JC, Hubbard KG. 1997. Application of geostatistics to evaluate partial weather station networks. *Agricultural and Forest Meteorology* **84**: 225–271.

Azevedo EB, Santos Pereira L, Itier B. 1998. A physically-based model for local climate simulation in islands environments: a GIS application. In CD-ROM *Proceedings of the GIS PlaNET 98 (I International Conference and Exhibition on Geographic Information).* Produced and developed by Imersiva, Lisbon.

Benzi R, Deidda R, Marrocu M. 1997. Characterization of temperature and precipitation fields over Sardinia with principal component analysis and singular spectrum analysis. *International Journal of Climatology* **17**: 1231–1262.

Berndtsson R. 1989. Topographical and coastal influence on spatial precipitation patterns in Tunisia. *International Journal of Climatology* **9**: 357–369.

Bigg GR. 1991. Kriging and intraregional rainfall variability in England. *International Journal of Climatology* **11**: 663–675.

Blennow K. 1998. Modelling minimum air temperature in partially and clear felled forests. *Agricultural and Forest Meteorology* **91**: 223–235.

Blennow K, Persson P. 1998. Modelling local-scale frost variations using mobile temperature measurements with a GIS. *Agricultural and Forest Meteorology* **89**: 59–71.

Bolòs M, Vilà J, Albentosa LM, Gómez A, Sala M, Bolòs O, Vidal T. 1983. *Gran geografia comarcal de Catalunya. Geografia general*, vol. 17. Fundació Enciclopèdia Catalana: Barcelona.

Bolstad PV, Swift LL, Collins F, Régnière J. 1998. Measured and predicted air temperatures at basin to regional scales in the southern Appalachian mountains. *Agricultural and Forest Meteorology* **91**: 161–176.

Burrough PA, McDonnell RA. 1998. *Principles of Geographical Information Systems*. Oxford University Press: New York; 333.

Clavero P, Martín Vide J, Raso Nadal JM. 1996. Atles climàtic de Catalunya. Termopluviometria, Generalitat de Catalunya (Departament de Politica Territorrial i Obres Públiques), Institut Cartogràfic de Catalunya i Departament de Medi Ambient, Barcelona.

Chessa PA, Delitala AM. 1997. Objective analysis of daily extreme temperatures of Sardinia (Italy) using distance from sea as independent variable. *International Journal of Climatology* **17**: 1467–1485.

Chuvieco E, Salas J. 1996. Mapping the spatial distribution of forest fire danger using GIS. *International Journal of Geographical Information Systems* **10**: 333–345.

Debazac EF. 1983. Temperate broad-leaved evergreen forest of the Mediterranean region and middle east. In *Temperate Broad-Leaved Evergreen Forest: Ecosystem of the World*, vol. 10, Ovington JD (ed.). Elsevier: New York/Amsterdam; 107–122.

Dozier J. 1989. Spectral signature of alpine snow cover from the Landsat thematic mapper. *Remote Sensing of Environment* **28**: 9–22.

Driscoll DM, Yee Fong JM. 1992. Continentality: a basic climatic parameter re-examined. *International Journal of Climatology* **12**: 185–192.

Eastman JR. 1997. *Idrisi. Users' Guide*. Clark University: Worcester; 178.

Emberger L. 1952. *Sur le quotient pluviothermique*. C.R. Académie Science: Paris; 2508–2510.

Gessler PE, Moore ID, McKenzie NJ, Ryan PJ. 1995. Soil-landscape modelling and spatial prediction of soil attributes. *International Journal of Geographical Information Systems* **9**: 421–432.

Hamilton JE, Lennon P, O'Donnell B. 1988. Objective analysis of monthly climatological fields of temperature, sunshine, rainfall percentage and rainfall amount. *Journal of Climatology* **8**: 109–124.

Hargy VT. 1997. Objectively mapping accumulated temperature for Ireland. *International Journal of Climatology* **17**: 909–927.

Hubbard KG. 1994. Spatial variability of daily weather variables in the high plains of the USA. *Agricultural and Forest Meteorology* **68**: 29–41.

Hutchinson MF. 1995. Interpolating mean rainfall using thin plate smoothing splines. *International Journal of Geographical Information Systems* **9**: 385–403.

Lanzante JR. 1996. Resistant, robust and non-parametric techniques for the analysis of climate data: theory and examples, including applications to historical radiosonde station data. *International Journal of Climatology* **16**: 1197–1226.

Llasat C. 1997. Meteorologia agrícola i forestal a Catalunya, Secció d'estadística i informació agrària (Gabinet tècnic), Departament d'Agricultura, Ramaderia i Pesca, Generalitat de Catalunya; 298.

Martín Vide J. 1987. Característiques climatològiques de la precipitació en la franja costera mediterrània de la Península Ibèrica, Tesi Doctoral. Institut Cartogràfic de Catalunya, Generalitat de Catalunya, Barcelona; 245.

Menz G. 1997. Regionalization of precipitation models in east Africa using meteosat data. *International Journal of Climatology* **17**: 1011–1027.

Narumalani S, Jensen JR, Althausen JD, Burkhalter S, Mackey HE. 1997. Aquatic macrophytic modeling using GIS and logistic multiple regression. *Photogrammetric Engineering and Remote Sensing* **63**: 41–49.

Phillips DL, Dolph J, Marks D. 1992. A comparison of geostatistical procedures for spatial analysis of precipitation in mountainous terrain. *Agricultural and Forest Meteorology* **58**: 119–141.

Piñol J, Lledó MJ, Escarré A. 1991. Hydrological balance of two Mediterranean forested catchments (Prades, northeast Spain). *Journal des Sciences Hydrologiques* **36**(2): 95–107.

Pons X. 1996. Estimación de la Radiación Solar a partir de modelos digitales de elevaciones. Propuesta metodológica. In *VII Coloquio de Geografía Cuantitativa, Sistemas de Información Geográfica y Teledetección*, Juaristi J, Moro I (eds). Vitoria-Gasteiz; 87–97.

Pons X. 1998. *Manual of MiraMon. Geographic Information System and Remote Sensing Software* (http://www.creaf.uab.es/miramon). Centre de Recerca Ecològica i Aplicacions Forestals (CREAF): Bellaterra; 150.

Rothermel R, Wilson RA, Morris GA, Sackett SS. 1986. Modelling moisture content of fine dead wildland fuels. Input to the BEHAVE fire prediction system, USDA Forest Service, Intermountain Research Station. Research Paper INT-359, Ogden.

Sánchez Palomares O, Sánchez Serrano F, Carretero Carrero M^aP. 1999. *Modelos y cartografía de estimaciones climáticas termopluviométricas para la España peninsular*, Ministerio de Agricultura, Pesca y Alimentación. Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria: Madrid; 192.

Tomlin CD. 1990. *Geographic Information Systems and Cartographic Modelling*. Prentice Hall: Englewood Cliffs, NJ; 249.

Vogt JV, Viau AA, Paquet F. 1997. Mapping regional air temperature fields using satellite-derived surface skin temperatures. *International Journal of Climatology* **17**: 1559–1579.

World Meteorological Organization. 1967. A note on climatological normals. Technical note 84, WMO-208.

