Comparison of regression techniques for mapping fog frequency: application to the Aragón region (Northeast Spain)

S. M. Vicente-Serrano(1), J. I. López-Moreno(1), M.I. Vega-Rodríguez(1), S. Beguería(2)
and José M. Cuadrat(3)

1) Instituto Pirenaico de Ecología, CSIC (Spanish Research Council), Campus de Aula Dei, P.O. Box 202,
Zaragoza 50080, Spain

2) Estación Experimental de Aula Dei CSIC (Spanish Research Council), Zaragoza, Spain.

3) Departamento de Geografía. Universidad de Zaragoza, Zaragoza, Spain.
contact: svicen@ipe.csic.es

Abstract. We compare different spatial interpolation techniques in mapping the monthly frequency of fogs in the Aragón region (Northeast Spain). The local and spatially complex nature of the fogs makes them more difficult to map than other climatic variables such as precipitation and temperature. We found clear seasonal differences in the quality of the obtained maps. The localized nature of spring and summer fogs, mainly restricted to valley bottoms in mountainous areas, gives rise to several limitations. The modeling of fog frequency is more complex than that of other climate variables; to improve the model predictions, it is necessary to consider topographic variables that simulate the terrain structure. Moreover, the highly complex nature of the relationship between fog frequency and terrain means that simple linear models perform poorly in accounting for the role of geographic and topographic variables in determining the spatial distribution of fog frequency. The inclusion of non-linear relationships between fog frequency and terrain variables in the models following a General Additive Model procedure leads to an improvement in model performance because the flexibility of GAMs enables the inclusion of non-linear relationships and the generation of response-curve shapes that detail the exact relationship between the dependent variable and predictors throughout the entire range of the variable.

Key-words: Fog frequency, fog mapping, interpolation, ordinary least squares regression, splines, general additive models, non-linearity, Aragón

1. Introduction

Reliable regional estimates of climate parameters are crucial for territorial management; thus, several techniques have been developed to convert temporally discrete climatic data to continuous fields and obtain reliable maps of climate variables. Many previous studies have mapped climate parameters such as precipitation (Ninyerola *et al.*, 2000, 2007a; Brown and Comrie, 2000), temperature (Vicente-Serrano *et al.*, 2003; Ninyerola *et al.*, 2007b), evapotranspiration (Vicente-Serrano *et al.*, 2007; Martínez-Cob, 1996), and snow depth (López-Moreno and Nogués-Bravo, 2005). The climate variables that have received the

greatest attention are precipitation and temperature, as maps of these variables are of great interest for numerous applications. In comparison, other climate variables have received much less attention, as some are difficult to model spatially. For example, wind direction and intensity are difficult to model because of the highly variable nature of wind flows and their interaction with relief at fine spatial scales (Jiménez *et al.*, 2007).

Among the less intensively analyzed and modeled climate variables is fog, despite the fact that its occurrence has ecological, economic, and social implications. It is well known that fog and low cloud affect marine transport (Talley, 1999) and aviation, as a reduction in visibility affects both the safety and efficiency of air-traffic management (Allan *et al.*, 2001; Terradellas and Cano, 2007). Road traffic is also hampered by fog, leading to an increase in the risk of accidents and limiting driving speeds (Kolisetty *et al.*, 2006). Fog is also important for the quality of life and health of citizens because of its association with respiratory problems (Varughese *et al.*, 2005; Bell and Davis, 2001) and harboring of harmful particles aerosols and trace elements (Rattigan *et al.*, 2002). Fog is also important from an ecological perspective: in arid and semi-arid regions, fog is an important source of humidity, enabling the growth of vegetation (Braun-Blanquet and Bolos, 1957; Caceres *et al.*, 2007). Finally, fog is an important source of water in arid and semiarid regions, with several technical measures having been developed to harvest water from fog for use in irrigation and domestic supply (Caceres *et al.*, 2004; Abdul-Wahab, 2007).

Despite the high levels of interest in obtaining the spatial distribution and frequency of fog, there have been no attempts to develop reliable spatial models of fog frequency or obtain accurate maps of its spatial distribution within a given region.

Fogs are usually highly localized in their distribution, which is largely determined by relief and its interaction with air masses. This leads to complex relationships between topographic and geographic characteristics (e.g., the presence of water bodies, rivers, depressions) and the

occurrence of fog, resulting in turn in the highly localized behavior of fog banks, complicating any attempt to map their distribution.

Given the important role of topography in determining fog occurrence, the mapping of fog must be based on methods that combine available fog data obtained from meteorological stations with geographic and topographic information. In this study, we compare different spatial interpolation techniques in mapping the monthly frequency of fogs in the Aragón region (Northeast Spain). For this purpose, we employ various regression-based methods assisted by geographic information systems that enable the inclusion of auxiliary geographic and topographic data in creating prediction models.

2. Methods

2.1. Data

We used fog-frequency data recorded at 450 meteorological observatories in the Aragón region over the period 1950–2002. In this dataset, the designation of a day with fog is determined subjectively by observers. The National Institute of Meteorology of Spain employs the following categories and criteria based on visibility: i) mist (visibility between 500 and 1000 m), ii) moderate fog (visibility between 50 and 500 m), and iii) dense fog (visibility less than 50 m). In our analysis, we did not make this distinction between different types of fog on the basis of visibility; the data were summarized within monthly intervals, counting the monthly frequency of fogs in days. The analyzed data are fragmentary and incomplete, commonly comprising very short series of less than 5 years duration. Any long time-series of fog frequency would therefore be unreliable, and the maps were compiled using data for a common period (1980–2002), for which a minimum of 10 years of data was required to calculate monthly averages. Data were used from a total of 258 observatories (Figure 1). A random sample of 80% of the observatories was used to make interpolations; the

remaining 20% were reserved for validation. The data set has a very good spatial coverage, including data in a wide range of elevations (Figure 2).

2.2. Interpolation procedures

Three regression-based methods were tested in mapping fog frequency: ordinary least squares regression (OLSR), generalized linear models (GLMs), and general additive models (GAMs).

2.2.1. Ordinary least squares regression

In contrast with geostatistical and local methods, regression procedures use all of the available weather stations in compiling climatic maps. These methods are based on the creation of dependence models between climatic data and other potential predictors in predicting the values of climatic variables in unsampled regions. The geographic coordinates of the weather stations (latitude and longitude) and geographic (e.g., distance to water bodies) and/or topographic variables (e.g., elevation, aspect, slope) are used for this purpose. The value of a climatic variable at an unsampled point is calculated using the following transference function:

$$z(x) = b_0 + b_1P_1 + b_2P_2 + \dots + b_nP_n ,$$

where z is the predicted value at point x , $b_0 \dots b_n$ are the regression coefficients, and $P_1 \dots P_n$ are the values of the different predictors at point x . The inclusion of these variables in a geographic information system (GIS) facilitates their calculation and helps to reveal their interrelation with climatic data. The most commonly used variables include those that affect the spatial distribution of climate, including elevation, longitude, latitude, distance to the ocean, and areas of topographic shadow (Ninyerola *et al.*, 2000, 2007a; Vicente-Serrano *et al.*, 2003, 2007).

These variables are usually sufficient in mapping climate variables that show gradual spatial variations; however, for variables that show very large spatial variability and notable local

features, such as the frequency of fogs, other spatially varying predictors that record local geographic and topographic characteristics are potentially useful in modeling. Such variables are commonly generated from digital elevation models (DEM). The potential predictors used in the present analysis are listed in Table 1. For elevation, available from a DEM at a resolution of 1000 m, which is the variable that most strongly influences the spatial distribution of climate parameters (Daly *et al.*, 1994, 2002), we performed transformations using low-pass filters with radii of 2.5, 5, 10, and 25 km to measure the wider influence of elevation on fog frequency. We also calculated the maximum elevation, minimum elevation, and range in elevation using the same filters and radii to take into account the possible effect of topographic gradients, which reflect the terrain structure (valleys, depressions, etc.).

To define the terrain aspect, which is important in modeling the contrasting climatic conditions of northern and southern slopes and in taking into account the effect of slope exposure to wind flow on the spatial distribution of climate variables, we used a theoretical model of incoming solar radiation following Ninyerola *et al.* (2000) (northern and southern slopes have low and high values of incoming solar radiation, respectively). Digital coverage of annual mean incoming solar radiation was created using the Mira-Mon GIS and the DEM of Pons (1997). The slope angle of the terrain was included as a potential predictor to take into account its possible role in the accumulation of cold air in valley bottoms, as this usually leads to a higher frequency of fogs upon shallow slopes than upon steep slopes. Low-pass filters with radii of 2.5, 5, and 10 km were also applied to these variables.

The distances to main rivers were also included, as they represent a source of humidity that can favor the presence of fogs. Finally, because fog frequency is strongly favored by the presence of valley bottoms and depressions, we used a method that models these topographic characteristics at different spatial scales. For this purpose, the topographic wetness index (TWI, $\ln(\alpha/\tan\beta)$), which combines the local upslope contributing area (α) and slope (β), has

generally been used to model those areas favored to retain soil moisture due to their topographic characteristics (commonly valley bottoms and the centers of depressions) (Sorensen *et al.*, 2006); such areas are also usually likely to record a high frequency of fogs. The index was calculated using ArcGis software and the available DEM. Low-pass filters with radii of 2.5, 5, and 10 km were also applied to these variables.

Figure 3 shows the spatial distribution of elevation and the TWI filtered with radii of 2.5 and 10 km. It is clear that valley bottoms and depressions show the highest TWI values. Following Brown and Comrie (2002), we calculated the cross-products of elevation, slope, and TWI, as these variables may cover spatial variations associated with their interactions.

The spatial database was sampled at those locations corresponding to the climatic stations. A regression analysis of this dataset was then performed for each month of the year using the software SPSS (Statistical Package for Social Sciences) version 12. The normality of each variable was tested using the Kolmogorov–Smirnov test. In those cases for which normality was rejected at a statistical significance of $p > 0.05$, the variables were transformed using natural logarithms. Given the likely occurrence of correlations between several of the potential predictors, a forward stepwise procedure was selected in performing the regression. A conservative value of 0.01 was set for a variable to enter the stepwise procedure, as recommended by Hair *et al.* (1998); however, because colinearity problems can arise even with correlations as low as 0.3, a condition index test (SPSS software, version 12) was applied to the regression models. This test flags those models for which the uncertainty coefficient is too large because of multicollinearity, and recommends the use of a model with fewer parameters.

2.2.2. Generalized linear models

The generalized linear model is a generalization of ordinary least squares regression. Each terrain variable is transformed according to the best non-linear relationship with the

dependent variable (frequency of fog days) and the transformed variable included in the ordinary least squares regression. This enables an improvement in the quality of the models because non-linear relationships are also taken into account. In the present study, only simple non-linear (potential, exponential, etc.) relationships were taken into account for transformations. The best adjustment following non-linear relationships was calculated by means of bivariate relationships using the software TableCurve 2D©; this was only performed if the difference between the linear relationship between the dependent variable and each of the potential predictors and the non-linear relationship between the dependent variable and each of the potential predictors was greater than $R^2 > 0.05$. In such a case, the non-linear relationship was preferred and the potential predictor was transformed according to the obtained equation. As an example, Figure 4 shows the relationship between the fog frequency in January and the RANG₅. The relationship between the variables is better quantified following a non-linear exponential relationship rather than a linear model. Thus the non-linear model explains more than 5% more of the spatial variance than the linear model.

2.2.3. Generalized additive models

GAMs are non-parametric extensions of GLMs that estimate response curves with a non-parametric smoothing function rather than parametric terms (Hastie and Tibshirani, 1987). This approach enables the user to explore shapes of predictor responses to the gradient in the dependent variable (fog frequency in this case), enabling in turn the accurate fit of statistical models to highly non-linear relationships and the detection of abrupt changes in the responses of many natural processes (Lehmann *et al.*, 2002). Thus, a GAM can be stated as

$$g(E(Y)) = PL = \alpha + f_1(X_1) + f_2(X_2) + \dots + \varepsilon$$

where each predictor variable, X_n , is fitted using a smoothing function $f_n(X_n)$. Consequently, a GAM involves the addition of different functions fitted to the potential predictors in order to predict Y values. Data are fitted with respect to the partial residuals, which are the residuals

remaining after removing the effects of all predictor variables. A detailed description of how GAMs are fitted to the data in relation to the employed algorithms can be found in Hastie and Tibshirani (1987).

Predictors were selected using a newly developed method for GAMs (Lehmann *et al.*, 2005) called “cross-selection”, as implemented in the software GRASP (Lehmann *et al.*, 2002; Maggini *et al.*, 2006). Using this method, the statistic used to include a predictive variable is tested at each step, validating the predictions on five independent subsets.

As an example of this approach, Figure 5 shows the response curves of the three predictor variables selected by the GAM in predicting the mean frequency of fog during January. According to the obtained model, fog frequency decreases with $ELEV_{10}$, especially from the lowest altitude to 1000 m a.s.l.; thereafter, the rate of decrease in frequency is much more moderate. Longitude is also considered a significant predictor, assigning higher frequencies to the lowest and highest values. Finally, the third predictor, $RANG_5$, also shows a marked non-linearity in its response curve, with an essentially null influence in the model up to 1000 m (response curve constant around 0) and a marked decrease in the expected frequency beyond this threshold. However, the low number of cases for which $ELEV_{100} > 1500$ and $RANG_5 > 1000$ produces a noticeable increase in the confidence intervals, thereby making it necessary to consider with caution the predictions over terrain that possess these characteristics.

2.2.4. Local corrections

Because of the global nature of the regression-based techniques, some local features remain unrepresented in the model. Global methods are inexact interpolators, as the predicted value $z(x)$ does not coincide with the real climatic data recorded at the weather stations: there exists a known error in the final prediction (residual). Previous studies used the following correction method (interpolation of residuals) to obtain exact climatic data from weather

stations (e.g., Ninyerola *et al.*, 2000, 2007; Agnew and Palutikof, 2000; Brown and Comrie, 2002; Vicente-Serrano *et al.*, 2007):

$$\text{residual} = \text{observed data} - \text{predicted data} .$$

A local adaptive method, *splines* with tension ($\varphi = 400$), was selected for the interpolation of the residuals obtained between the monthly observed and predicted data for each month and the three regression procedures (Mitasova and Mitas, 1993). We used a relatively high value of the tension parameter (400) in order to obtain interpolated residuals close to the values observed at the stations. This approach was employed because it has proved to be more suitable than other local techniques such as inverse-distance weighting or geostatistics (Ninyerola *et al.*, 2007a; Vicente-Serrano *et al.*, 2003; Nalder and Wein, 1998). The sum of residual interpolation and predicted regression maps modified the initial results of the model, and real values were obtained at the locations of the weather stations:

$$\text{observed data} = \text{predicted data} + \text{residual interpolation}.$$

2.3. Validation

To validate the maps, we used the data from the 20% of randomly selected weather stations reserved for subsequent tests. We calculated the mean absolute error (MAE), which is a measure of the average error of the interpolation. We also used the agreement index (Willmott's D), a relative and bounded measure, to assess the map quality because it retains mean information and does not amplify outliers (Willmott, 1982). The index also scales with the magnitude of variables, thereby enabling comparisons of the different monthly maps of fog frequency independently of differences in the magnitude and range of the variable for each month.

Willmott's D is calculated according to the following equation:

$$D = 1 - \frac{\sum_{i=1}^N (P_i - O_i)^2}{\sum_{i=1}^N (|P_i'| + |O_i'|)^2},$$

where N is the number of observations, O is the observed value, P is the predicted value, $P_i' = P_i - \bar{O}$, and $O_i' = O_i - \bar{O}$, where $\bar{O}$ is the mean of observed values. Values close to 1 represent a good agreement between the observed and predicted data.

3. Results

Table 2 lists the R^2 values obtained from the monthly models following the OLSR, GLM, and GAM methods before splines adjustment. The models generally explain a low percentage of the total variance, especially for the spring and summer months. The methods that consider non-linear relationships between predictors and dependent variables (GLMs and GAMs) provide better results than OLSR; nevertheless, even the highest R^2 values obtained for the winter months do not exceed 60–70% of the explained variance (the GAM method). This finding indicates the important contribution of local components in the spatial distribution of fog frequency and the need to consider these effects via the local interpolation of the residuals obtained between observed and modeled data.

Table 3 shows the predictors included in each model. In general, the number of variables included in GAMs is much lower than those in OLSR and GLM. The number of variables included in the models is higher in the spring and summer months; however, this does not lead to an improvement in the quality of the model predictions.

Elevation, TWI, and the cross-product between these variables were selected in most of the models obtained by OSLR and the GLMs. In the GAMs, the variable included most frequently was the elevation range within a radius of 5 km. The correlation of this variable with TWI_5 is $R = -0.83$ ($p < 0.01$); therefore, the two variables are comparable in that they both record the topographic characteristics favorable to a high frequency of fogs, such as the valley

bottoms and depressions. The reliability of the obtained maps varies markedly as a function of the employed method and month (Table 4). Regardless of the selected method and residual correction, the best fit was found for winter months. The models generally showed acceptable performance in the winter months, whereas in summer the Willmott's D values are very low, indicating the poor quality of the obtained maps. Considering residual correction and independent of the method selected, in most cases the MAE values decrease after residual correction, especially for winter months. The D values also increase following this correction, confirming the need for residual correction to take into account the local behavior of fogs that cannot be represented by the models.

The three tested regression-based procedures yield similar average monthly D values (0.56, 0.58, and 0.61 for OLSR, GLM, and GAM, respectively); nevertheless, if the winter months are analyzed independently, GLM and GAM clearly provide better results than OLSR, thereby providing evidence that complex relationships between fog frequency and topographic/geographic variables are better recorded using non-linear procedures. Few differences exist between GAM and GLM methods. Thus, there are not statistically significant differences between the distribution of errors (observed-predicted) in the test stations between the GLM and GAM methods, according to a one-way ANOVA with Tamhane post-hoc test ($p = 0.93$). Nevertheless, GAM usually provides better D values and lower MAE values than GLMs. These results, together with the fact that GAM is more automatic and time-efficient, made GAM the preferred choice in mapping fog frequency in the analyzed region.

Figure 6 shows the maps of average wintertime fog frequency obtained by OLSR, GLMs, and GAMs. Although the spatial patterns are similar, the OLSR map shows an excessive smoothing of the spatial distribution of fogs. The maps obtained using GLMs and GAMs

adapt better to the topography and the distribution of valley bottoms and the main topographic depressions of Aragón, showing a more realistic spatial distribution of fogs within the region. Figure 7 shows the monthly maps of fog frequency obtained for each month using GAMs and after locally correcting the models via the interpolation of residuals. The spatial distribution of fog frequency shows noticeable spatial differences; notable seasonal contrasts are also important. The highest frequency of fogs is recorded in the winter months (November–February), mainly in the central and eastern parts of the study area. The main reason of this seasonal behaviour is that low elevations and depressed topography favours the accumulation of cold air and the occurrence of fog during the passing of frequent wintertime anticyclones. During the spring and summer months, the spatial patterns of fog frequency are highly localized, with the highest frequencies being recorded in the northernmost elevated areas where the relief favors the development of thermal inversions, mainly in the valley bottoms. In some mountainous locations, fog occurrence may also be a consequence of cloudiness retained over slopes exposed to the wind. This behavior causes excessively patchy maps and makes it difficult to map fog frequency, thereby explaining the low R^2 values obtained for the models during these months. In contrast, wintertime maps provide an accurate account of the observed fog frequency in the region.

4. Conclusions

In this paper, we compared different regression-based techniques in mapping the frequency of fogs in a region with complex topography. To the best of our knowledge, this is the first approach in the scientific literature to objectively map this climatic variable. We demonstrated that the local and spatially complex nature of fogs makes them more difficult to map than other climatic variables such as precipitation and temperature, for which it is common to obtain high R^2 values for regression models, usually higher than 0.7 (Ninyerola *et al.*, 2007a,

b). R^2 values higher than 0.7 have been obtained previously for the Aragón region in monthly regression models of precipitation and temperature (Vicente-Serrano *et al.*, 2003; Vicente-Serrano *et al.*, 2007; López-Moreno *et al.*, 2007). In contrast, fog-frequency models for this region yield R^2 values between 0.48 and 0.72 for the winter months, and only consider non-linear relationships by means of GAMs.

The high degree of seasonality of fog frequency makes it difficult to obtain reliable maps for most of the months of the year. In this study, we identified clear seasonal differences in model performance. The localized nature of spring and summer fogs, being mainly restricted to valley bottoms within mountainous areas, introduces several limitations to our analyses. The low R^2 values obtained by the linear and non-linear models during these months (0.12 to 0.52) indicate that the resulting maps are largely unreliable. During winter, fog has a wider spatial distribution, enabling the compilation of more reliable fog-frequency maps.

In terms of the usefulness of the obtained maps, the fact that largely non-reliable maps were obtained for spring and summer is not a serious problem because the most widespread fogs are recorded in winter, when fogs have the highest spatial density and extent, longest duration, and greatest affect on road and air traffic and the health of citizens. In contrast, springtime and summertime fogs are highly localized, of low spatial density, and of relatively short duration, only appearing in the night and early morning and having a low economic and social impact.

In terms of the obtained models, among the terrain variables considered for modeling, the topographic variables that simulate the terrain structure in the form of valleys, depressions, crests, etc. (e.g., TWI and elevation range) are included in the majority of the simple linear and complex non-linear models. Accordingly, the modeling of fog frequency is a more complex task than that of modeling other climate variables. Precipitation and mean

temperature are usually well modeled in terms of elevation, aspect, influence of water bodies, and latitude (Vicente-Serrano *et al.*, 2003).

Other variables show more complex patterns, with a greater dependence on local conditions. This has been observed in the mapping of minimum temperature, for which the common occurrence of thermal inversions makes it difficult to obtain reliable estimates (Daly *et al.*, 2002), and the mapping of snow depth, which is strongly affected by topographic features in mountainous areas (López-Moreno and Nogués-Bravo, 2006). Fogs are strongly influenced by thermal inversions, which are a common occurrence in the center of the study area; this explains the high frequency of fogs in the winter months in association with topographic features. Previous studies sought to take thermal inversions into account. For example, Daly *et al.* (2002) divided a set of observatories into two sub-samples (those above and below a certain elevation) and modeled them independently to take into account the effect of thermal inversions in creating minimum-temperature models. Nevertheless, the inclusion of terrain variables such as TWI and the elevation range within a given radius enabled us to consider those topographic characteristics that favor the occurrence of fogs and facilitated the creation of the models via a unique step.

The highly complex nature of the relationship between fog frequency and terrain means that simple linear models are insufficient in terms of accounting for the influence of geographic and topographic variables on the spatial distribution of fog frequency. The same finding has been obtained in mapping the spatial distribution of snow depth in North Aragón (López-Moreno and Nogués-Bravo, 2006). Although ordinary least squares regression commonly provides excellent results when mapping “ordinary” climate variables such as precipitation and temperature, such linear models are largely inflexible in terms of considering the complex relationships between climate and topography when the variable behavior is spatially complex and when local terrain features have a greater effect on the spatial distribution. Variables such

as snow depth and fog frequency are strongly influenced by local features. Model performance is improved if non-linear relationships between fog frequency and terrain variables are included in the models (following a GLM procedure), although the attained increase in the explained variance is not high (about 5% more than linear models). Nevertheless, greater flexibility is achieved with GAMs, leading to an increase (of around 20% more than linear models) in the amount of variance explained by the models and an improvement in the final quality of the maps. The flexibility of GAMs enables the user to include non-linear relationships and obtain response-curve shapes that detail the exact relationship between the dependent variable and predictors throughout the entire range of the variable.

Nevertheless, although non-linear models improve the fog-frequency maps compiled using linear models, they are unable to capture all of the local diversity. For this reason, local corrections via residual interpolation are required to improve the maps. Improved performance is obtained by the combination of advanced non-linear regression procedures and local interpolation to obtain reliable spatial estimations of highly complex climate variables such as fog frequency.

Acknowledgements

We would like to thank the National Institute of Meteorology (Spain) for providing the precipitation data base used in this study. This work has been supported by the following research projects: CGL2005-04508/BOS (financed by the Spanish Commission of Science and Technology and FEDER), PIP176/2005 (financed by the Aragón Government), and “Programa de grupos de investigación consolidados” (BOA 48 of 20-04-2005), also financed by the Aragón Government.

References

- Abdul-Wahab, S.A., Al-Hinai, H., Al-Najar, K.A. and Al-Kalbani, M.S. (2007): Fog water harvesting: Quality of fog water collected for domestic and agricultural use. *Environmental Engineering Science* 24: 446-456.
- Agnew, M.D. and Palutikof, J.P. (2000): GIS-based construction of base line climatologies for the Mediterranean using terrain variables. *Climate Research* 14:115–127
- Allan, S.S., Gaddy, S.G., Evans, J.E. (2001): *Delay and Casualty Reduction at the New York City Airports Using Terminal Weather Information Systems*, Lincoln Laboratory Project Report ATC-291, Massachusetts Institute of Technology, Lexington, MA, 59 p.
- Bell, M.L. and Davis, D.L. (2001): Reassessment of the lethal London fog of 1952: Novel indicators of acute and chronic consequences of acute exposure to air pollution. *Environmental Health Perspectives* 109: 389-394.
- Borrough, P.A. and McDonnell, R.A. (1998): Principles of geographical information systems. Oxford University Press, Oxford.
- Braun-Blanquet, J. and Bolós, O. (1957): Les groupements vegetaux du bassin de l'Ebre. *Anales de la Estación Experimental de Aula Dei*, 5, nº 1-4.
- Brown, D.P. and Comrie, A.C. (2002): Spatial modeling of winter temperature and precipitation in Arizona and New Mexico, USA. *Climate Research* 22: 115–128.
- Caceres, L., Delatorre, J., Gomez-Silva, B., Rodriguez, V. and McKay, C.P. (2004): Atmospheric moisture collection from a continuous air flow through a refrigerated coil tube. *Atmospheric Research* 71: 127-137.
- Caceres, L., Gomez-Silva, B., Garro, X., Rodriguez, V., Monardes, V. and McKay, C.P. (2007): Relative humidity patterns and fog water precipitation in the Atacama Desert and biological implications. *Journal of Geophysical Research-Biogeoscience* 112. Art. No. G04S14.
- Daly, C., Neilson, R.P., Phillips, D.L. (1994): A statistical-topographical model for mapping climatological precipitation over mountainous terrain. *Journal of Climate and Applied Meteorology* 33:140–158.
- Daly, C., Gibson, W.P., Taylor, G.H., Johnson, G.L. and Pasteris, P. (2002): A knowledge-based approach to the statistical mapping of climate. *Climate Research* 22: 99–113.
- Hair, J. F., Anderson, R. E., Tatham, R. L. and Black, W. C. (1998): *Multivariate Data Analysis*. Prentice Hall International, 799 pp.
- Hastie, T. and Tibshirani, R. (1987): Generalised additive model: some applications. *Journal of the American Statisticians Association* 82: 371-386.
- Jiménez, P. A., González-Rouco, J. F., Montávez, J.P., Navarro, J., García-Bustamante, E. and Valero, F. (2007): Wind surface regionalization in a complex terrain region. *Journal of Applied Meteorology and Climatology*. In press.
- Kolisetty, V.G.B., Iryo, T., Asakura, Y. and Kuroda, K. (2006): Effect of variable message signs on driver speed behavior on a section of expressway under adverse fog conditions - A driving simulator approach. *Journal of Advanced Transportation*, 40: 47-74.
- Lehmann, A., McOverton, J. and Leathwick, J.R. (2002): GRASP: generalized regression analysis and spatial prediction. *Ecological Modelling* 157: 189-207.
- Lehmann, A., Leathwick, J.R. and Overton, J.McC. (2005): *Grasp V. 3.2 User's Manual*. Swiss Center For Faunal Cartography, Neuchâtel, Switzerland. <http://www.cscf.ch/grasp>
- López-Moreno, J.I. and Nogués-Bravo, D. (2005): A generalized additive model for modelling the spatial distribution of snowpack in the Spanish Pyrenees. *Hydrological Processes* 19: 3167-3176.

- López Moreno J.I. and Nogués Bravo D. (2006): Interpolating snow depth data: a comparison of methods. *Hydrological processes* 20(10): 2217-232.
- Lopez-Moreno, J.I., Vicente-Serrano, S.M. and Lanjeri, S. (2007): Mapping snowpack distribution over large areas using GIS and interpolation techniques. *Climate Research* 33: 257-270.
- Maggini, R., Lehmann, A., Zimmermann, N.E. and Guisan, A. (2006): Improving generalized regression analysis for the spatial prediction of forest communities. *Journal of Biogeography* 33: 1729-1749.
- Martínez-Cob, A. (1996): Multivariate geostatistical analysis of evapotranspiration and precipitation in mountainous terrain. *Journal of Hydrology* 174: 19–35.
- Mitasova, H., and Mitas, L. (1993): Interpolation by regularized spline with tension. *Mathematical Geology*, 25, 641–655.
- Nalder, I.A. and Wein, R.W. (1998): Spatial interpolation of climatic Normals: test of a new method in the Canadian boreal forest. *Agricultural and Forest Meteorology*, 92: 211-225.
- Ninyerola, M., Pons, X. and Roure, J.M. (2000): A methodological approach of climatological modelling of air temperature and precipitation through GIS techniques *International Journal of Climatology* 20: 1823-1841.
- Ninyerola, M., Pons, X. and Roure, J.M. (2007a): Monthly precipitation mapping of the Iberian Peninsula using spatial interpolation tools implemented in a Geographic Information System. *Theoretical and Applied Climatology* 89: 195-209.
- Ninyerola, M., Pons, X. and Roure, J.M. (2007b): Objective air temperature mapping for the Iberian Peninsula using spatial interpolation and GIS. *International Journal of Climatology* 27: 1231-1242.
- Pons X (1997) Estimación de la radiación solar a partir de modelos digitales de elevaciones: propuesta metodológica. In: Juaristi J, Moro I (eds) *Modelos y sistemas de información en geográfica*. Asociación de Geógrafos Españoles, Vitoria, p 87–97.
- Rattigan, O.V., Mirza, M.I., Ghaury, B.M., Khan, A.R., Swami, K., Yang, K. and Husain, L. (2002): Aerosol sulfate and trace elements in urban fog. *Energy and Fuels* 16: 640-646.
- Sorensen, R., Zinko, U. and Seibert, J. (2006): On the calculation of the topographic wetness index: evaluation of different methods based on field observations. *Hydrology and Earth System Sciences* 10: 101-112 2006
- Talley, W.K. (1999): The safety of sea transport: Determinants of crew injuries. *Applied Economics* 31, 1365–1372.
- Terradellas, E. and Cano, D. (2007): Implementation of a Single-Column Model for Fog and Low Cloud Forecasting at Central-Spanish Airports. *Pure and Applied Geophysics*. 164: 1327–1345.
- Varughese, S., Teschke, K., Brauer, M., Chow, Y., van Netten, C. and Kennedy, S.M. (2005): Effects of theatrical smokes and fogs on respiratory health in the entertainment industry. *American Journal of Industrial Medicine* 47: 411-418.
- Vicente-Serrano, S.M., Saz, M.A. and Cuadrat, J.M., (2003), Comparative analysis of interpolation methods in the middle Ebro valley (Spain): application to annual precipitation and temperature. *Climate Research* 24: 161-180.
- Vicente-Serrano, S.M., Lanjeri, S. and López-Moreno, J.I. (2007): Comparison of different procedures to map reference evapotranspiration using geographical information systems and regression-based techniques. *International Journal of Climatology*, 27: 1103-1118.
- Willmott CJ (1982) Some comments on the evaluation of model performance. *Bulletin of the American Meteorological Society* 63:1309–1313.

Figure 1. Map of the study area showing the distribution of fog-frequency observatories used in this study.

Figure 2. Altitudinal distribution of the fog frequency observatories

Figure 3. Spatial distribution of elevation in meters (A) and TWI filtered with radii of 2.5 km (B) and 10 km (C).

Table 1. Potential predictors used for regression modeling.

ELEV	Elevation (in meters)
ELEV _x	Average elevation within x, where x has a radius of 2.5, 5 and 10 km
MIN_ELEV _x	Minimum elevation within x, where x has a radius of 2.5, 5 and 10 km
MAX_ELEV _x	Maximum elevation within x, where x has a radius of 2.5, 5 and 10 km
RANG _x	Elevation range within x, where x has a radius of 2.5, 5 and 10 km
DIST_RIVERS	Distance to rivers
DIST_MEDIT	Distance to the Mediterranean Sea
DIST_BIS	Distance to the Bay of Biscay
LONG	Longitude
LATIT	Latitude
RADI	Solar incoming radiation
RADI _x	Solar incoming radiation , within x, where x has a radius of 2.5, 5 and 10 km
SLO	Slope
SLO_ELEV	Slope x Elevation
SLO_ELEV _x	Slope x Elevation, within x, where x has a radius of 2.5, 5 and 10 km
PEND_RAG _x	Slope x Elevation range, within x, where x has a radius of 2.5, 5 and 10 km
TWI	TWI
TWI_ELEV	TWI x Elevation
TWI_ELEV _x	TWI x Elevation, within x, where x has a radius of 2.5, 5 and 10 km
TWI _x	TWI within x, where x has a radius of 2.5, 5 and 10 km
TWI _x _ELEV	TWI, within x, where x has a radius of 2.5, 5 and 10 km x Elevation

Figure 4. Upper: relationship between fog frequency (in days) in January and RANG₅ (in meters). Also shown are the equations of the best simple non-linear model and the linear model. Lower: relationship between fog frequency in January and the linearized variable RANG₅.

Figure 5. Response curves of the three predictor variables selected by the GAM in predicting the mean fog frequency during January. Units for elevation and range are in meters and for longitude in decimal degrees

	OLSR	GLM	GAM
January	0.42	0.458	0.722
February	0.378	0.413	0.592
March	0.211	0.255	0.399
April	0.072	0.062	0.442
May	0.121	0.103	0.313
June	0.232	0.277	0.522
July	0.133	0.129	0.302
August	0.162	0.154	0.260
September	0.152	0.128	0.230
October	0.133	0.116	0.384
November	0.295	0.399	0.476
December	0.436	0.476	0.672

Table 2. R^2 values for the monthly models of fog frequency obtained from OSLR, GLM, and GAM.

	OLSR	GLM	GAM
January	ELEV ₅ , TWI ₁₀ , LONG	MAX-ELEV, LONG, RAD ₁₀	ELEV ₁₀ , LONG, RANG ₅
February	MAX_ELEV ₅ , LONG, TWI ₁₀ , SLO	MAX_ELEV ₁₀ , LONG, RANG ₅	LONG, LAT, RANG ₅
March	DIST_BIS, SLO, TWI_ELEV, LATIT, RANG ₅ , SLO_ELEV ₁₀	DIST_BIS, LONG, LATIT, TWI_ELEV, SLO, RANG ₅ , MIN_ELEV _{2.5}	LONG, RANG ₅ , ELEV _{2.5} , ELEV ₅
April	TWI_ELEV, TWI_ELEV ₅	TWI_ELEV, TWI_ELEV ₅	LONG, RANG ₅ , ELEV ₁₀ , ELEV
May	TWI_ELEV, TWI_ELEV ₁₀ , SLO, SLO_ELEV ₁₀	TWI_ELEV, TWI_ELEV ₁₀ , SLO, SLO_ELEV ₁₀	LONG, RANG ₅ , ELEV ₁₀ , ELEV
June	RADI, RAD ₅ , SLO, TWI_ELEV, TWI_ELEV ₁₀	TWI ₁₀ _ELEV, RADI, RAD ₅ , SLO, TWI_ELEV, SLO_ELEV, SLO_ELEV ₅	LONG, RAD ₅ , SLO, MIN_ELEV _{2.5} , ELEV ₁₀
July	TWI ₅ _ELEV	TWI ₅ _ELEV	DIST_RIVERS, LONG, RAD ₅ , SLO, ELEV_MIN _{2.5} , ELEV ₁₀
August	TWI_ELEV ₅ , TWI_ELEV _{2.5}	TWI_ELEV ₅ , TWI_ELEV _{2.5}	DIST_RIVERS, LONG, TWI _{2.5} , ELEV _{2.5} , ELEV ₅
September	TWI ₅ _ELEV, MAX_ELEV ₅ , LONG, TWI ₁₀	TWI ₅ _ELEV, RANG ₅ , LONG	LONG, RANG ₅ , ELEV _{2.5}
October	RANG ₅ , LONG, SLO, TWI_ELEV ₁₀	RANG ₅ , LONG, SLO, TWI_ELEV ₁₀	LONG, RANG ₅
November	MAX_ELEV ₅ , LONG	MAX_ELEV ₅ , LONG, LATIT, DIST_BIS, MIN_ELEV ₁₀	MIN_ELEV ₅ , LONG, RANG ₅
December	MAX_ELEV ₅ , LONG, TWI ₁₀	MAX_ELEV ₅ , LONG, RAD ₁₀	LONG, RANG ₅ , ELEV ₁₀

Table 3. List of predictors included in each model. The names of variables correspond to those in Table 1.

Figure 6. Maps of wintertime (November–February) fog frequency (in days) compiled using OLSR, GLM, and GAM.

Before residual correction														
MAE	January	February	March	April	May	June	July	August	September	October	November	December	Average	Winter
OLSR	1.47	0.90	0.56	0.42	0.48	0.26	0.16	0.19	0.51	1.04	1.54	1.87	0.78	1.45
GLM	1.40	0.84	0.62	0.39	0.46	0.26	0.16	0.19	0.53	1.04	1.56	1.85	0.77	1.41
GAM	1.30	0.90	0.60	0.36	0.45	0.25	0.17	0.16	0.49	1.08	1.39	1.56	0.73	1.29
Average	1.39	0.88	0.59	0.39	0.46	0.26	0.16	0.18	0.51	1.05	1.50	1.76	0.76	1.38
D	January	February	March	April	May	June	July	August	September	October	November	December	Average	Winter
OLSR	0.71	0.61	0.44	0.37	0.34	0.51	0.31	0.70	0.39	0.31	0.57	0.71	0.50	0.65
GLM	0.79	0.67	0.42	0.30	0.33	0.51	0.34	0.71	0.46	0.31	0.61	0.74	0.52	0.70
GAM	0.81	0.67	0.44	0.36	0.34	0.53	0.38	0.72	0.47	0.33	0.69	0.80	0.54	0.74
Average	0.77	0.65	0.43	0.34	0.34	0.52	0.34	0.71	0.44	0.32	0.62	0.75	0.52	0.70
After residual correction														
MAE	January	February	March	April	May	June	July	August	September	October	November	December	Average	Winter
OLSR	1.22	0.81	0.52	0.39	0.48	0.26	0.13	0.19	0.50	1.04	1.46	1.52	0.71	1.25
GLM	1.18	0.81	0.51	0.38	0.46	0.27	0.12	0.20	0.48	1.04	1.32	1.28	0.67	1.15
GAM	1.02	0.80	0.48	0.34	0.47	0.21	0.14	0.19	0.43	1.07	1.33	1.21	0.64	1.09
Average	1.19	0.82	0.54	0.37	0.49	0.26	0.14	0.24	0.48	1.09	1.42	1.46	0.71	1.22
D	January	February	March	April	May	June	July	August	September	October	November	December	Average	Winter
OLSR	0.76	0.81	0.48	0.37	0.39	0.53	0.36	0.76	0.47	0.33	0.62	0.80	0.56	0.75
GLM	0.81	0.83	0.47	0.33	0.39	0.49	0.38	0.76	0.54	0.33	0.71	0.90	0.58	0.81
GAM	0.87	0.83	0.48	0.38	0.38	0.58	0.42	0.80	0.58	0.35	0.73	0.91	0.61	0.84
Average	0.79	0.79	0.46	0.32	0.35	0.49	0.40	0.71	0.52	0.31	0.66	0.84	0.55	0.77

Table 4. Mean absolute errors and D values for different monthly models before and after residual correction. Also shown are the average monthly values and average values for the winter months (November–February).

Figure 7. Monthly maps of fog frequency (in days) obtained from GAMs and local interpolation of residuals.