Analysis of the atmospheric macro-physical using spatial methods

J E Andrades-Grassi¹, L Cuesta-Herrera², H A Torres-Mantilla³, and J Y López-Hernández¹

¹ Universidad de Los Andes, Mérida, Venezuela
² Universidad Católica del Maule, Talca, Chile
³ Universidad de Santander, Bucaramanga, Colombia

E-mail: ledyz.cuesta@alu.ucm.cl

Abstract. The central western area of Venezuela has an unequal distribution of precipitation. Due to its agricultural importance, it is necessary to plan water accounting and this requires an evaluation of spatial and temporal variability of precipitation and an estimate of local geophysical effect from the relief. In this research we use an iterative computationally lattice approach to perform a confirmatory analysis of the variability and the spatial correlation structure in monthly precipitation stations. Spatial correlograms and pooled empirical semivariogram were applied to evaluate the most appropriate spatial weighting matrix to estimate the Moran’s I. The altitude effect over monthly rainfall was estimated through spatial regression algorithm which determine the predominant spatial process in each slice. A homogeneous spatial stochastic process with positive spatial autocorrelation is evidenced. There is a trend towards a higher frequency of spatial error and spatial auto-regressive processes between the months of June and August whilst there are not dominant process between October and December. This response is caused by the dynamics of the intertropical convergence zone, which generates a seasonal effect on precipitation. These estimations allow decision-making in modeling and will lead to an improvement for analysis and forecasting in areas strongly affected by climate change and water stress.

1. Introduction

Central western Venezuela is a key area for the country’s agricultural production. However it has a climate predominantly dry with high temperatures and low rainfall given by its location in the center of it is the Quibor valley depression which is surrounded by the coastal and Andean mountain systems. This zone has an environment of scarce highly heterogeneous precipitations that oscillate between 250 mm and 2200 mm, due firstly to the influence of the inter-tropical convergence zone (ICZ), which generates a seasonal component. Secondly, at a local scale the topography plays an important role, causing orographic rains in those areas where the terrain is oriented in such a way that it forces the wind to rise, cooling adiabatically until the steam condenses and forms clouds that cause rain [1].

The phenomenon described above is an important geophysical phenomenon called relief precipitation relationship. The spatial and temporal variability of rainfall is an important problem in modeling water accounting in agricultural areas with high variability, experiences in
other dry areas have shown the crucial importance of modeling auto-correlation with dependence on the orographic effect \[2\].

The description of the behavior of rainfall requires estimation of dependency structure \[3\] through the spatial autocorrelation (SA), defined as the degree to which objects or activities in a geographic unit are like other objects or activities in nearby geographic units \[4\] and the spatial heterogeneity (SH), refers to the variation of the relationships between variables in space \[5\]. That autocorrelation structure with dependence on the orographic effect is unknown for the central western area of Venezuela and this knowledge will lead to an improvement in water accounting. The objective of this study was to identify the stochastic process of monthly rainfall in central-western Venezuela.

2. Methodology

Confirmatory analysis is a rigorous way of testing the absence or presence of certain statistical properties, which have not been discovered in exploratory analysis \[6\]. The study area correspond to 42 monthly precipitation stations from five states of Venezuela: Trujillo, Lara, Yaracuy, Cojedes, and Carabobo as shown in Figure 1, between the period 1949-2000, in total, 624 temporal observations were available, which are called spatial analysis slices. A spatial confirmatory analysis was applied, their objective is to model the nature of the autocorrelation and the type of stochastic process that predominates and the distance of the correlation structure.

![Figure 1. Selected stations of monthly precipitation.](image)

This data structure is called spatio temporal pooled data. To determine the existence of SA and distance, a univariate exploration of the monthly precipitation was carried out for each of the spatial slices. Using the spdep \[7\], gstat \[8\], ncf and maptools libraries of the R programming language, the distance-dependent SA \[8\] was evaluated by means of the spatial correlogram and the empirical pooled semivariogram as shown in Figure 2. This made it possible
to guide the selection between the different spatial weighting matrices (SWM) to evaluate the presence of global SA in each slice by estimating Moran’s I coefficient, choosing a weighting matrix based on the euclidean distance method, where all the neighbors of the selected range have the same weight. As suggested by [9], an ordinary least squares (OLS) model is fitted on monthly precipitation using altitude as a covariate. Subsequently, on the model residuals, SA tests are carried out using Moran’s I coefficient.

Figure 2. Decision algorithm of spatial regression process.

Unlike Moran’s I, the tests described below not only serve to identify if the spatial Slice has SA, but also to evaluate the type of inherent stochastic process, either spatial-lag \( y = \rho Wy + X\beta + \varepsilon \), where \( \rho \) corresponds to the autoregressive coefficient of the spatial lag of the variable \( y \) and \( \beta \) the coefficient of the exogenous variable \( X \) or spatial-Err \( y = X\beta + \mu \), \( \mu = \lambda W \mu + \varepsilon \), where \( \lambda \) corresponds to the coefficient of moving averages of the lag of the synthetic variable \( \mu \) and \( \beta \) the coefficient of the exogenous variable \( X \) [9], this type of model is used by [10–12] in climatic variables.

The processes described are spatial analogs of the moving average and autoregressive processes and their presence are shown by the statistics associated with LM-ERR test or spatial error test and LM-LAG or spatial lag test. Both statistics have robust forms to the presence of local dependence called LM-EL and LM-LE respectively and are used when both the LM-ERR and the LM-LAG are significant. These tests are based on the principle of lagrange multipliers (LM) and their test statistics are asymptotically distributed as a chi-square distribution with one degree of freedom. The test performed using the Spatial Autoregressive process and moving averages (SARMA) statistic jointly evaluates the presence of both classes of processes [13,14].

In addition to the spatial autocorrelation contrasts, the Breusch-Pagan homocedasticity statistic was estimated to evaluate whether the data have SH in the residuals; and the Shapiro-Wilk statistic to evaluate the normality of the residuals (the non-normality of the residuals would indicate a model specification problem) [3]. Subsequently, the spatial-lag \( y = \rho Wy + X\beta + \varepsilon \) models were applied, the estimation method used was maximum likelihood or spatial \( y = X\beta + \mu \), \( \mu = \lambda W \mu + \varepsilon \), as in the spatial-lag model, the method for estimating the parameters was the maximum likelihood, using the monthly precipitation (mm) of the Slice as endogenous variable and the altitude (meters) as exogenous [9]. With fitted models the quality of each one of them and their residuals was evaluated.

3. Results
From the univariate spatial exploration of the correlograms, the variation has a wave behavior in space, where 2 peaks are manifested, also this is observed in the pooled semivariogram, one located around 27 km and a second peak located at 80 km, which are present along the study
time. Due to this, two SWM were designed, the first at 80 km and the second at 100 km. The 27 km matrix was obviated due to the lack of spatial neighbors it generates inconsistency in the maximum likelihood estimation of the spatial parameters with the chosen SWM as shown in Figure 3.

![Figure 3.](image)

(a) Variation of the spatial correlogram in time; (b) pooled semivariogram; (c) 80 km weighting matrix; (d) 100 km weighting matrix.

Based on the selected SWM as shown in Table 1, the results of the Moran’s I analysis indicate the presence of positive SA, for 70% of the cases; in the 80 km and 100 km matrices; however, in the latter there were 12 cases of negative SA processes (2%). When analyzing the results of the OLS regression, 15% of the regressions were significant and in 10% an HE process was identified, which indicates an apparently spatially homogeneous process. The Shapiro-Wilk test indicated a rejection of the null hypothesis of normality on the residuals in more than 75% of the cases for both matrices.

As for the autocorrelation contrasts, a higher frequency is manifested for all the tests where these are significant for the 80 km matrix, this is indicative that the SA manifests itself optimally at this distance. The temporal behavior of the SA contrasts of the residuals, it indicates a well-defined SA process according to Moran’s I. The residuals of the model showed positive SA, under a SWM scheme of 80 km. The remaining autocorrelation contrasts define the type of stochastic process in time, in which it is highlighted in monthly slices from October to December are those in which the spatial-Err, spatial-lag and SARMA processes, being more frequent in monthly slices from June to August, but the monthly frequency in which these contrasts are significant do not determine a dominant process. An increase in the frequency of significance of the LM-ERR and LM-LAG tests was identified, starting in slices early 1970s.
Table 1. Quality indicators of the OLS regression.

| Test                                | SWM 80 km | SWM 100 km |
|-------------------------------------|-----------|------------|
| Moran’s I significant               | 438       | 438        |
| OLS regression significant          | 98        | 98         |
| Significant Breusch and Pagan       | 64        | 64         |
| Significant and positive Moran’s I residuals | 477 | 439 |
| Shapiro-Wilk Significant            | 469       | 469        |
| OLS regression                      | 98        | 98         |
| Significant LM-ERR                  | 364       | 315        |
| Significant LM-LAG                  | 363       | 320        |
| Significant LM-EL (error)           | 384       | 247        |
| Significant LM-LE (lag)             | 327       | 327        |
| Significant SARMA                   | 393       | 364        |

The adjustment of the spatial-lag spatial models $y = \rho Wy + X\beta + \varepsilon$ and spatial-Err $y = X\beta + \mu, \mu = \lambda W\mu + \varepsilon$ which significance test results for all slices are summarized in Table 2. It indicate a more frequent process of the spatial-Err type, in which was achieved with a higher lambda frequency and significant altitude, with a low frequency of an SH process.

Table 2. Frequency of significance of the parameters of the spatial models.

| Test                                | SWM 80 km | SWM 100 km |
|-------------------------------------|-----------|------------|
| Intercept significant               | 273       | 266        |
| Altitude significant                | 95        | 97         |
| $\rho$ significant and positive     | 322       | 276        |
| $\rho$ significant and negative     | 3         | 0          |
| Wald significant                    | 397       | 333        |
| Significant Breusch-Pagan           | 32        | 32         |
| AIC less than AIC OLS               | 413       | 362        |
| Spatial-lag                         |           |            |
| Intercept significant               | 477       | 464        |
| Altitude significant                | 132       | 123        |
| $\lambda$ significant and positive  | 341       | 294        |
| $\lambda$ significant and negative  | 1         | 4          |
| Wald significant                    | 409       | 360        |
| Significant Breusch-Pagan           | 29        | 32         |
| AIC less than AIC OLS               | 418       | 374        |

In the analysis based on 80 km SWM the frequency of achieving significant $\rho$ and $\lambda$ is maximized in the months of June to August, without a clear dominator, this again is indicative that SA is more frequent for these months; a different behavior is manifested in the rest of the months, since it is $\lambda$ of the stochastic spatial-Err process with the SWM of 80 km, the dominant process at a monthly level. The spatial dominant process has a low HE with a high frequency of linear SA. This response is caused by the dynamics of the ICZ, which generates a seasonal effect on precipitation. Globally, the ICZ is generated by the intense sun and hot water from the equator that heat the air, generating a wide belt of low pressure and high humidity constituted by upward air currents, where large masses of warm and humid air from the north and south intertropical zone [15]. Despite the low number of spatial observations, the dominant stochastic process is spatial-Err type with an effective SWM of 80 km, until 1985, after that date there is no clear dominator of the spatial structure and its matrix.
Comparatively, the altitude has a lower frequency of significance with respect to the results of the parameters $\rho$ and $\lambda$, it is maximized in the months of September to December where it reaches a maximum of 6 months with significant altitude and decreases substantially in the months of June and July. This can be explained by: (a) a non-linear correlation process between monthly precipitation and altitude; (b) poor specification of the spatial model; (c) a random and independent process and (d) the maximum likelihood estimation process is inefficient, due to the low number of spatial observations per slice ($n = 42$). In order to introduce a more complex model from the point of view of the precipitation-altitude relationship, the size of the sample must be increased, since for the estimation of the model parameters to be adequate, there must be a sufficient amount of data per each parameter of the model, which subjectively can vary from 10 to 30 data for each parameter [16].

An important element to consider is the problem proposed by [17–19] who discuss the effect of non-normality in the data which affects the Moran’s I indicator of spatial autocorrelation, they indicate that if the data are under a SAR process (Spatial-Lag), the indicator underestimates the spatial autocorrelation when the parameter $\rho$ moves away from 0, since the data do not have a normal distribution. It is possible that a SAR process may be masked by the occurrence of extreme values, therefore it would be equally convenient to evaluate the possibility of a transformation of the frequency distribution of the precipitation and altitude data; this recommendation is proposed [19] as a solution to face non-linearity in the functional form, since if it occurs between the explanatory variables and the unknown parameters of the model, it is found with a very poor fit, with results lacking in any sense.

A consistent estimate of the variance/covariance matrix over a geostatistical approach would demand a high number of data points [20] while lattice approach let to obtain results with fewer number of observations since it refers to spatial regression models.

4. Conclusions
A linear spatial stochastic process with homogeneity was determined. The orographic effect was shown by the presence of two spatial autocorrelation peaks located around 27 km and 80 km. In slices until 1985 the dominant stochastic process is spatial-Err type with an effective spatial weighting matrix of 80 km. The seasonal effect caused by the dynamics of the intertropical convergence zone was measured through high frequency of spatial autocorrelation significative tests in month slices from June to August.

These results should be taken with caution because the data can affect the response of the estimate, which is due to the non-normal structure and the low number of samples from each of the spatial slices. In subsequent studies, an analysis should be considered in the selection of the spatial weighting matrix, and the use of non-linear indicators of spatial autocorrelation, temporal space and spatial correlation. A geostatistical approximation could be used, but the implementation requires high number of observations. For this reason, a lattice-type spatial approximation is a flexible alternative, which can be used for low number of observations, since when considering the spatial unit as a discrete element.

References
[1] Martelo M T 2003 La Precipitación en Venezuela y su Relación con el Sistema Climático (Caracas: Dirección de Hidrología, Meteorología y Oceanografía-Dirección General de Cuencas Hidrográficas-Ministerio del Ambiente y los Recursos Naturales)
[2] Subyani A M 2004 Geostatistical study of annual and seasonal mean rainfall patterns in southwest Saudi Arabia / Distribution géostatistique de la pluie moyenne annuelle et saisonnière dans le Sud-Ouest de l’Arabie Saoudite Hydrological Sciences Journal 49(5) 803-817
[3] Gujarati D, Porter D 2010 Econometría (México: McGraw-Hill/Interamericana Editores)
[4] Goodchild M F 1987 A spatial analytical perspective on geographical information systems International Journal of Geographical Information System 1(4) 327-334
[5] LeSage J P 1999 The Theory and Practice of Spatial Econometrics (Toledo: University of Toledo)
[6] Hipel K W, McLeod A I 1994 Time Series Modelling of Water Resources and Environmental Systems (Netherlands: Elsevier Science)
[7] Bivand R, Hauke J, Kossowski T 2013 Computing the jacobian in gaussian spatial autoregressive models: An illustrated comparison of available methods Geographical Analysis 45(2) 150-179
[8] Bivand R, Pebesma E, Gómez-Rubio V 2008 Applied Spatial Data Analysis with R (New York: Springer)
[9] Anselin L 2005 Exploring Spatial Data with Geoda™: A Workbook (Urbana: University of Illinois)
[10] Al-Ahmadi K, Al-Ahmadi S 2014 Spatiotemporal variations in rainfall-topographic relationships in southwestern Saudi Arabia Arabian Journal of Geosciences 7 3309-3324
[11] Hasanean H, Almazrou M 2015 Rainfall: Features and variations over saudi arabia, a review Climate 3(3) 578-626
[12] Subyani A M 1999 Topographic and seasonal influences on precipitation variability in southwest saudi arabia JKAU: Earth Sci 11(1) 89-102
[13] Burridge P 1980 On the cliff-ord test for spatial autocorrelation among regression residuals Geographical Analysis 4 267-284
[14] Anselin L 1988 Spatial Econometrics: Methods and Models (Boston: Kluwer academic publishers)
[15] Adams J 2010 Vegetation-Climate Interaction: How Plants Make the Global Environment (Berlin: Springer)
[16] Hair J F, Black W C, Babin B J, Anderson R E, Tatham R L 2009 Multivariate Data Analysis (Porto Alegre: Bookman Editora)
[17] Dormann C, McPherson M, Araújo M, Bivand R, Bolliger J, et al. 2007 Methods to account for spatial autocorrelation in the analysis of species distributional data: a review Ecography 30(5) 609-628
[18] Li H, Calder C A, Cressie N 2007 Beyond Moran’s I: testing for spatial dependence based on the spatial autoregressive model Geographical Analysis 39(4) 357-75
[19] De Antonio M G 1999 Econometría Espacial: Algunos Aspectos Generales (Madrid: Universidad Complutense de Madrid)
[20] Oliver M A, Webster R 2015 Basic Steps in Geostatistics: The Variogram and Kriging (New York: Springer International Publishing)