Global Advanced Research Journal of Agricultural Science (ISSN: 2315-5094) Vol. 4(2) pp. 118-126, February, 2015.
Available online http://garj.org/garjas/index.htm

Full Length Research Paper

Influence area determination of rain gauge stations in the coffee region, Caldas department, Colombia

Julián A Valencia A[1], Oscar G Castillo R [2], John F Betancur [3], Andrés J Peña Q [2]

[1]Master Student. College of business and Economic Sciences, University of Manizales. (Manizales, Caldas, Colombia).
[2]AgWeatherNet, Washington State University.(Prosser, WA, United States)
[3]Research Center for Environment and Development, University of Manizales. (Manizales, Caldas, Colombia)

Accepted 17 February, 2015

According to statistics of DANE (2012), Caldas department has suffered a significant reduction in the coffee planted area; it was 92,000 hectares in 2005 and 79,000 in 2012. However, this crop is still of great importance for the region as it represents about 6% of the provincial gross domestic product (GDP). According to Rojas and Gast (2013), decrease in planted area could be explained by reduction in coffee production in recent years. This reduction was related to wet environmental conditions (excessive rainfall) during last years, and it is largely explained by the presence of four "La Niña" events during this time frame. In order to define the area of action for a weather station in an Andean region, records were used of daily rainfall alone last 25 years, taken from a network with 21 weather stations. Daily data were analyzed by using geostatistics, taking into account the experimental semivariograms to define the range of action of a weather station. Specifically, the concept used was the range, calculated from the experimental semivariogram. Results show that, on average, in the Andean region, a rainfall station has between 6 km of a radius influence (daily data) and 15 km of radius (annual data).

Keywords: Rain, coffee, range, influence area, semivariogram

INTRODUCTION

According to official reports from DANE (2012), the department of Caldas has undergone a significant reduction in the area planted in coffee, from 92,000 hectares in 2005 to 79,000 in 2012. However, this crop is still of great importance to the region as it represents about 6% of departmental gross domestic product (GDP). It also generates great contributions to national productivity in the sense that the volumes of coffee produced in the region represent over 10% of the annual coffee production in the country. Decreasing department production could affect not only the regional yields but also the national economy. According to Rojas and Gast (2013), decrease in planted area could be explained by reduction in coffee production in recent years. This reduction was related to wet environmental conditions (excessive rainfall) during last years, and it is largely explained by the presence of four "La Niña" events during this time frame (Figure 1).

Several authors agree that climate elements are a factor threatening cropping systems. In order to decrease the negative effects of climate on crop systems some strategies have been designed to reduce the exposure of the agricultural production. These are focused on reducing the probability that farmers lost production as a result of the weather conditions. (Peña .et al., 2012; Ramírez et al., 2012.). Such as Arce(2013) argued, this is the reason guild producers are increasingly interested in conducting and

*Corresponding Author's Email: julianandres.valenciaarbelaez@gmail.com

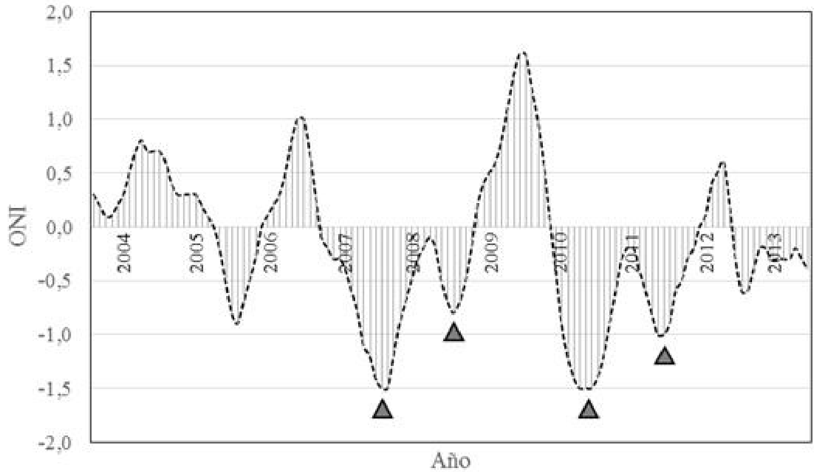

Figure 1. Oceanic Niño Index (ONI) between 2004 and 2013. Between 2008 and 2012 there was four La Niña (▲) events

updating the agroclimatic zoning. These zonings are based on the knowledge of the physical environment (soil and climate), for which it is necessary the monitoring of variables that characterize the ambient. However, unlike soil variables (that are evaluated once every certain period); by its high temporal variability, climatic elements need to be constantly sampled. It is accomplished by installing sensors that record information over time. In the specific case of precipitation, to obtain records a sensor that quantifies the volume of water passing through a given area, located one meter above the ground surface (gauge) is used.

Despite simplicity of the rain gauges, even nowadays, in developing countries, the probability of having a network of rain gauges was low. This situation has changed as a result of studies that show the effects of climate on agricultural production. However, for mountainous areas, where Colombian coffee cultivation is currently present, there is high uncertainty regarding the number of weather stations to be installed to have adequate coverage. In this paper, rainfall dataset of Caldas department are used to determine the area of mean influence of rainfall stations located in mountain conditions at different scales time.

Because the spatial data is an important input to make decisions in earth science, like meteorology, geophysics, and other sciences, interpolation is a necessity. Interpolation consists of estimating values where it was not observed, using values observed in neighboring points (Montoya et al., 2000). Nevertheless, the degree of certainty of the data generated by any of the existing techniques of interpolation depends on the quality of the sampling, which results in meteorology stations number per unit area. One of the strategies used to determine the ideal number of stations is through reducing the error variance in the data interpolated via kriging (Pardo-Iguzquiza, 1998; Rojas and Mora, 2009). In this paper, we established, in a more accurate way, the optimum sampling distances. It was possible by using the regionalized variable principle (Viera, 2002), and elementary analysis of the experimental semivariogram (Van Es et al., 1989; Van Fagroud and Meirvenne, 2002).

METHODOLOGY

Dataset

Daily rainfall records of 21 meteorological stations measured during the last 25 years, located in the Caldas department (Table 1) were used. Dataset belongs to climate database of the meteorological service of the National Federation of Coffee Growers of Colombia.

Based on the daily data series, decadal, monthly and annual rainfall series were gotten, resulting in four timescales for analysis. From the daily database, 120 days were selected randomly, trying to have a representative sample of each month. From the decadal database, 72 decades were randomly selected. From the monthly database, 60 months were selected, and 20 years were selected from the annual database. In this case the Random function of Microsoft Excel ® 2010 was used.

Geostatistics

Geostatistics is a methodology for surface generation by interpolation. But, unlike more simple methods such as Inverse Distance Weighted (better known as IDW) and Thiessen polygons (González, et al. 2007), a dependence spatial analysis is done before the interpolation. This previous analysis, known as spatial structure of the data

Table 1. Location of stations used in the study

Station	Department	Town	Altitude	Latitude (N)		Longitude (W)	
			m	Degrees	Min	Degrees	Min
AGRONOMIA	CALDAS	MANIZALES	2088	5	3	75	30
CENICAFE	CALDAS	CHINCHINA	1310	5	0	75	36
CUATRO ESQUINAS	CALDAS	AGUADAS	1900	5	40	75	25
EL DESCANSO	CALDAS	MARMATO	1803	5	30	75	37
EL RECREO	CALDAS	PALESTINA	1430	5	2	75	39
GRANJA LUKER	CALDAS	PALESTINA	1031	5	4	75	41
GUAYMARAL	CALDAS	AGUADAS	1600	5	39	75	27
JAVA	CALDAS	MANIZALES	1778	5	1	75	32
LA ARGENTINA	CALDAS	RIOSUCIO	1420	5	28	75	42
LA JULIA	CALDAS	FILADELFIA	1650	5	18	75	34
LA LINDA	CALDAS	PACORA	1750	5	33	75	32
LA MANUELITA	CALDAS	RIOSUCIO	1460	5	22	75	41
LA PALMA	CALDAS	PALESTINA	1165	5	1	75	41
LA PASTORITA	CALDAS	VICTORIA	1122	5	19	74	58
LA SELVA	CALDAS	MANIZALES	1312	5	5	75	36
LA SIERRA	CALDAS	CHINCHINA	1440	4	59	75	38
NARANJAL	CALDAS	CHINCHINA	1381	4	58	75	39
RAFAEL ESCOBAR	CALDAS	SUPIA	1307	5	27	75	38
SANTA HELENA	CALDAS	MARQUETALIA	1395	5	19	75	0
SANTAGUEDA	CALDAS	PALESTINA	1026	5	4	75	40
LA ARGENTINA	CALDAS	PALESTINA	1354	5	2	75	41

(structural analysis), is usually done through the variogram, semivariogram or correlogram (Giraldo, 2000). In this case, we use experimental semivariograms, one of the tools used to analyze the spatial dependence of one variable on a defined area. Semivariogram or semivariance function, which characterizes the spatial dependence of properties of a process is estimated by the method of moments (Giraldo, 2000) as shown in equation 1

$$\bar{\gamma}(h) = \frac{\sum(Z(x+h) - Z(x))^2}{2n}$$

Equation 1. Model used for the semivariogram

Where z(x) is the value of the variable in a place x; z(x + h) is a previous sample, separated from x by a distance h; while n is the number of pairs that are separated by this distance. To interpret the semivariogram we part from the view that the smaller distance between two sampling sites, the greater the similarity or spatial correlation between observations and thus less variance. (Giraldo, 2000).

The calculation of the semivariance, experimental semivariogram model fit and the same was done using GS+ v 10 (Gamma Design Software, 2004), following the methodology used by Peña et al (2009). Half the maximum distance between two sampling points was used as a range to calculate the semivariance and cumulative rainfall data throughout the network in the proposed scales (daily, decadal, monthly and yearly) were considered. Although there are various theoretical models of semivariance that can be adjusted to experimental semivariogram, which are divided into unbounded (linear, logarithmic, potential) and dimensional (spherical, exponential, Gaussian) (Warrick et al., 1986), it's the second group which guarantees the increasing of covariance is finite and, therefore, are widely used when there is evidence showing good fit (Giraldo 2000), thereby we decided to use for this work a model of exponential type.

The theoretical model for each Semivariograms was preset, but the coefficient of determination (r 2), the relationship between sill and nugget (equation 2) and range as the distance from which two observations are independent were taken into account.

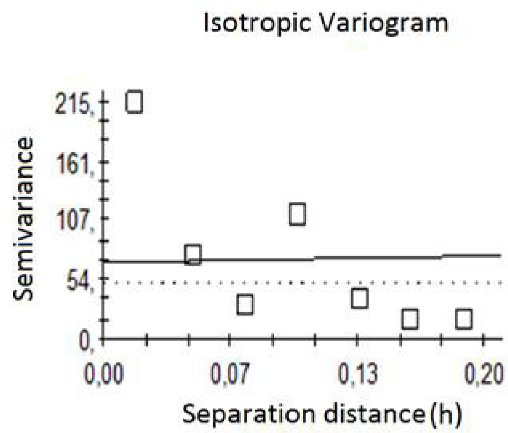

Figure 2. Characteristics of a pure nugget effect, evidenced for the rain values on January 30, 2005

$$\frac{C}{(Co+C)}$$

Equation 2. Relationship between the sill and the nugget. Where C is the structural variance, Co is the nugget and Co + C is the sill

The range has a particular interest, i.e., the required minimum separation between two sampling points for the values of a variable to be spatially independent. This experimental semivariogram variable indicates the region in which there is spatial dependence (Gonzales, et al. 2006); this means when the range is small the sampled phenomenon is spatially independent and vice versa. However its practical use is related to the establishment of optimum meteorological, geological and soil sampling distance variables (Peña, 2009) so it is considered as a key element in defining the average radius of action of any spatial sampling. The range does not always explicity appear in the semivariogram. For the exponential model, the range is equal to the distance for which the semivariogram takes a value equal to 95% of the sill (Equation 3).

$$\gamma(h) = C_0 + C_1\left(1 - \exp\left(\frac{-3h}{a}\right)\right)$$

Equation 3. Exponential model, where the effective range is a/3

It is assumed that the radius is near zero when the Semivariograms feature discontinuity point at the origin of the semivariogram (nugget effects), being indicative of which part of the spatial structure is concentrated under lower distances than the observed ones. It is indicative of lack of spatial correlation between the observations of a variable. (Figure 2.)

Basic statistics

With the values obtained form range and proportion from the exponential model generated by GS+ version 10, and which were considered of significant importance, even above the R2 (which was not taken into account in the present study) the value of influence area in square kilometers for each weather station evaluated was obtained by multiplying the value of the range, assuming a degree of latitude equals 110.86 Km and the area of influence has shaped a circle.

The range value (as a basis for determining the area of influence of each season) for each time scale was analyzed through basic statistics; specifically the percentiles 25, 50 and 75 (as values that divide the distribution into 100 equal parts) to generate the "BoxPlot" and determine which is the approximate coverage of each season in different time scales and well defined levels of accuracy and uncertainty for each scenario, considering the rainfall variable.

Determination of the areas covered in the department

The generation of areas of influence involved the creation of an area around each point representing the geographic location of each station. In this case, the methodology was based on the construction of a sect of polygons, depicting "buffer" which in this case is the area of influence of each station, which is built using the range value found from the analysis of semivariogram. According to Burrough & McDonnell (1998) buffers are useful when you want to

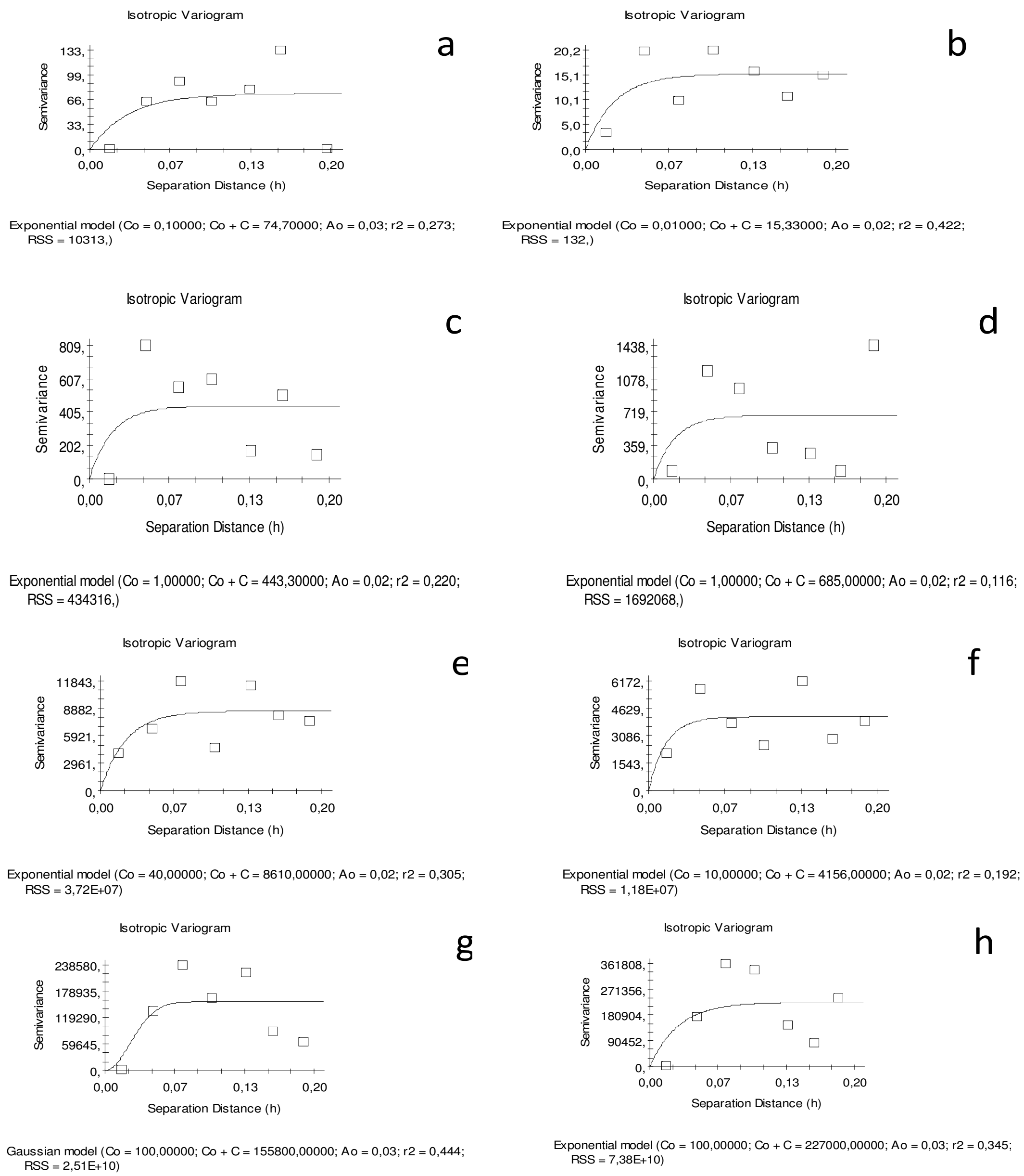

Figure 3 Example of semivariogram obtained for: Daily **Scale)** March 11, 2007, **b)** 22 April 2004. decadal scale **c)** 2 2004 **d)** 1997 14 monthly scale **e)** January 1997 **f)** June 1991. Annual Scale **g)** 1992 **h)** 2011 belonging to 21 stations.

resolve issues such as defining which entities are inside or outside the specified area of influence.

The tool called a “Dissolved Buffer”built into the ArcGIS software version 10.2 to create a new polygon coverage area of influence around coverage features specified input was used. This tool was used to identify or define the area within a distance "x" around each weather station at different timescales mentioned. Percentile values (25, 50 and 75) were used to create multiple rings as influence areas in order to know the representativity of weather stations in function of rainfall.

RESULTS

Geostatistics

As results of geostatistical analysis we have two basic outputs, experimental semivariograms for each time scale analyzed (Figure 3) and summary tables with information from each experimental semivariogram (Table 2).

Table 2. Example values semivariogram parameters for data obtained for daily scale, decadal, monthly and yearly.

Date	Scale	R2	Proportion	Residue	Nugget	Sill	Range
03/11/2007	daily	0,273	0,999	10313	0.10	74.7	0,096
04/22/2004	daily	0.422	0,999	132	0.01	15.3	0,069
10/12/2011	daily	0.084	0,999	13141	0.10	118.8	0,039
02/2004	decadal	0,220	0,998	434,316	1.00	443.3	0,057
14/1997	decadal	0.116	0,999	1692068	1.00	685.0	0,054
32/2007	decadal	0,223	0,999	7653290	1.00	1858.0	0,057
01/1997	monthly	0,066	0,998	20900000	10.00	5033.0	0,003
06/1991	monthly	0.833	1,000	32200000	10.00	21120.0	0.846
11/1990	monthly	0.311	1,000	5809946	1.00	2248.0	0.081
1992	annual	0.364	0,999	3,05E + 10	100.00	153,400.0	0.081
2011	annual	0.345	1,000	7,38E + 10	100.00	227,000.0	0.084
1990	annual	0.396	0,999	2,70E + 10	100.00	185,200.0	0,069

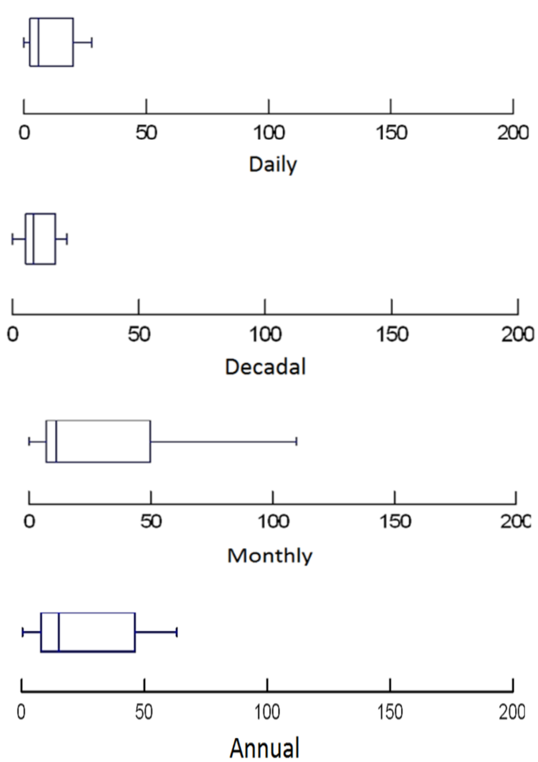

Figure 4. Boxplot indicating the area of influence at different time scales in 21 weather stations in the department of Caldas

Descriptive Statistics

The 25th, 50th and 75th percentiles for the range variable were obtained by generating the experimental semivariogram for randomly selected dates at each time scale (Figure 4). In this case the 25th percentile means that 25% of the cases analyzed have a radius smaller than *cases* "X" action range, whereas P50 is interpreted as half the events analyzed for each time scale have lower values than this (Figure 4).

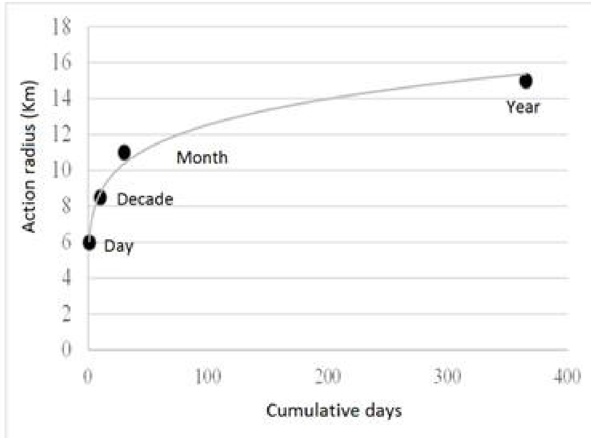

Figure 5. Action radius as a function of time scale analyzed

Table 3. Areas covered (potential and real) as analyzed by the median

	Area (km^2)		
Scale	**Actual cover**	**Cover potential**	**Coffee potential**
Daily	1014	2375	5257
decadal	1489	4767	5257
monthly	2008	7982	5257
annual	2753	14844	5257

It was found to radius that action radius (range) a rainfall station in the Andean region is dependent on the time scale analyzed; ie, the area covered is less when it comes to daily data and is higher when annual data are being analyzed (Figure 5).

The same figure shows that the radius of action of a rainfall station is 6 Km at daily scale, at decadal scale is 8.5 Km, it's 11 Km on a monthly basis and 15 km on an annual scale. This does not mean that you should make measurements of rainfall only scales exceeding one month time to achieve have a large coverage area, it really means is that depending on the purpose of the rainfall network should have more or less rain gauges per unit area. For example, if the objective is to calculate agricultural water balances for irrigation purposes for coffee cultivation, where daily data is needed, in the department of Caldas 47 well-distributed rainfall stations to meet this need would be needed. In the case of water balances for monitoring the water content in the soil, where data are needed on decadal scale, a total of 24 well-distributed rain gauges are needed, while if wants to make a monthly monitoring the cumulative rainfall only 14 rain gauges well distributed on the coffee zone of the department would be required; while if it wants to reference the cumulative annual rainfall only 8 precipitation stations well distributed over the coffee zone of Caldas (Table 3) would be required.

The big difference between the actual area covered and covered potential area to decadal, monthly and yearly scales show that to solve the problem with these time scales, rather than installing stations should redistribute the current rainfall network, as there is redundant information in all time scales (Figure 11).

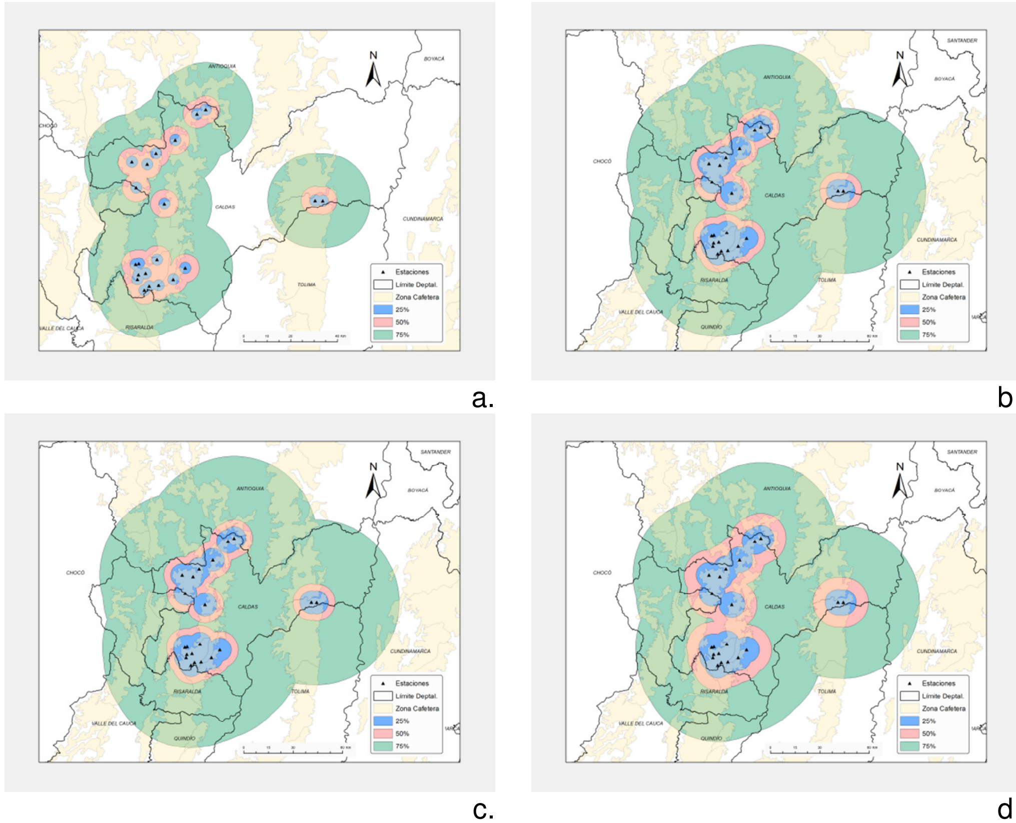

Figure 11. Buffer used to indicate the area of influence daily scale in the 21 weather stations in the department of Caldas. Daily scale, b. Decadal scale, c. Monthly scale, d. Annual scale.

DISCUSSION AND CONCLUSIONS

Clearly, though the use of experimental semivariogram to determine areas of influence of meteorological stations, biological and hydrological sciences, and especially in soil science has not been reported, its use has yielded very good results when you want to optimize resources in soil samples (Peña et al, 2009), so that the costs of a detailed soil analysis can be much smaller scale without affecting the quality of spatial analysis inherent to it. Cambardella et al (1994) conducted studies to determine the spatial variability in soil properties in the state of Iowa, USA, using semivariogram and the ratio of nugget regarding the sill, as an indicator of the spatial dependence expressed in percentage. He found that a ratio of less than 25% indicates a strong spatial dependence, between 25 and 75% reported moderate spatial dependence and greater than 75% indicated weak spatial dependence.

The experimental semivariogram provides important for subsequent variable interpolation information; in this case use of sill-nugget ratio and range to determine optimal sampling distances. In this work it was found that, by citing an example, for agricultural applications, where models require the use of data on a daily scale, the area of influence of an average rainfall station in the equatorial Andes is less than 120 km^2 and that on a monthly scale, as it is generally used in hydrological applications, the influence area is less than 400 km^2. The area of influence of monthly cumulative rainfall from a rain gauge installed in the Andean region is consistent with the suggestion by WMO (Arsenault et al, 2013), while for the daily scale this value is very high, suggesting that in some cases the dynamics of atmospheric variables in mountain conditions is not known and most jobs with which we make a reference in our country come from totally dissimilar sites. For example, Arsenault and Brissette (2013) found that the area of influence of a rain gauge in flat areas ranges between 1600 km^2 and 4000km^2. It should be noted that even at annual scale, when the biggest areas of influence of a rain gauge in the Andean region are found, this area is less than 1000 km^2.

It is clear that, improving the rain interpolation in the Andean conditions is necessary to increase the density of stations installed, as claimed by Grimes, Et al (1999), this requires a financial investment. An optimal network is needed to ensure high certainty when making decisions at the departmental level. Currently, the rainfall network of Caldas covers 43% of the area planted with coffee, if it wants to make decisions at the annual scale, but if the decision is made on decadal scale, as it is currently done in the National Federation of Coffee, the coverage is only 19%. However, the current network has high redundancy, so the network could increase its coverage through station repositioning, as shown in Table 3.

ACKNOWLEDGEMENTS

The National Coffee Research Center -Cenicafé FNC- for providing climate information to generate the present study.

REFERENCES

Arce RR (2013). Ordenamiento Territorial y Cambio ClimáticoMetodología para incorporar Cambio Climáticoy Gestión del Riesgo de Desastres en procesos de OT. Programa "Adaptación de la agricultura y del aprovechamiento de aguas de la agricultura al cambioclimático en los Andes - Programa AACC 2010-2013. BN: 12.2005.2131.51. Lima - Perú

Arsenault R, Brissete F (2013). Determining the Optimal Spatial Distribution of Weather Station Networks for Hydrological Modeling Purposes Using RCM Datasets: An Experimental Approach Ecole de Technologie Superieure, Montreal, Quebec, Canada.

Cambardella C, Moorma T, Novak J, Parkin T, Karlen D, Turco y R Konopka A (1994). Field-scale variability of soil properties in central Iowa soils. Soil Sci. Soc. Amer. J. 58: 1501-1511.

Departamento Nacional de Estadística (DANE). (2012). Informe al Congreso de laRepública, 24, 47, 68. República de Colombia

Gamma Design Software (GDS) (2004). GS+. Geostatistics for the environmental sciences. Gamma design software. Plainwell, MI.

Giraldo HR (2000). Introducción a la geoestadística: Teoría y aplicación. Universidad Nacional de Colombia, sede Bogotá.

González J, Guerra F, Gómez H (2007). Conceptos básicos de geoestadística en geografía y ciencias de la tierra: manejo y aplicación. Geoenseñanza Vol. 12, núm.1: 81-90, Universidad de los Andes. Venezuela

Krstanovic PFy, Singh VP (1992). Evaluation of Rainfall Networks Using Entropy. Water Resource Manager 6:279-293.

Moreno Pérez JA, Giráldez Cervera JV (2005). Diseño óptimo de una red meteorológica en Andalucía. Congreso Andaluz del Agua

Pardo-Iguzquiza E (1998). Optimal selection of number and location of rainfall gauges for areal rainfall estimation using geostatistics and simulated annealing. Journal of Hydrology, 210, 206-220.

Peña R, Rubiano Y, Peña A, Chaves B (2009). Variabilidad espacial de los atributos de la capa arable de un Inceptisol del piedemonte de la cordillera Oriental (Casanare, Colombia). Agronomía Colombiana, Vol. 27, núm. 1.

Peña, A., Ramirez, V., Valencia, J., Jaramillo, A.2012. La lluvia como factor de amenaza para el cultivo del café en Colombia. Avances Técnicos Cenicafé. No. 415.

Ramirez, V., Jaramillo, A., Peña, A., Valencia, J. 2012. El brillo solar en la zona cafetera colombiana, durante los eventos El Niño y La Niña. Avances Técnicos cenicafé. No. 421

Rojas, J. M., & Gast, F. (2013). Cenicafé a través de 75 años. In F. Gast, Manual del cafetero colombiano. Tomo I: 27-46. Manizales: Cenicafé.

Rojas, M. I., & Mora, L. E. (2009). Diseño óptimo de redes pluviométricas. Revista Forestal Venezolana, 53: 9-22.

Ruíz Muñoz, D. 2004. Manual de estadística. Universidad Pablo de Olavide. Sevilla, España. Editorial Eumed.net

Warrick A.W, Myers D.E, Nielsen D.R (1986) Geostatistical methods applied to soil science. Methods of soil analysis, part 1, 2nd edn Physical and mineralogical methods. Agronomy Monograph 9, pp 53–82