XVIII International Conference on Water Resources
CMWR 2010
J. Carrera (Ed)

REGIONAL PARAMETERIZATIONS OF STOCHASTIC MODEL SIMULATION ON HIGH RESOLUTION TEMPORAL RAINFALL: CASE STUDY BOGOTÁ (COLOMBIA)

Hugo Rico T. *, Ana Santos R. †, Eddy Herrera D. ††, Nelson Obregón N. †††

* Pontificia Universidad Javeriana, Bogotá, Colombia
E-mail: hrico@javeriana.edu.co
†Universidad Nacional de Colombia, Bogotá D.C. Colombia.
E-mail: acsantosr@unal.edu.co
††Pontificia Universidad Javeriana, Bogotá D.C. Colombia,
E-mail: eherrera@javeriana.edu.co
††† Javeriana University Geophysical Institute. Pontificia Universidad Javeriana
E-mail: nobregon@javeriana.edu.co

Key words: Rainfall, Neyman-Scott model, Parameter estimation, Regionalization of parameters

1 INTRODUCTION

It happens frequently that in a region one cannot count on rainfall records in a continuous way, neither spatially nor timely. Different time scale resolutions are of great need in studies of floods, reservoir design, rainfall sewage systems and rainfall runoff modeling. Especially institutions in charge of precipitation measurement are not distributed in a uniform way, which suggests that to obtain precipitation values or parameters associated to this phenomenon, it is necessary to turn to a spatial interpolation model to regionalize the above mentioned parameters and this way esteem a value in a place for which no records exist.

In order to approximate the precipitation values for scales larger than on a daily basis, it is common to use stochastic models. In literature, two punctual models of temporal disaggregation of rainfall data are acknowledged: the model of Barlett-Lewis and the one of Neyman-Scott. The model of Neyman Scott is not a physical model; it is a statistical model that is also known as the composed model of Neyman Scott, in which each process in the generation of synthetic rainfall is generated with a different probabilistic structure.

In the present investigation, the second one was used. This model is composed of a series of parameters that allow generating synthetic series of precipitation at different timely resolutions.

The model of Neyman-Scott is conceived for punctual analysis, so it is proposed to study the parameters' spatial variation from the precipitation records of 20 rain gauges located in the city of Bogotá, Colombia. First, an optimization process for estimating the parameters in each station was carried out, using the Levenberg-Marquardt method as a search algorithm. Afterwards, the parameter regionalization was conducted, with the purpose of finding evidence of spatial variation and to verify if a relation with factors like orography or timely resolution exists.

• En matemáticas y computación, el algoritmo de Levenberg-Marquardt, también conocido como el método de mínimos cuadrados amortiguados, se utiliza para resolver problemas de mínimos cuadrados no lineales. Estos problemas de minimización surgen especialmente en el ajuste de curvas de mínimos cuadrados.

2 CASE STUDY

The city of Bogotá is located in South America, in the center of Colombia, in the oriental chain of the Andes Mountains. Its location is given by the coordinates of 4° 35' 53" North and 74° 4' 33" West, at an altitude of 2630 meters above sea level. Bogotá has a total area of 1,776 square kilometers (see Figure 1). The rainfall season is characterized by two periods, being the rainy seasons, the periods between April and May, and from September to November. The average temperature is about 13°C, ranging from 7 to 18°C.

Sources of information

This project used the data of 20 rain gauges operated by the Institute of Hydrology, Meteorology and Environmental Studies (Instituto de Hidrología, Meteorología y Estudios Ambientales, IDEAM) and the city's Institution in charge of the Aqueduct and Sewage System (Empresa de Acueducto y Alcantarillado de Bogotá, EAAB). The available gauges are located concentrated within the urban area of the city of Bogotá, although coverage on a rural level also exists. The time period of the project is the first rainy season of May, 1995. Due to the fact that there are no precipitation records for some years, only a scale of aggregation of an hour could be used to show the benefits of this methodology. In this case, only an hourly scale was used, but still work was continued with different levels of aggregation (3 hours, 6 hours, 12 hours, 1 day) to try to find the best results. This is interesting, since in the city of Bogotá the spatial variability of rainfall is high.

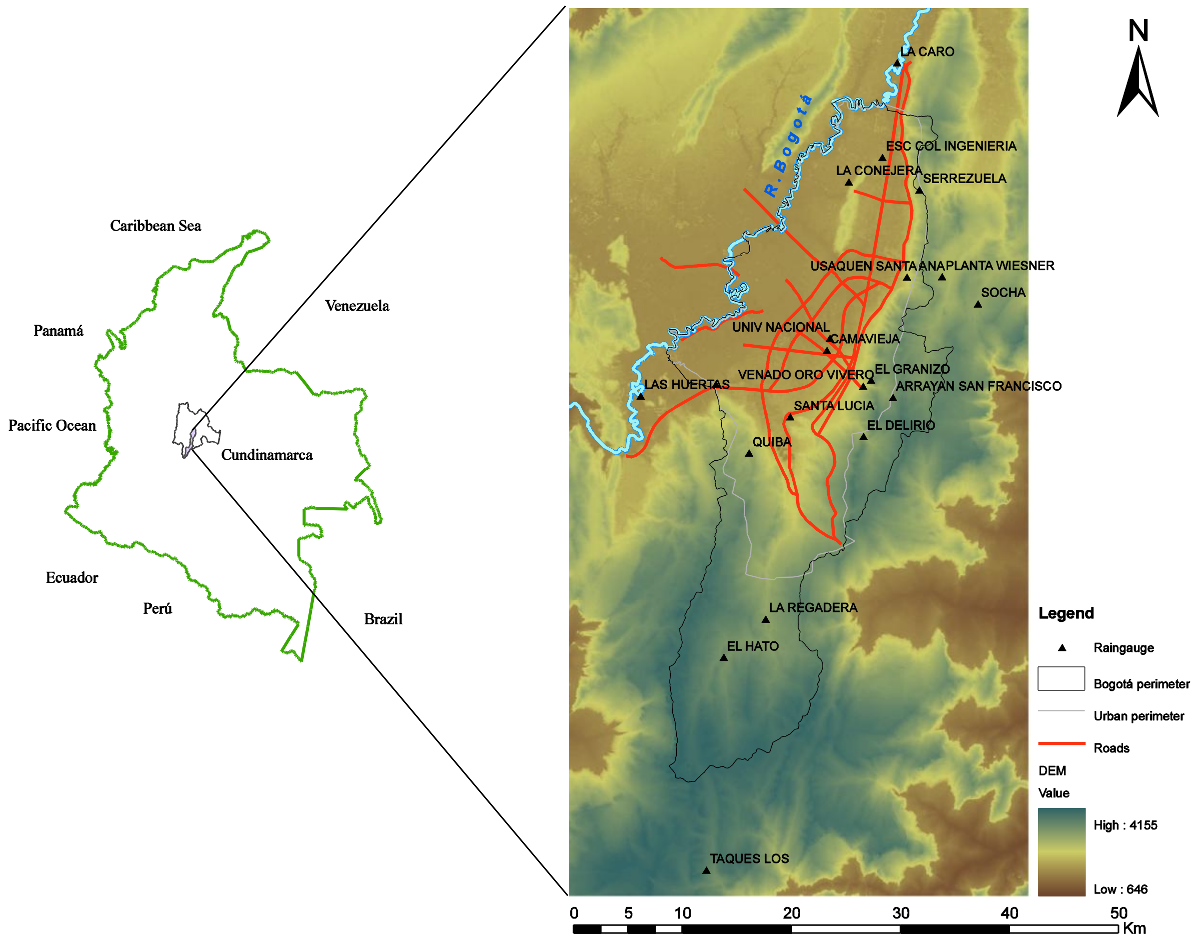

Figure 1: Localization of the project area

.

3 RAINFALL MODEL

Definition of the Neyman-Scott model

The stochastic simulation is a pulse model [1] and [2]. The Neyman-Scott model is described by three stochastic processes, which are all independent:

1. A process that defines the origin of the storms.
2. A process that determines the number of rain cells generated in each storm.
3. A process that defines the origin of the cells.

The model considers that the beginning of each pulse does not coincide with the origin of the rainfall, apart from the origin of each pulse being independent from the existence of another pulse within the same event, because of which a superposition of pulses can exist. Since it is a stochastic model, it is important to have in consideration that various "realizations" with different seeds have to be carried out.

The Neyman-Scott model is composed of five density function *f*, each one being associated to a parameter (Table 1):

Process	Parameter	Distribution
Rate of storm origin arrival	λ	Poisson
Mean number of cells in each storm	$1/\mu_c$	Geometric
Mean intensity of the pulse	$1/\mu_x$	Exponential
Mean waiting time for the rain cells after the storm origin	β	Exponential
Mean cell duration	η	Exponential

Table 1: Parameters of the Neyman-Scott model

Since the precipitation data are normally measured in discrete time intervals, it was considered the total accumulated precipitation in disjoint time intervals in fixed time scales ***h***. That means

Y: es la variable aleatoria que mide la cantidad de agua que cae en un intervalor de tiempo *h*

$$Y_i^{(h)} = \int_{(i-1)h}^{ih} Y(v)\,dv \qquad (1)$$

The expressions that describe the model are expressed through the theatrical second-order properties [3]:

$$E\left[Y_i^{(h)}\right] = \frac{\lambda}{\eta}\mu_c\mu_x h \qquad (2)$$

$$Var\left[Y_i^{(h)}\right] = \frac{2\lambda}{\eta^3}\left(\eta h - 1 + e^{-\eta h}\right)\times\left(2\mu_c\mu_x^2 + \frac{\beta^2\mu_x^2\mu_c(\mu_c - 1)}{\beta^2 - \eta^2}\right) - 2\lambda\left(\beta h - 1 + e^{-\beta h}\right)\frac{\mu_x^2\mu_c(\mu_c - 1)}{\beta^2\left(\beta^2 - \eta^2\right)} \qquad (3)$$

• De la ecuación (1) a la (4) es nuestro sistema de ecuaciones.
• Reemplazar los valores iniciales en las ecuaciones.
• Vamos a tomar *h* como menos de 1 pero no tan cercano a cero. Podria ser 0,5.

$$Cov\left[Y_i^{(h)}, Y_{i+k}^{(h)}\right] = \frac{2\lambda}{\eta^3}\left(1 + e^{-\eta h}\right)^2 e^{-\eta(k-1)h}\times\left(\mu_c\mu_x^2 + \frac{\beta^2\mu_x^2\mu_c(\mu_c - 1)}{2\left(\beta^2 - \eta^2\right)}\right) - 2\lambda\left(1 - e^{-\beta h}\right)^2 e^{-\beta(k-1)h}\times\frac{\mu_x^2\mu_c(\mu_c - 1)}{2\beta^2\left(\beta^2 - \eta^2\right)} \qquad (4)$$

Obtaining and optimizing the parameters

The method of moments [1], [4], [5] and [6] is one of the most frequently used one in solving this kind of problem, the search and estimation process of the five parameters of the model for different temporal resolutions ***h*** is supported by the conservation of statistical properties.

In this case, the five first moments of the timely series of observations were determined and associated with the theoretical moments of the Neyman-Scott model, which is shown in equation (2,3 and 4), for each station.

The objective of the estimation of the parameters is to determine a set of parameters that meets the conditions (2), (3) and (4) and minimize the error function given in (5).

$$\Theta(\lambda,\mu_c,\mu_x,\beta,\eta)=\sum_{i=1}^{p}\left(1-\frac{M_i}{\hat{M}_i}\right)^2 \qquad p\geq 5$$

(5)Esto es la función de error!!!!

To find the parameters, Levenberg- Marquardt's algorithm to resolve linear systems was used, a method based on Newton's method [7], which optimizes an object function being subject to constraints. This method has been used in different investigations and good results have been obtained [8], which is characterized by the method's rapidity, which show convergence at 1000 iterations in a couple of seconds (6 seconds on average).

While a different objective function could have been used, literature refers to this objective function, which gives good results and quick convergence for the parameters that are to be calibrated [8].

This objective function is sensible respecting the seed value, because of which the inadequate selection of input variables can produce better results. This is why this work seeks to find the initial seed based on different works previously conducted in this topic [3] and [4]. The variation of simulated parameters in this investigation coincides with the results in other investigations, as is shown in Table 2.

	λ [h^{-1}]	η [h^{-1}]	β [h^{-1}]	μ_c	μ_x [$mm \cdot h^{-1}$]
Maximum	0,13172	1,93119	1,76218	163,13353	1,24822
Mean	0,05527526	1,8235495	1,598633	39,911237	0,2857265
Minimal	0,01723	1,71236	1,27779	1,84417	0,01456

Table 2: Range with which the parameters differ in the simulation

Table 4 shows parameters variation in each of 20 stations of the area of interest

Rain gauge	λ [h^{-1}]	η [h^{-1}]	β [h^{-1}]	μ_c	μ_x [$mm \cdot h^{-1}$]
PLANTA WIESNER	0,03789	1,81357	1,68820	30,94994	0,15665
SOCHA	0,07701	1,72569	1,60217	28,77420	0,11425
ARRAYAN SAN FRANCISCO	0,13172	1,89434	1,76218	1,84417	1,24822
EL DELIRIO	0,10996	1,71236	1,43654	92,42760	0,02296
EL HATO	0,07740	1,92780	1,63650	61,94869	0,06301
LA REGADERA	0,06033	1,93119	1,60858	10,36994	0,47698
TAQUES LOS	0,12774	1,78379	1,43933	141,56981	0,01846
LA CONEJERA	0,02899	1,83763	1,71935	11,95864	0,37064
USAQUEN SANTA ANA	0,02437	1,80754	1,63480	10,06140	0,32064
SERREZUELA	0,03394	1,83348	1,59924	10,96835	0,26980
VENADO ORO VIVERO	0,03200	1,82466	1,61305	10,44186	0,30010
EL GRANIZO	0,01723	1,78281	1,68012	10,02581	0,35151
LA CARO	0,03105	1,84073	1,60714	11,06652	0,28364
ESC COL INGENIERIA	0,02968	1,87353	1,65463	10,01687	0,42053
UNIV NACIONAL	0,04509	1,84129	1,63587	12,08802	0,30903
CAMAVIEJA	0,03775	1,90089	1,69827	13,98053	0,37705
BOSA BARRENO No 2	0,03548	1,80562	1,56070	145,73562	0,01621
SANTA LUCIA	0,04127	1,80111	1,59355	11,50410	0,20443
QUIBA	0,08191	1,76494	1,52465	163,13353	0,01456
LAS HUERTAS	0,02731	1,76802	1,27779	9,35914	0,37586

Table 3: Variation of the statistical parameters of the probability function used in the experiment. Aggregation scale 1 hour. Month May 1995

Simulations and synthetic series

With the results about parameters synthetic series were generated for each station. Figure 3 shows comparison between observed series and simulated series for the Venado de Oro station. Figure 4 shows comparisons between moments in observed series and synthetic series.

In a time scale of 1 hour, which was the one that was analyzed, it was found that the parameters found by the inverse model, adjust to the moments of the time series, except in the forth moment (second coefficient of the autocorrelation function), where the moments were overestimated, as can be observed in Figure 2.

The error variation is compared using three performance criteria for the simulations. RMSE, which represents the mean squared error, is a simple summing measurement of the residuary, which results in a good performance approximation for the model. In this case, an approximation close to zero indicates a good performance, which is the case in the mean, the second and third coefficient of the autocorrelation function. In the mean and the first moment of the autocorrelation function, the model does not yield good results. The MSRE is a method that is very sensitive to big errors, which indicated the relative error, and like in the previous case the variance and the first coefficient of the autocorrelation function are not similar to the observed data. Finally, the Pearson product moment correlation coefficient is used, in which, like the previous cases, the values close to zero indicate a good performance and those close to one a bad estimation, and describe the proportion of the statistical variance of the set of observed data.

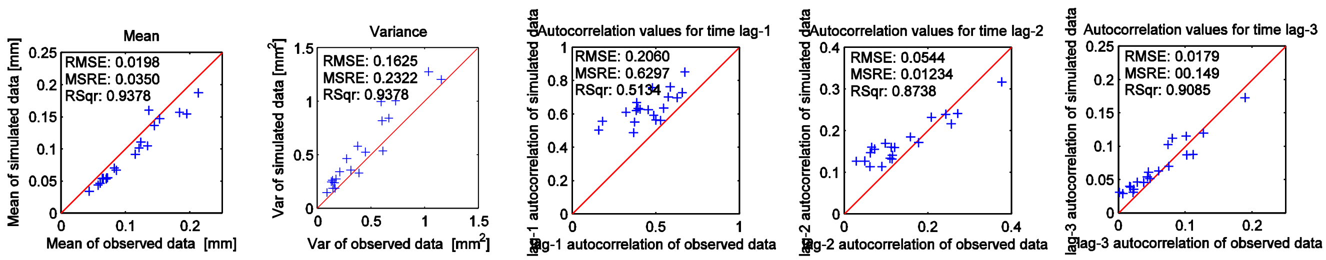

Figure 2: Dispersion diagram for the observed and simulated moments with the Neyman-Scott model for the 20 rain gauges used in the experiment at a temporal resolution of an hour.

With the results about parameters synthetic series were generated for each station. Figure 3 shows comparison between observed series and simulated series for the Venado de Oro station. This figure presents one of the many simulation types that were carried out, which because of spatial limitations cannot be included in the report. For this example it can be noted that the model can simulate the texture of the observed series.

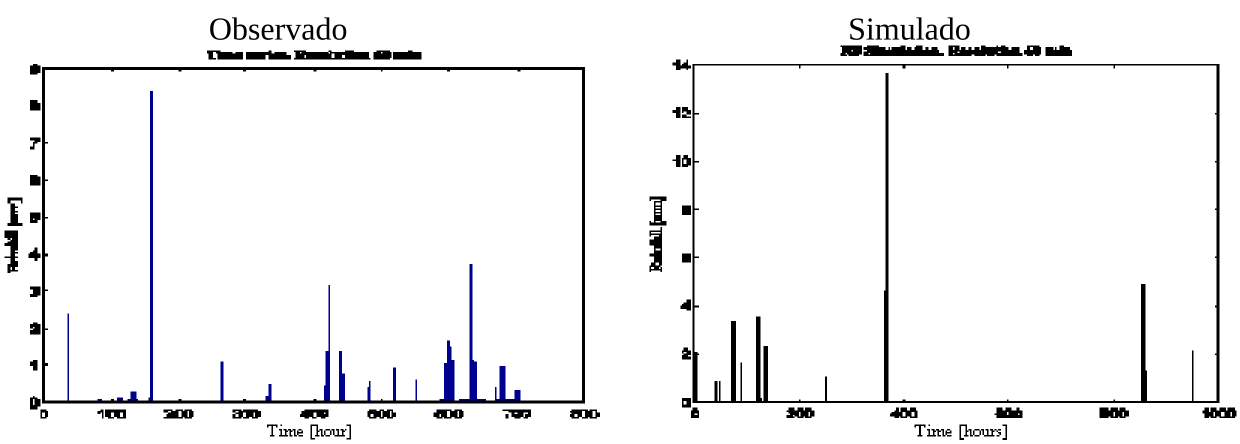

Figure 3: Comparison of stochastic simulation of synthetic series of precipitation from the model of Neyman-Scott and observed seires for the raingauge "VENADO ORO VIV

Rain gauge	Mean [mm]		Variance [mm²]		Autocorrelation values for time lag-1 [-]		Autocorrelation values for time lag-2 [-]		Autocorrelation values for time lag-3 [-]	
	Obs.	Sim.	Obs.	Sim.	Obs.	Sim.	Obs.	Sim.	Obs.	Sim.
1	0,1209	0,1013	1,0371	1,2755	0,5473	0,6351	0,1145	0,1320	0,0223	0,0307
2	0,1530	0,1467	1,1546	1,2026	0,6277	0,6972	0,1784	0,1712	0,0472	0,0513
3	0,1360	0,1601	0,6115	0,5360	0,5298	0,5607	0,1205	0,1596	0,1023	0,0870
4	0,1448	0,1363	0,3887	0,3282	0,6746	0,8515	0,3770	0,3168	0,1892	0,1728
5	0,1840	0,1567	0,6660	0,8403	0,5005	0,5631	0,1094	0,1334	0,0380	0,0457
6	0,1950	0,1545	0,7302	1,0036	0,3667	0,4875	0,0901	0,1136	0,0233	0,0353
7	0,2132	0,1871	0,4498	0,5250	0,5757	0,7019	0,2430	0,2394	0,1274	0,1195
8	0,0825	0,0699	0,6051	0,8149	0,4559	0,6257	0,0500	0,1271	0,0017	0,0308
9	0,0574	0,0435	0,1406	0,2605	0,3947	0,6332	0,0726	0,1554	0,0433	0,0541
10	0,0708	0,0548	0,1599	0,2361	0,4057	0,6273	0,0985	0,1695	0,0755	0,0699
11	0,0723	0,0550	0,1754	0,2739	0,3825	0,6203	0,1124	0,1601	0,0450	0,0611
12	0,0436	0,0341	0,2084	0,3401	0,3859	0,6676	0,0629	0,1468	0,0171	0,0399
13	0,0702	0,0529	0,1299	0,2468	0,3216	0,6109	0,0659	0,1599	0,0602	0,0628
14	0,0856	0,0667	0,2752	0,4622	0,1807	0,5569	0,0291	0,1264	0,0182	0,0388
15	0,1153	0,0915	0,3766	0,5796	0,4879	0,5918	0,1153	0,1405	0,0282	0,0460
16	0,1342	0,1047	0,5933	0,9950	0,3741	0,5507	0,0622	0,1134	0,0069	0,0288
17	0,0605	0,0464	0,0894	0,1453	0,4799	0,7574	0,2088	0,2319	0,0808	0,1119
18	0,0638	0,0539	0,1449	0,1857	0,6581	0,7287	0,2561	0,2159	0,0742	0,1024
19	0,1239	0,1103	0,3117	0,3546	0,5867	0,7624	0,2711	0,2407	0,1015	0,1152
20	0,0653	0,0543	0,1657	0,1861	0,1582	0,5028	0,1588	0,1848	0,1117	0,0875

Table 4: Variation of the synthetic series' precipitation statistics in different time scales for the city of Bogotá during a period between 01-05-1995 and 31-05-1995

4 Regionalization of the parameters

Applying Inverse distance weighting method (IDW), parameters regionalization was done. (See Figure 4). In this case the parameters were interpolated in places where there was no information. The advantages of this methodology lie in the regionalization of the statistical parameters, in which seems to exist information where there are no stations that measure the rainfall intensity, and the results are similar to those of other methodologies that have been performed in the city of Bogotá, where thunderstorms were regionalized [9]. Although the parametrical sensibility analysis was not executed, the experiment was continued with different time scales, exploring other optimization alternatives.

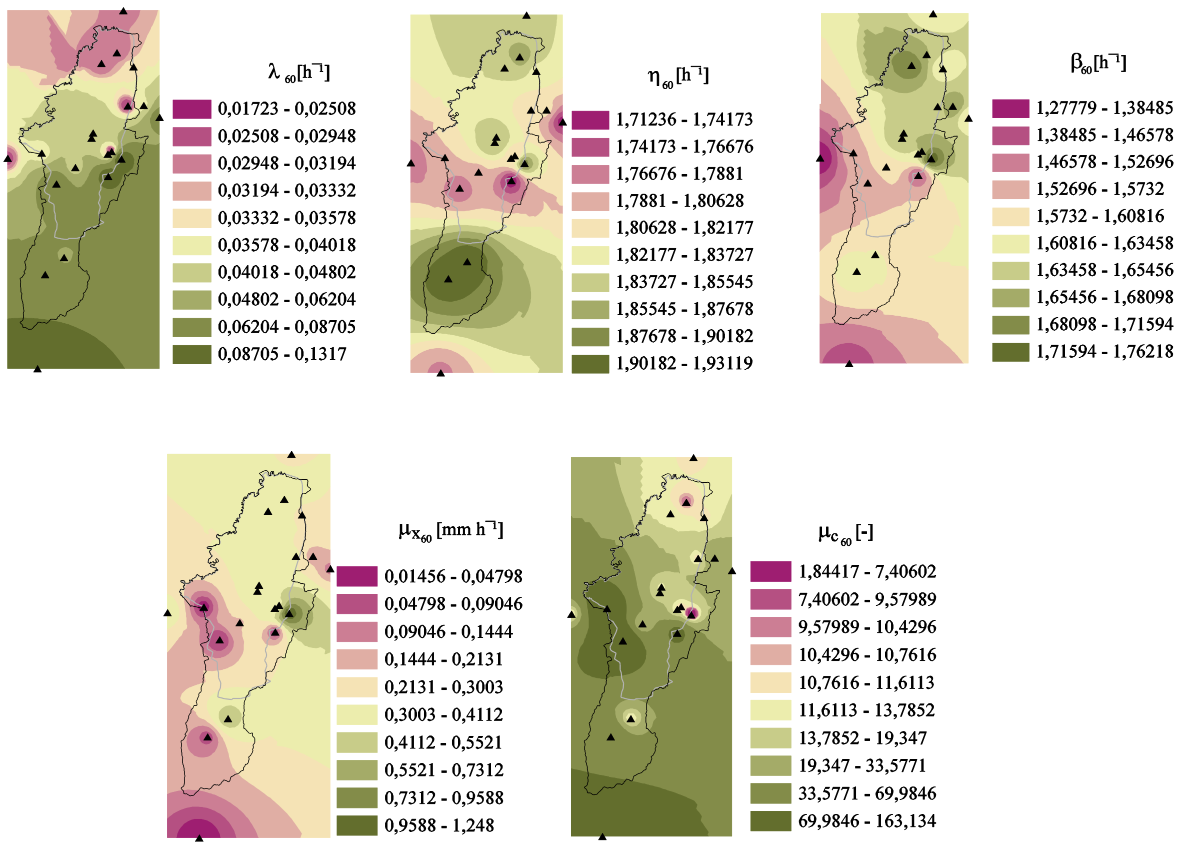

Figure 4: Regionalization of the parameters of the Neyman-Scott model for the month of May, 1995.

5 CONCLUSIONS

It was found that the parameters are very sensible at different time scales, which means that the regionalization varies depending on the scale. The seed's sensibility is great, so it is recommended that a previous analysis be executed to establish more detailed ranges of variation for the parameters.

The estimation of the parameters for the Neyman-Scott model requires that the process be examined carefully. In the optimizing process, different solutions for each set of initial points can be found.

The contribution of this investigation is based on the functionality of having the parameters of a disaggregation model in a space, which was the product of a modeling experiment, so some considerations can be inferred in the moment of realizing the structural designs or the planning of water resources in cities.

An approach of the disaggregation of precipitation is based on the existence of historical records on an hourly time scale with similar characteristics as the place that is being considered. The stochastic models for the disaggregation, based on punctual processes are used and recognized within the hydrological community. It is proposed to investigate the capacity of the Neyman-Scott model as a tool to disaggregate temporally the precipitation, while calibrating the parameters of the model within a two-dimensional space, and afterwards regionalize those parameters and verify their values in certain points.

6 REFERENCES

[1] Rodríguez-Iturbe, Cox, D. R., Isham, V. Some modelsfor rainfall based on stochastic point processes. *Proc. R. Soc*. London A 410, p. 269 – 288. 1987a.

[2] Rodriguez-Iturbe, I., Cox, D.R., Isham, V. A point process model for rainfall: further developments. Proceedings of the Royal Society of London Series A 417 (1853), 283– 98. 1988.

[3] Rodríguez-Iturbe I., Gupta V. and Waymire K. Scale considerations in the modeling of temporal rainfall. *Water Resources Research*. Volumen 20. N° 11. pp 1611-1619. 1984.

[4] Rodriguez-Iturbe, I., Febres De Power, B., Valdes, J. Rectangular pulses point process models for rainfall: analysis of empirical data. *Journal Geophysical Research* 92 (D8), 9645–9656. 1987b.

[5] Entekhabi, D., Rodriguez-Iturbe, I., Eagleson, P.S. Probabilistic representation of the temporal rainfall process y a modified Neyman–Scott rectangular pulses model: parameter estimation and validation. *Water Resources Research* 25, 295–302. 1989.

[6] Cowpertwait, P.S.P. Further developments of the Neyman– Scott clustered point process for modeling rainfall. *Water Resources Research* 27, 1431–1438.Corr: Vol. 28, p. 1497. 1991.

[7] Lourakis, Manolis I.A. A brief description of the Levenberg-Marquardt Algorithm implemented. Institute of Computer Science, Foundation for Research and Technology - Hellas (Forth), Heraklion, Crete, Greece. Frebruary 2005.

[8] Favrea, Anne-Catherine, Musyb Andre´, Morgenthalerc Stephan. Unbiased parameter estimation of the Neyman–Scott model for rainfall simulation with related confidence interval. Journal of Hydrology 286 (2004) 168–178

[9] IRH, Ingeniería y Recursos Hídricos Ltda. Ingenieros Consultores. Estudio para el análisis de tormentas en la sabana de Bogotá. Santafé de Bogotá, Noviembre de 1995