CAPÍTULO

18

The planning of the precipitation

stations network for the territory of Ecuador

POR

CÉSAR ALONSO ZULETA PADILLA

Introduction

The rational planning of a meteorological network is a subject that has been studying for years but by nowadays it has not been solved satisfactory yet. (Czelnai, 1963), (Hoffman, 1970). The purpose of the present work, in the first part is to show the mathematical-statistics principles of a meteorological field, mainly for that the behavior of a meteorological element can be explained by the stochastics processes theory (Gulyás, 1980).

To know better the areal and temporal distribution of a meteorological, element, many scientists around the world have been working on it and several methods have been appeared. One of the most successful is the so-called Optimal Linear Interpolation Method (Gandin, 1970) that was applied to planning rationally the precipitation stations network in the territory of Ecuador.

The measurements' obtained data of the quantity of precipitation is not always exactly, they get many errors as the punctuality of the machine, the visual error at the moment of measuring, etc. In addition to the above, we have to consider other factors as the climatic differences within the studied zones, e.g. the orographic factors, the distances to the sea, etc.

Materials and methods

Materials

INAMHI (Instituto Nacional de Meteorología e Hidrología, Ecuador) Meteorological Year Books and FORTRAN Program.

The Method

The used methods are descriptive, analytic and applied. The investigation begins with general conditions of quantitative approach to the planning of meteorological stations network.

According to (OMM, 2011) each network must be optimized to provide data and function properly at a reasonable cost. Most optimization methods are based on data from an existing network. They are based on both temporal and spatial statistical analyzes of time series. The determination of redundant stations allows network administrators to consider possible options for optimizing them, for example by eliminating redundant stations to reduce costs or using resources to establish stations at sites where observations are required for more effectively achieve network objectives. The density and distribution of the climatological stations to be established in a terrestrial network of a given area depend on the meteorological elements to be observed, the topography and the use of the lands in the zone and the information needs of the specific climatic elements in question. The stations should be located so as to provide representative climatic characteristics that fit all types of terrain. The density of the station should depend on the purpose for which the observations are made and the use to be made of the data. When planning a terrestrial network, an intermediate solution between the ideal density of the stations and the resources available to install, operate and manage those stations is often necessary.

The applied method for this investigation is the named The Optimal Linear Interpolation that is given by the following expression:

$$\hat{f}_o = \sum_{i=1}^{n} p_i \varphi_i + q, \;\; i = 1, 2, \dots, n \qquad (1.1)$$

Where:
$\hat{f}_o$ is the value to be estimated
p_i are weighting factor
q is an addition factor
The weighting factors p_i are determined to minimize the interpolation error S characterized by the mean quadratic deviation between the estimated and real values:

$$S^2 = E\left\{\left(\sum_{i=1}^{n} p_i \varphi_i - \hat{f}_o\right)^2\right\} \qquad (1.2)$$

The weighting factors have to satisfy the following condition:

$$\frac{\partial S^2}{\partial p_i} = 0 \qquad (1.3)$$

The application of the optimal interpolation becomes easy if the examined meteorological field is homogeneous and isotropic because the correlation function depends only on the distance. Then we obtain a linear system of equations given by the following formula:

$$\sum_{j=1}^{n} p_j \mu_{ij} + p_i \zeta_i \;\; (4) \qquad (1.4)$$

Where:
μ_{ij} is the correlation value
ζ_i is the relative error of the interpolation
The interpolation error is given by the formula:

$$S^2 = 1 - \sum_{i=1}^{n} p_i \mu_{i0} \qquad (1.5)$$

The Mathematical-Statistics Structure

According to (Gandin, 1970), the mathematical – statistics structure is given by the following process:
Let *fi* be the value of some meteorological element showed with *f* in the r_i point, in other words, $f_i = (r_i)$.

The expected value, $m_i = Ef_i$ point out the average magnitude of the random variable's value. The expected value is estimated with the mean, which is the unbiased-estimation. Let's suppose that there are N observation in n points related to me f element. Then the mean gives the following expression:

$$\widehat{m}_i = \frac{1}{N}\sum_{k=1}^{n} f_{ik}, \;\; i = 1, 2, \dots, n \qquad (2.1.)$$

Where f_{ik} is the k observation in the i point.
The values of the random variable oscillate around the expected value. The value of the oscillation is given by the variance, D^2, and by the standard deviation, $\sqrt{D^2}$.

$$D_i^2 = D^2 = E[(f_i - Ef_i)^2], i = 1, 2, \dots, n \qquad (2.2)$$

Its estimation can be calculated by the following formula

$$\widehat{D}_i^2 = \widehat{m}_i = \frac{1}{N-1}\sum_{k=1}^{n}(f_{ik} - \widehat{m}_i)^2 \;\; i = 1, 2, \dots, n \qquad (2.3)$$

In the case of a multi-dimensional standard deviation, we must know the stochastic relation between them. In order to know it, the covariance function is used given by the following formula:

$$b_{ij} = b(\underline{r}_i, \underline{r}_j) = E\left[(f_i - Ef_i)(f_j - Ef_j)\right], i, j = 1, 2, \dots, n \qquad (2.4)$$

Its estimation is given by the next formula

$$\widehat{b}_{ij} = \widehat{b}(\underline{r}_i, \underline{r}_j) = \frac{1}{N}\sum_{k=1}^{n}(f_{ik} - \widehat{m}_i)(f_{jk} - \widehat{m}_j) \;\; i, j = 1, 2, \dots, n \qquad (2.5)$$

It is obvious, that the covariance of the same components is:

$$b_{ii} = b(\underline{r}_i, \underline{r}_j) = D_i^2, i = 1, 2, \dots, n \qquad (2.6)$$

It is more frequently to use the correlation function defined by the expression:

$$\mu_{ij} = \mu_{ij}(\underline{r}_i, \underline{r}_j) = \frac{b_{ij}}{D_i D_j}, i, j = 1, 2, \dots, n \qquad (2.7)$$

Its estimation is given by the next for formula:

$$\widehat{\mu}_{ij} = \widehat{\mu}_{ij}(\underline{r}_i, \underline{r}_j) = \frac{\widehat{b}_{ij}}{\widehat{D}_i \widehat{D}_{j_j}}, i = 1, 2, \dots, n \qquad (2.8)$$

The theory of the homogeneity and isotropy help us in knowing the mathematical-statistical structure of a meteorological field.

A meteorological field is considered homogeneous in relation to the covariance (correlation) function if the cov (corr) function depends only on the distances of the points; i.e.

$$b(\underline{r}_i, \underline{r}_j) = b(\underline{r}_l, \underline{r}_q) \quad \text{if } |\underline{r}_i - \underline{r}_j| = |\underline{r}_l - \underline{r}_q| \qquad (2.9)$$

A meteorological field is considered isotropic in relation to the covariance (correlation) function if the cov (corr) function depends only on the vector and the $p = |\underline{r}_i - \underline{r}_j|$, i.e.:

$$b(\underline{r}_i, \underline{r}_j) = b(\underline{r}_l, \underline{r}_q) \quad \text{if } |\underline{r}_i - \underline{r}_j| = |\underline{r}_i - \underline{r}_l| \qquad (2.10)$$

In addition to it, it is necessary that the expected value must be equal in all points. We consider this point as part of the theory of homogeneity and isotropy.
As a resume, we can say that:
- The first moment (the expected value) is equal in all points

$m_i = m = const$

- The second moments related to the mean (cov, corr functions) depend only on the

$b_{ij} = b(\rho)$ distance and thus $D_i^2 = D^2$
$\mu_{ij} = \mu(\rho)$
Where $p = |\underline{r}_i - \underline{r}_j|$

Results and discussion

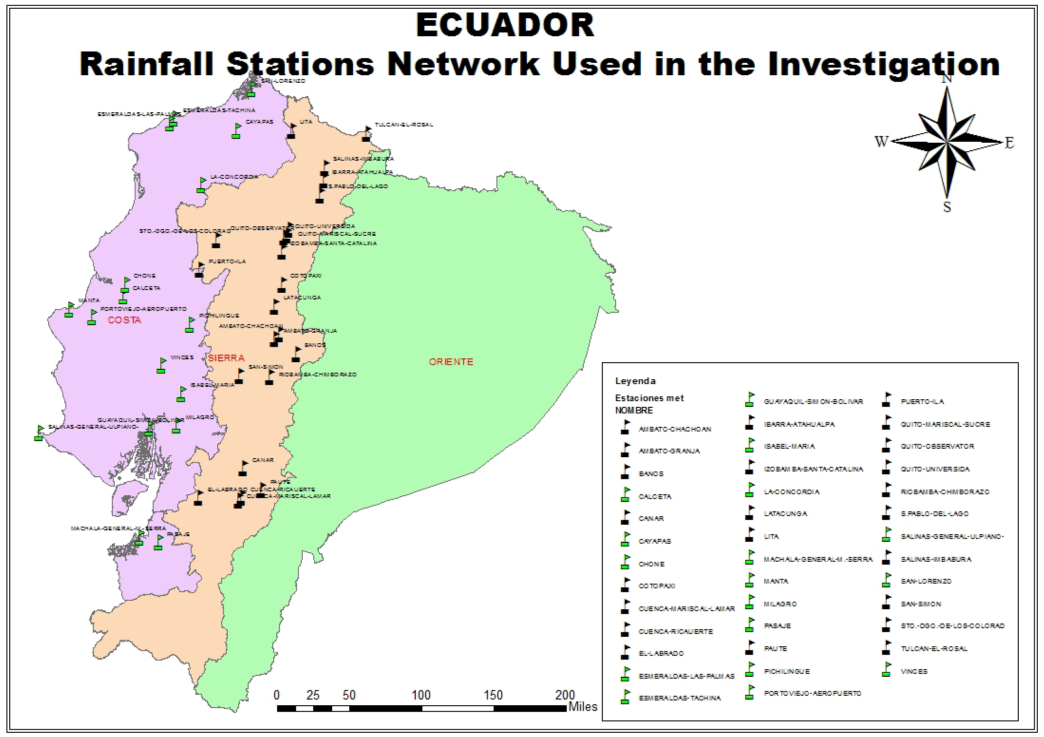
Particularities of the Precipitation's Distribution in Ecuador

In the present investigation, Ecuador's two zones were examined, the Coast and the Highland (Sierra).

In the Coast, there are two different regions; one of them is characterized for to be too dry with small annual precipitation amounts, the Humboldt Current, which brings cool and dry air to the Ecuadorian seaside, causes it. This region is situated just next to the sea and its dimension is about 10-30 km E-W. The another region is very rich in rainfall, the annual amounts are around 1500-4000 mm from a 15 years observation database, this part is situated at the center of the Coast where generally, the period from January to April is the wettest.

In the Highland the distribution of the annual precipitation shows two maximums which fall to November and April. July and August are the driest months of the years, in general, the quantity of the precipitation decreases from North to South. We cannot find a strongest relation between the orography and the distribution of precipitation. (INAMHI, 1976-1981)

For each examined zone 20-20 daily precipitation station's data were used for a period of five years from 1975 to 1979, thus we had 155 or 150 or 141 precipitation data for every station. These stations are geographically irregular situated in order to obtain most representative values. The **GRAPHIC 1**. shows the distribution of the precipitation stations in the two examined regions used in this paper. As it can see from the distribution map, the analysis was based on the stations' available data. The main purpose of the investigation was to optimize the number of stations from the territorial and spatial point of view.



GRAPHIC 1
Map of Rainfall Stations Network. *Author*: Own elaboration

The calculated correlation values are strongest in regions with more quantity of precipitation and weakest when it origins from combined zones (arid and wet zones).

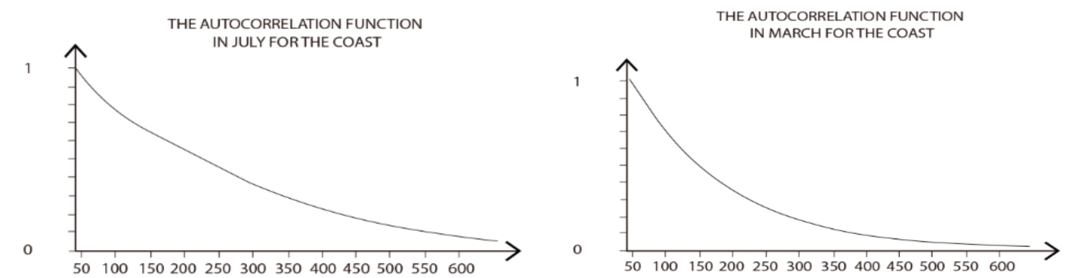
The variation of the correlation values in respect to the decreasing is moderate. The function that describes the distribution is named Autocorrelation Function (Box, 1976) and can be calculated in every case by the following way:

$$\mu(\rho) = ae^{-\alpha_1\rho} + (1-a)e^{-\alpha_2\rho}$$

Where:
$\propto_1, \propto_2, a$ are empiric factors
p is the distance
And it is needed that $\propto_1 » \propto_2$

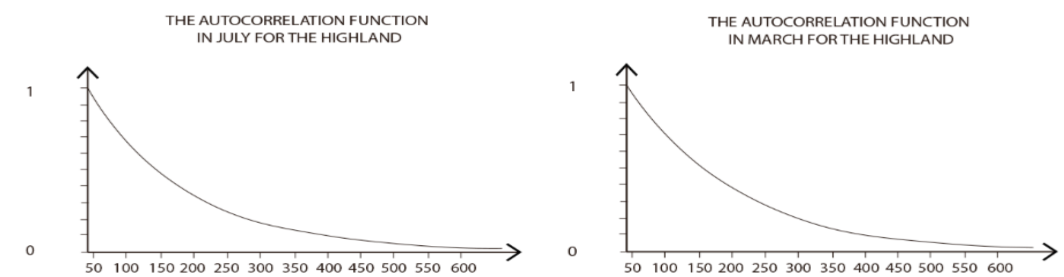
The approximation was made on basis of that the first member of the equation decreases fastest to zero but it has a highest coefficient. The **2.A.** and **2.B. GRAPHICS** show the Autocorrelation Function for the Coast in March and July.

The **2.A. AND 2.B. GRAPHICS** show the Autocorrelation Function for the Coast in March and July



GRAPHIC 2A., 2B.
Author: Own elaboration

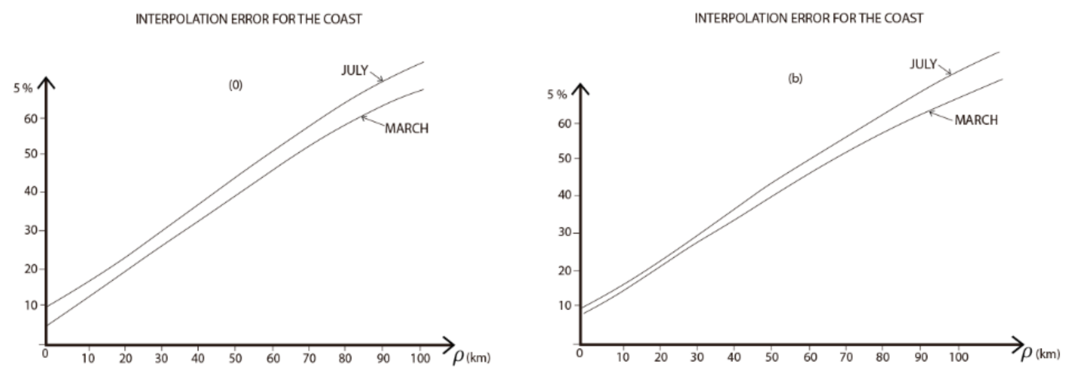
The **3.A. AND 3.B. GRAPHICS** show the Autocorrelation Function for the Highland in March and July.



GRAPHIC 3.A, 3.B.
Author: Own elaboration

If we examine the autocorrelation function for every month, we can see that these functions have a great degree of similarity when they origin from a same season, in this case the variation is low. The variation becomes to be highest when the season changes from wet to dry (or vice versa). From the graphics above, the experimental function in July decreases slower than in March, it means that the correlation values are higher in July than the ones in March. Knowing the weighting factors, the interpolation error can be calculated and on basis of a determined accuracy, the optimal distances between stations can be deduced.

The scale of the interpolation errors between the different stations are very widely, thus we averaged the results for every zone depending on the number of stations we like to interpolate with. The interpolation was made using the data of 1, 2, 4, 6 and 10 stations. In every case, we choose the stations which are situated closest to the station to be interpolated. (Hoffmann, 1971). In the following means the smallest distance between the station to be interpolated and the closest stations. The **4.A. GRAPHIC** show the interpolation error for the Coast and the 4.b. for the Highland.



GRAPHIC 4.A, 4.B.
Author: Own elaboration

We can see from these graphics that the averaged error grows almost as a linear function. In the Coast, the error in July is smaller than in March. We can deduct from it that the optimal distance between stations has to be higher in July than in March. For the Highland, the situation is the opposite. To determine the optimal distances we used these graphics, for instance: if we permit an error of max 20 % the optimal distances for the Coast are 20 and 30 km in March (wet season) and July (dry season) respectively. For the Highland, at the same error, the optimal distances are 25 and 20 km for March and July respectively.

Determination of the Stations' Optimal Density

To determine the optimal density of the stations we put the interpolation error values of every station in a map and with the aid of the isohyets, we defined the territories where it is necessary to grow the number of stations in order to decrease the error degree. The more accuracy in the precipitation data we want the more the number of the station should be. Of course, if the accuracy of the data is given the density of the stations network could be calculated. (Haszpra, 1985). For instance, to obtain an error around the 20% the network in the Coast should be to grow with 15 stations in March and with 8 stations in July (35 and 28). In the case of the Highland with a 20% of error, the network should be to grow with 8 stations in March and with 12 stations in July (45 and 35).

Conclusions

At the present, 120 precipitation stations operate in the Coast and 248 in the Highland. In both cases, the number of stations should be decreased to obtain a small degree of error.

The application of the Optimal Linear Interpolation Method seems to be the best option in the moment of planning the distribution of precipitation stations due to the basis of the error percentages. The Optimal Linear Interpolation Method is a consistent tool that allows overcoming the subjectivity of the traditional application of the empirical method of interpolation of climatic data based on the manual plotting on a topographic map of the place studied.

During the considered period it was possible to appreciate that there is seasonality for the two zones, situation that should be used to run automatic stations and their number will depend on the given time, achieving with this, objectivity and economic saving.

These result cannot be considered finally because the configuration has not taken into account in this paper, it would imply the introduction of new parameters as economics, social and technical elements.

Recommendations

The application of automatized meteorological stations is extremely recommendable to optimize the use of them and justify the obtained results due to the variability of the climatological patterns around the year in the territory of Ecuador.

The inclusion of more parameters as the territorial configuration is advisable to apply in a further investigation.

References

Box, G. a. (1976). Holden – Day. *Time Series Analysis: Forecasting and Control*.

Czelnai, R. D. (1963). On the determination of the the rational density of the temperature – measuring networks. *Idojárás, 67*, 257 - 267.

Gandin, L. S. (1970). The planning of a meteorological stations network. *WMO, Technical Note 111.*, 35.

Gulyás, O. (1980). Matematikai statisztika meteorológusok számara. *OMSZ Kisebb kiadványai*.

Haszpra, L. (1985). On the density and placement of the Hungarian precipitation chemistry network. *Időjárás 89.*, 1 - 8.

Hoffman, J. (1970). La distancia crítica para la interpretación de datos y la reducción de las estadísticas de precipitación al mismo período. *Meteorológica*.

Hoffmann, J. (1971). Características de las series de precipitaciones en la Argentina. *Meteorológica*.

INAMHI. (1976 - 1981). Series climatológicas y frecuencia de lluvia anual de varias localidades del Ecuador. *Anuarios Meteorológicos*.

OMM. (2011). *Guía de Prácticas Climatológicas.* Ginebra: OMM.