Rainfall regionalization Techniques to provide design values for infrastructure: Case of study in a binational basin of South America

Conference Paper · June 2015

Julio Isaac Montenegro Gambini
Universidad Nacional de Ingeniería (Peru)

E-proceedings of the 36th IAHR World Congress
28 June – 3 July, 2015, The Hague, the Netherlands
ISBN: 978-1-5108-2434-8
DOI: 10.13140/RG.2.1.1354.1364

RAINFALL REGIONALIZATION TECHNIQUES TO PROVIDE DESIGN VALUES FOR INFRASTRUCTURE: CASE OF STUDY IN A BINATIONAL BASIN OF SOUTH AMERICA

JULIO MONTENEGRO GAMBINI(1),
(1)Lecturer – National University of Engineering. Lima, Peru.
(1)Researcher, National Hydraulics Laboratory. Lima, Peru.
E-mail: jmontenegrog@uni.edu.pe – jmontenegrog@pucp.pe

ABSTRACT

This research aims to make a Hydrological regionalization in the binational basin of Catamayo-Chira river, one of the most affected basins in South America due to extreme events. We explored the most of existing information, seeking to estimate hydrological variables in lacking data sites where existing information is insufficient in quantity or quality; on the other hand, it is very important we have design values such as rainfall to propose hydraulic infraestructure. This technique allows better exploring in samples and, consequently, improving the estimates of variables checking the consistency of hydrological series and identifying the lack of observation using daily rainfall data provided by governmental institutions of Perú and Ecuador. With the use of statistical methods, hydrological regions are analogous to areal classes. In this present study, we applied to Catamayo-Chira river in order to analyses the patterns of variation and to delimit regions of uniform behavior and making comparisons with the maximum values of rainfall and runoff data using different methods. Regional homogeneity characteristics were evaluated following the Multidimensional strokes method by Andrews (1972) and index-flood method, described by Dalrymple (1960). We perform theoretical probability distributions obtaining the most representative maximum daily rainfall values and to assess several return periods, additionally within each region, the relationship between rainfall and El Niño phenomenon will be examined. The comparison of results between the maximum rainfall values obtained by frequency analysis in each of the rain gauges and by regionalization shows that the methodology is reliable and useful with very good correlation between the Real and the Regionalized values, in order to obtain design values for hydraulic infrastructure. Based on hydrometeorological analysis have been developed isohyets maps and 3D surfaces to assess the spatial distribution of rainfall regions.

Keywords: Catamayo, Chira, Regionalization, Rainfall, index-flood

1. INTRODUCTION

The Catamayo-Chira binational basin (Figure 1) occupies 17,199.18 km2 between Ecuadorian and Peruvian territory, the latter with more extension, in the department of Piura. Starting at the confluence of the rivers Catamayo and Macara, the river takes the name of Chira. Downstream receives inputs from tributaries and streams that are activated in wet seasons. Floods have intensity in a defined space and time, caused by the dynamics of the earth, where no man intervenes, cannot avoid acting as a lateral overflow of river water in times of heavy rainfall. The risk lies downstream of the dam, in the valley of the Chira River to the Pacific Ocean.

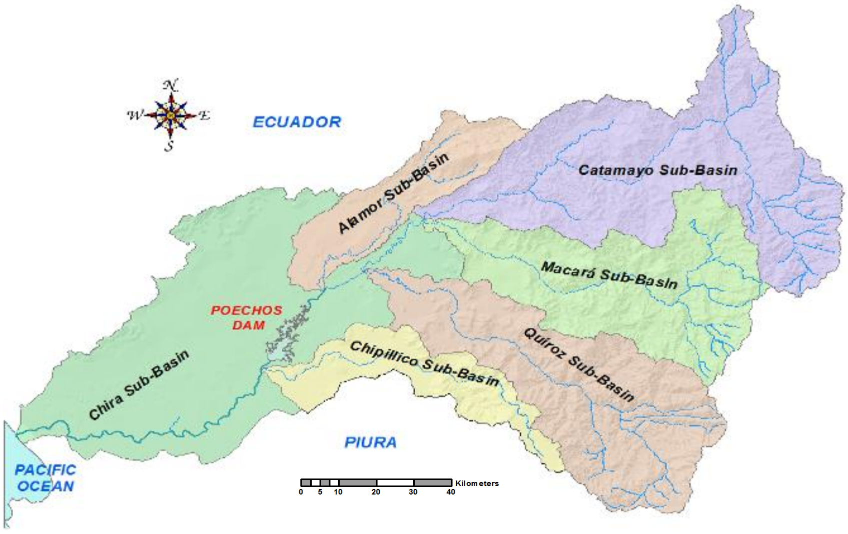
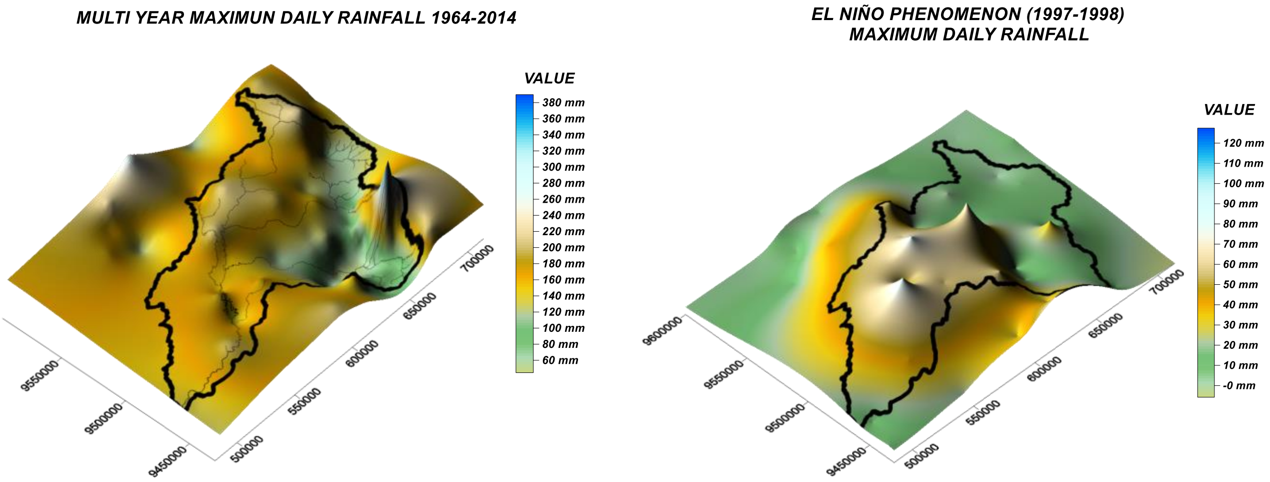


Figure 1. Catamayo Chira Binational Basin and Poechos Dam

The main objective of this study is analyze the behavior of the maximum daily rainfall in different stations with rain gauges in Catamayo-Chira basin, obtaining expressions that allow direct determination of a reliable and secure manner, in order to facilitate the development of hydrological studies and hydraulic design in sites where no it has sufficient amount of hydrological data and/or hydrometric. Additionally we are able to evaluate rainfall records from different gauges from Perú and Ecuador concerning quality, consistency, seasonality, among others and get maps regionalization of Daily Maximum Rainfall for a binational basin.

2. AVAILABLE INFORMATION

The basic information for hydrologic and hydraulic analysis is based on studies and measurements made by the Project Chira-Piura, Piura Regional Government and organizations like SENAMHI, ANA (Perú) and INAMHI, National Water Secretary Office (Ecuador) in the basin with data recorded from 1937 to 2014. The change in the weather conditions show clear tendency to increase with higher intensity rainfall therefore transport liquid and solid runoff over its territory, decreasing return periods of extreme events (Montenegro, 2013). The reservoir will be at risk Poechos to download large volumes that can affect the Chira River valley and cause flooding, damages on many economic and social sectors in the population. As for rain gauges in Catamayo sub basin (Ecuador) are 8 gauges, 9 gauges to Macara (Ecuador and Peru), 2 gauges for Alamor (Ecuador), for Quiroz (Peru) 15 stations, for Chipillico (Perú) 6 stations, for Chira (Peru) to the Pacific Ocean are five 4 stations. There are 19 auxiliar and aditional gauges close to Chipillico and Chira sub basins in Piura region and 5 in Ecuador close to Catamayo Sub basin. They have average monthly precipitation series of up to 50 years. Furthermore, there are inoperative stations that contains records in El Niño Phenomenon and representative extreme events in the basin, according to the quality and density we considered those measures in order to get more amplitude of the behavior.


MULTI YEAR MAXIMUN DAILY RAINFALL 1964-2014

EL NIÑO PHENOMENON (1997-1998)
MAXIMUM DAILY RAINFALL

Figure 2. Rainfall behavior over Catamayo Chira in 50 years and during Niño Phenomenon (1997-1998)

Three-dimensional maps based on geostatistical Kriging interpolation for normal years and El Niño. The graph shows the distribution multiyear upward territorial precipitation with altitude, which changes in El Niño years with higher rainfall in the middle of the basin (Fig. 3). Quantitative estimation of variables and parameters indicative of aspects of the use, management, control and preservation of water resources in river basins is a topic of great interest to the South American region. In this binational basin, we face the difficulty of not having the necessary hydrometeorological information, in quantity and quality, in order to estimate the flow rates to study the waterworks as Poechos dam, make emergency plans, or design new infraestructure. Using mapping regionalization hydrological reduces time that takes hydrological studies, ensuring high reliability in the estimation of the design variable associated with a given recurrence period.

In this area prevailing lack of basic information not only flows but also rainy disaggregated into smaller durations accumulated in a day, therefore the designer engineer chooses to use various synthetic methodologies for deriving the design flow (Q_{Tr}) to hydraulic works in general. The method for hydrologic design (Rodriguez, 2012) of such works is:

1. Have daily maximum rainfall (MDR) and proceed to statistical analysis - probability to determine the function that best fits the data of the sample.

2. Estimate the probable maximum precipitation for different durations and determine frecuency an durations with some approaches in hydrology.

3. Calculate the design flow based on process simulation rainfall – runoff by hydrologic model

It is for this reason that the motivation of this work is to be able to present mathematical expressions for determining the MDR according to spatial attributes (longitude, latitude and altitude). Each region that contains areas that are hydrologically similar, would enable the design engineers can determine the Q_{Tr}, eliminating the step 1 of the method for hydrologic design of hydraulic structures defined above. Likewise determining regions with similar hydrological characteristics enables better management of water resources are there in the sub - region west central South America, for this work MDR was considered as discrimination parameter to identify and verify that each region is hydrologically similar among themselves.

3. METHODS AND ANALYSIS

a) Data Validation:

A data samples is obtained and an assessment of data quality based on statistical and hydrological processes. First, we managed the Independence test Anderson that uses serial autocorrelation coefficient r '(for different delay k. In the case of analyzing a single record, then j = 1. The expression for the coefficient k serial autocorrelation delay is as follows:

$$r_k^j = \frac{\sum_{i=1}^{n_j-k}(P_i^j - \bar{P}^j)(P_{i+k}^j - \bar{P}^j)}{\sum_{i=1}^{n_j}(P_i^j - \bar{P}^j)^2} \quad for\ k = 1,2,\ldots,\frac{n_j}{3}$$

Where:

$$\bar{P} = \sum_{i=1}^{n_j} \frac{P_i^j}{n_j}$$

Furthermore, the limits at 95% confidence j_{kr} can be obtained with the equation:

$$r_k^j(95\%) = \frac{-1 \pm 1.96\sqrt{(n_j - k - 1)}}{n_j - k}$$

The graph of the estimated r_k^j (y-axis) against the delay times k (x-axis), along with their confidence limits, called correlogram of the sample. If no more than 10% of the values r_k^j beyond the limits of confidence, says that the P_i^j series is independent and therefore is a variable that follows the laws of probability. Figure 3 shows one of the results of this test in Catamayo Rain Gauge.

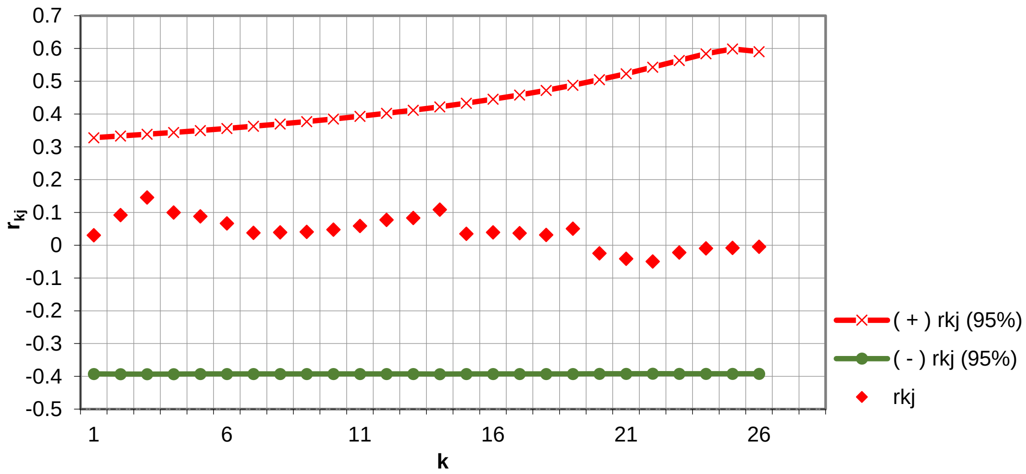


Figure 3. Independence test of Catamayo Rain gauge

Then we analyzed the presence of outliers (extreme values) applying the Mann–Whitney U test (also called the Mann–Whitney–Wilcoxon (MWW), a nonparametric test of the null hypothesis that two populations are the same against an alternative hypothesis, especially that a particular population tends to have larger values than the other. It has greater efficiency than the t-test on non-normal distributions, such as a mixture of normal distributions, and it is nearly as efficient as the t-test on normal distributions. The test involves the calculation of a statistic, usually called U, whose distribution under the null hypothesis is known. In the case of small samples, the distribution is tabulated, but for sample sizes above 20 using the normal distribution is good. Some books tabulate statistics equivalent to U, such as the sum of ranks in one of the samples, rather than U itself.

$$\mathrm{U}_1 = n_1 n_2 + \frac{n_{1,2}(n_{1,2}+1)}{2} - R_{1,2} \quad and \quad \mathrm{U}_1 + \mathrm{U}_2 = n_1 n_2$$

Where n_1 is the sample size for sample 1, and R_1 is the sum of the ranks in sample 1. The smaller value of U_1 and U_2 is the one used when consulting significance tables. Knowing that $R_1 + R_2 = N(N + 1)/2$ and $N = n_1 + n_2$. For large samples (N>20) , U is approximately normally distributed. In that case, the standardized value, mean and standard deviations are as follows:

$$z = \frac{U - U'}{\sigma_U}, \quad U' = \frac{n_1 n_2}{2} \quad and\ \sigma_U = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$$

Data outliers are significantly away from the trend of information and affecting a magnitude substantially of the statistical parameters of the series, particularly in small samples. To detect questionable data, upper and lower for each data set maximum flow stations thresholds analyzed according to the following equations often recommended by Ven Te Chow (1994) were calculated:

$$Y_H = y + K_n * \sigma_y$$
$$Y_H = y - K_n * \sigma_y$$

Where Y_H and Y_L is the upper and lower threshold for doubtful data in logarithmic units respectively, y is the arithmetic mean, σ_y is the standard deviation of the calculated logarithms and K_n is a value obtained in tables of Ven Te Chow (1994) for a normal distribution with 5% of significance level, also calculated as follows:

$$K_n = -3.62 + 6.28n^{\frac{1}{4}} - 2.49n^{\frac{1}{2}} + 0.49n^{\frac{3}{4}} - 0.037n$$

b) Frequency Analysis:

That is to define a common period of analysis of annual maximum daily precipitation in the "n" sites (n = 1,2,3, ..., m) that are within the homogeneous region, theoretical functions are used to adjust the observations and proceed to the selection of the probability function that best describes the study variable (Olmos et al, 2006). Given that the number of stations available with more than 25 years was acceptable records is that the AMS is adopted as a method of analysis that is based on the largest in each full year of registration event.

The Daily Maximum Precipitation data were fitted to a series of theoretical distributions commonly used in hydrologic studies by obtaining software EASYFIT Professional version 5.5, setting ranges of goodness of fit with Smirnov Kolmogorov, Anderson Darling and Chi-squared tests. The determination of maximum avenues was made by HYFRAN hydrologic model that allow notably statistical analysis of extreme events, using tests of homogeneity, independence summarizing the last procedures. Log Normal and Log Pearson III were the suitable distributions in almost all the records after testing, whose density functions are:

$$f(x) = \frac{1}{(x-x_0)\sigma_y\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{\ln(x-x_0)-\mu_y}{\sigma_y})^2} \qquad f(x) = \frac{1}{\alpha_1\Gamma(\beta_1)}\left[\frac{x-\partial_1}{\alpha_1}\right]^{\beta_1-1} e^{-\frac{x-\partial_1}{\alpha_1}}$$

Where: σ_y is the standard deviation, u_y the mean for log normal distribution. For the Log Pearson II we have λ_1, β_1 and δ_1 as the parameters of the distribution and (β_1) is the Gamma function. Figure 4 shows model selection probability distribution, La Esperanza rain gauge as follows:

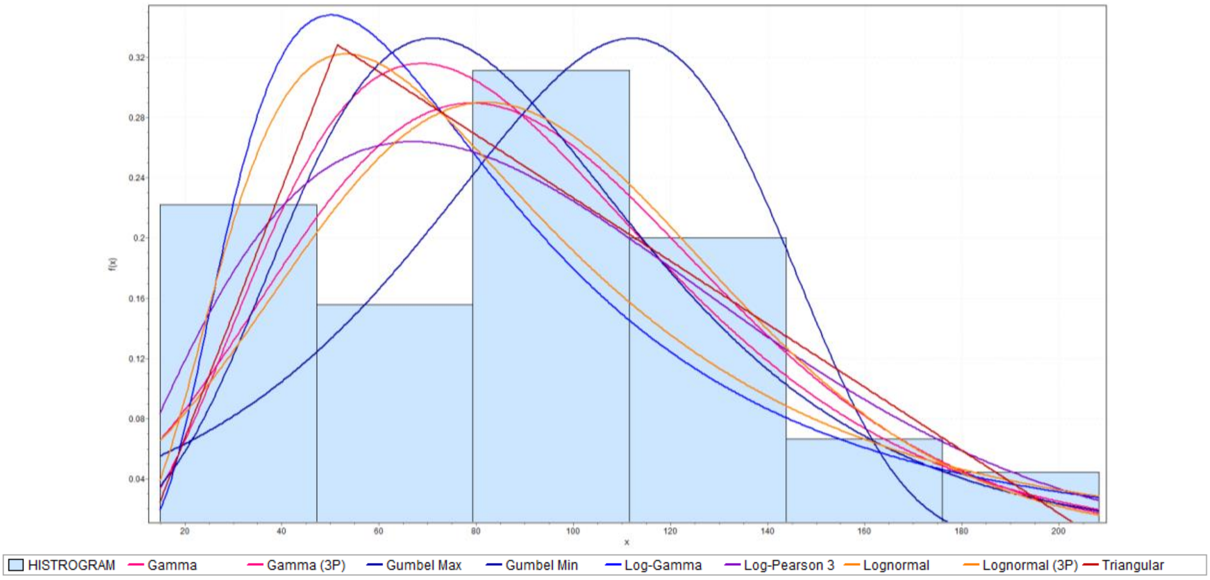


Figure 4. Theoretical distributions over frequency histogram of La Esperanza rain gauge (Piura-Perú)

b) Calculation of flood indexes:

The index flood method, developed at the beginning of 60's, is the most widely used due to its simplicity regional approach. The process begins by defining a common period of analysis of annual maximum daily rainfall in sites j (j = 1,2,3, ..., n) which are within the homogeneous region, after theoretical probability functions are used adjusted the observations, to obtain the maximum daily rainfall (MDR) for the return period of 2.33, which is considered equal to the average annual maximum rainfall. With the help of ε Errors obtained by the difference of measured and estimated rainfall, groups of stations that have the same error or similar ε will form regions.

On the other hand P_{TRn} values for TR = 2,5,10,20,50 and 100 years, so calculating rates or ratios increasing $X_{TRn}= P_{TRn}/P_n$ for all TR and n are estimated; with these values and grouping stations using the diagram shown in Figure 5, the regional frequency curves are obtained.

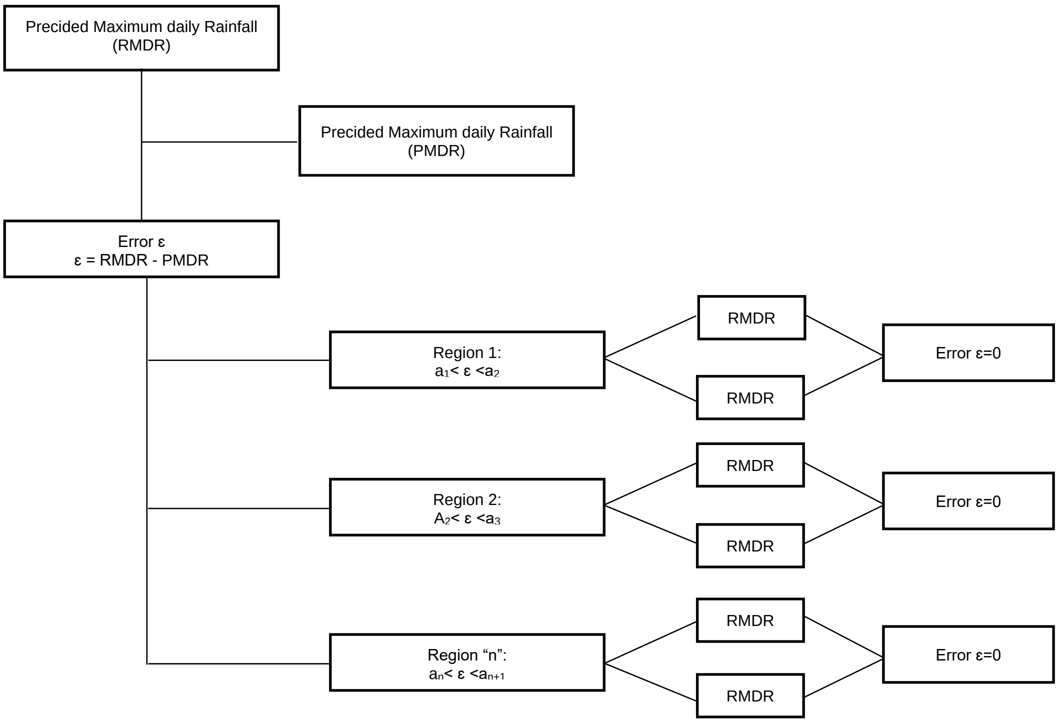


Figure 5. Sequence for selection of regions with similar hydrological characteristics.

b) Multidimensional strokes or Andrews curves:

Nathan and McMahon (1990) developed a technique of regionalization that solves the problems associated with the selection of proper technique clusters, the definition of the homogeneous region and the prediction of group membership that would belong to a new region. The method employs the technique of multiple regression to select the most appropriate geographical and climatic characteristics. The heterogeneity of the groups formed preliminarily evaluated by a technique of positioning proposed by Andrews in 1972, in which a point in the multidimensional space represented by a curve in two dimensions through the function given by the equation:

$$f(x) = \frac{X_1}{\sqrt{t}} + X_2 \sin(t) + X_3\cos(t) + X_4\cos(2t) + \cdots$$

Where X1, X2... are physiographic/meteorological features obtained from regression analysis and the function is evaluated in the range -π ≤ t ≤ π. The fact that this function preserves distances makes it an ideal technique for visual comparison to the formation of homogeneous groups. Clusters of regions with similar behaviors appear as a band of approximate curves from each other.

The fact that this function preserves distances makes it an ideal technique for visual comparison to the formation of homogeneous groups. Clusters of regions with similar behaviors appear as a band of approximate curves from each other. A skill of the method, which is not identified immediately, is that the results obtained depend on the type and order of the selected variables. The first variables are associated with low frequency components cyclic and recent high frequency. The low frequencies are easier to observe, in this way, that which X_1 represent the regression analysis meant more significant from a statistical point of view, the second X_2, and so on.

d) Regionalization of MDR:

We regionalized the Normalized variable P_n is regionalized i.e. the precipitation is 2.33 years of recurrence so that we can estimate the P_{Tr} in each site of interest within the region by the following equation (Rodriguez, 2012):

$$f(x) = (\text{Longitude})^A (Latitude)^B (Altitude)^B (10)^D$$

Where the parameters A, B, C and D are those that characterize each region hydrologically similar.

Then we are able to determine the flood index for each station based on: $X_{TR} = P_{TRN} / P_n$ for all period of recurrence and n sites $P_{2.33} = P_n$ corresponding to a TR = 2.33 years (P_{TRN}: rainfall for 2, 5, 10, 20, 50 and 100 years). Once defined $P_{2.33}$ we can define the possible homogeneous regions with similar hydrologic characteristics. Figure 6 shows the subdivision of the basin using the methodology proposed:

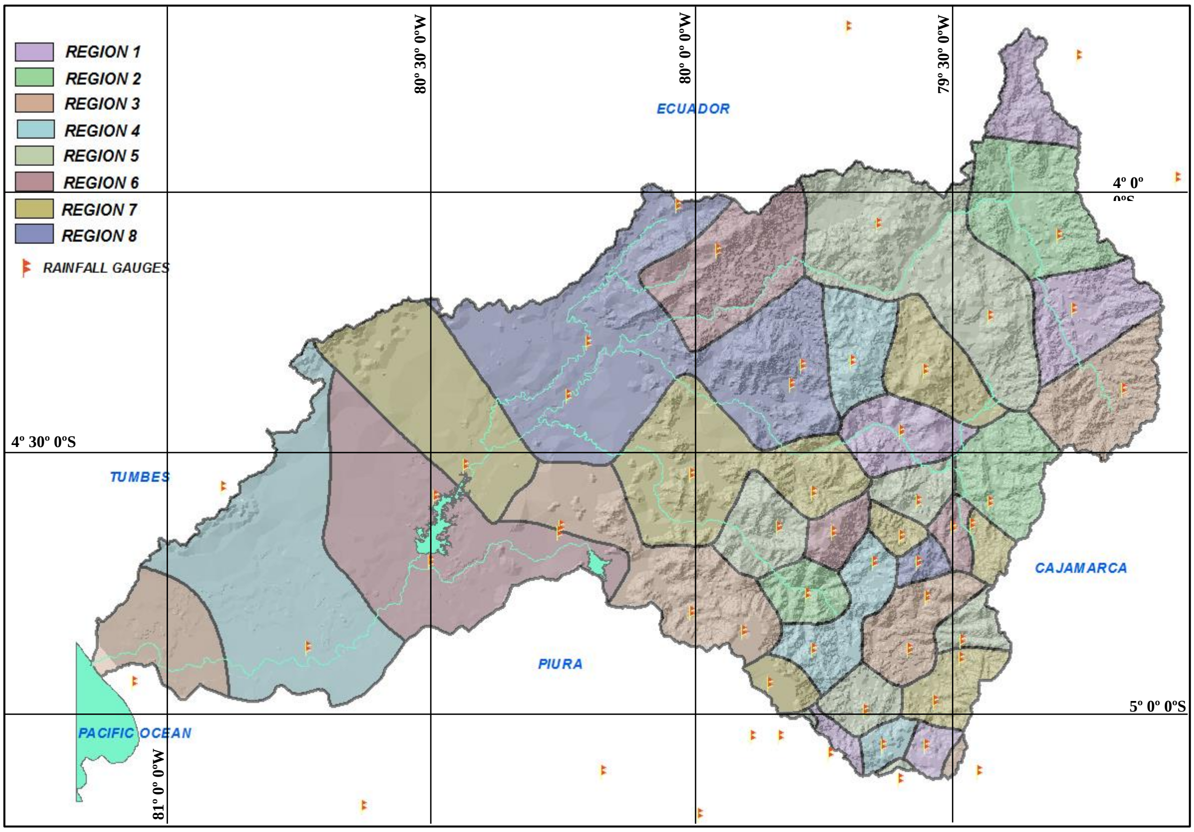


Figure 6. Regions with similar hydrological characteristics in Catamayo-Chira binational Basin

With these sub-regions and the averages of the indices for each period of recurrence, the frequency curve or respective regional flood-index is determined. The general and specific equations are shown below:

$$\frac{P\ (T_R)}{P\ (T_{Ri})} = a - b * Ln\left[-Ln(1-\frac{1}{T_R})\right] \qquad \frac{P\ (T_R)}{P\ (T_{Ri})} = a - b * Ln\left[-Ln(1-\frac{1}{T_R})\right]$$

Table 1. Sequence for regions with similar hydrological characteristics.

HOMOGENEOUS REGION	REGIONAL EQUATION
REGION 1	$\frac{P_{TR}}{P_{2.33}} = 1.12e^{0.311Ln(-Ln\left(1-\frac{1}{TR}\right))}$
REGION 2	$\frac{P_{TR}}{P_{2.33}} = 1.09e^{0.239Ln(-Ln\left(1-\frac{1}{TR}\right))}$
REGION 3	$\frac{P_{TR}}{P_{2.33}} = 1.11e^{0.316Ln(-Ln\left(1-\frac{1}{TR}\right))}$
REGION 4	$\frac{P_{TR}}{P_{2.33}} = 1.06e^{0.229Ln(-Ln\left(1-\frac{1}{TR}\right))}$
REGION 5	$\frac{P_{TR}}{P_{2.33}} = 1.13e^{0.245Ln(-Ln\left(1-\frac{1}{TR}\right))}$
REGION 6	$\frac{P_{TR}}{P_{2.33}} = 1.21e^{0.398Ln(-Ln\left(1-\frac{1}{TR}\right))}$
REGION 7	$\frac{P_{TR}}{P_{2.33}} = 1.11e^{0.251Ln(-Ln\left(1-\frac{1}{TR}\right))}$
REGION 8	$\frac{P_{TR}}{P_{2.33}} = 1.25e^{0.382Ln(-Ln\left(1-\frac{1}{TR}\right))}$

Defined regions within the binational basin, and tested using the method of multidimensional strokes, the method of the "Thiessen polygons" was used to divide the climatic regions. Joining with straight seasons to be analyzed, and then a perpendicular line is drawn and passing through the middle of the lines connecting the stations eventually join each of the intersections and polygons is obtained that delimit the area of action of the weather station.

Considering the graphs determined that according to the order of correlation (maximum and minimum) the variable "Latitude (degrees)" "Altitude (degrees)" and "Elevation (m.a.s.l.)" would have different positions of importance. The parameters that produce the highest absolute values of correlation coefficient are used in the Andrews equation to define the multidimensional strokes, we can show some plots for good understanding as follows:

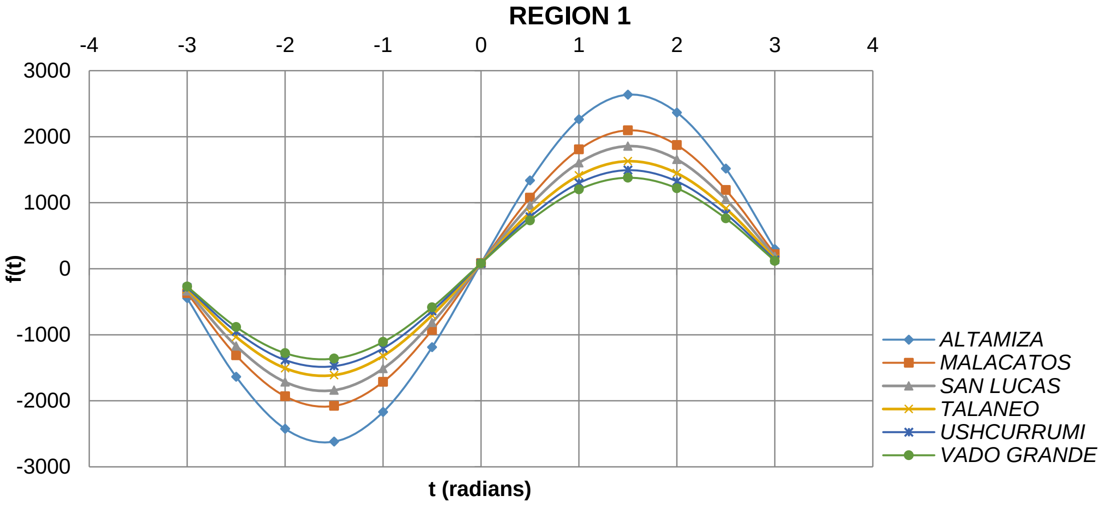


Figure 7. Multidimensional space and Andrews curves for each station in Region 1

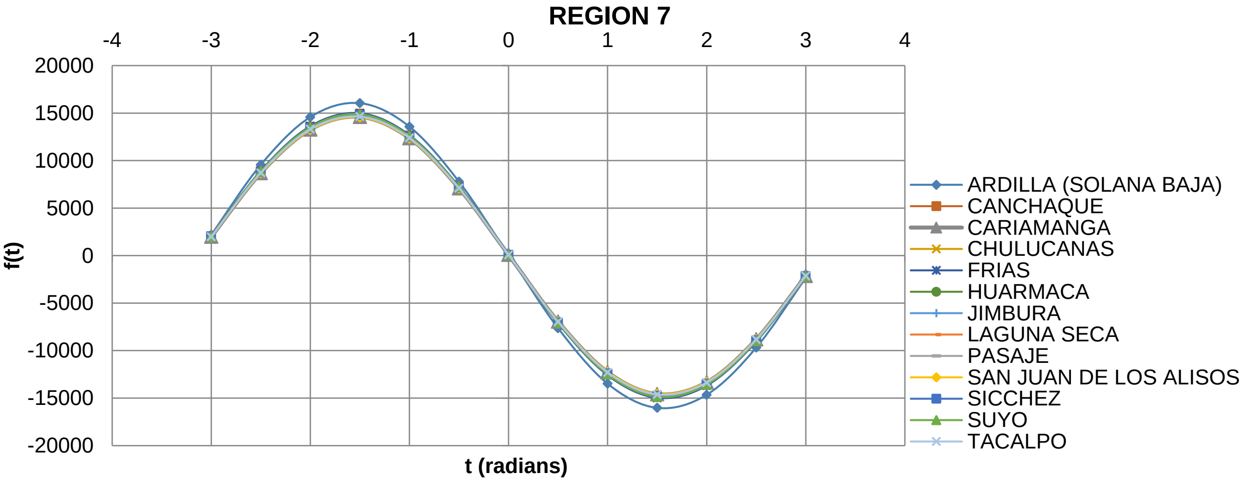


Figure 8. Multidimensional space and Andrews curves for each station in Region 7

In both countries, due to the high costs of installation, operation and maintenance of a meteorological network, especially in sites where there are necessities of design hydraulic infraestructure or make hydrological studies or emergency plans it becomes essential to optimize the available information. Taking into account that Piura, Tumbes and Loja, the main provinces in this basin have the high vulnerability and hazard of El Niño Phenomenon as an inter-annual event, these circumstances have led to the generation of these regions, which make use of information from basins with similar characteristics (Escalante Sandoval and Reyes Chávez, 2005).

So far, we have explored the most of the information existing, looking estimate hydrological variables in data lacking or where existing insufficient quantity or quality places. As mentioned Tucci (1993), regionalization is very useful to explore best spot samples and thus improve estimates of the variables; check the consistency of hydrological series and identify the lack of observation.

Regionalization process aims to evaluate the normalized correlation Variable $P_{2.33}$, i.e. precipitation for recurrence 2.33 years, depending on an attribute associated with the precipitation that has a spatial distribution. Once obtained this relationship, we are able to obtain the maximum daily rainfall of project to any point of interest in the region under study by $P_{TR} = X_{TR}$ ($P_{2.33}$). The expressions used to evaluate the correlation of the normalized precipitation and the results of comparisons are in table 2 and figure 9 as follows:

Table 2. Regional equations of each homogeneous region in Catamayo-Chira binational basin

HOMOGENEOUS REGION	REGIONAL EQUATION
REGION 1	$P_{2.33} = \dfrac{(10)^{139}}{(latit)^{0.26}(long)^{71.61}(alt)^{0.386}}$
REGION 2	$P_{2.33} = \dfrac{(alt)^{0.186}(10)^{94.05}}{(latit)^{2.132}(long)^{48.26}}$
REGION 3	$P_{2.33} = \dfrac{(alt)^{0.157}(10)^{72.85}}{(latit)^{1.908}(long)^{37.09}}$
REGION 4	$P_{2.33} = \dfrac{(alt)^{0.119}(10)^{65.65}}{(latit)^{1.776}(long)^{33.25}}$
REGION 5	$P_{2.33} = \dfrac{(alt)^{0.09}(10)^{54.86}}{(latit)^{1.54}(long)^{23.07}}$
REGION 6	$P_{2.33} = \dfrac{(alt)^{0.108}(10)^{46.23}}{(latit)^{1.484}(long)^{23.07}}$
REGION 7	$P_{2.33} = \dfrac{(alt)^{0.049}(10)^{38.49}}{(latit)^{1.029}(long)^{19.04}}$
REGION 8	$P_{2.33} = \dfrac{(long)^{0.646}(alt)^{-0.032}(10)^{1.206}}{(latit)^{1.037}}$

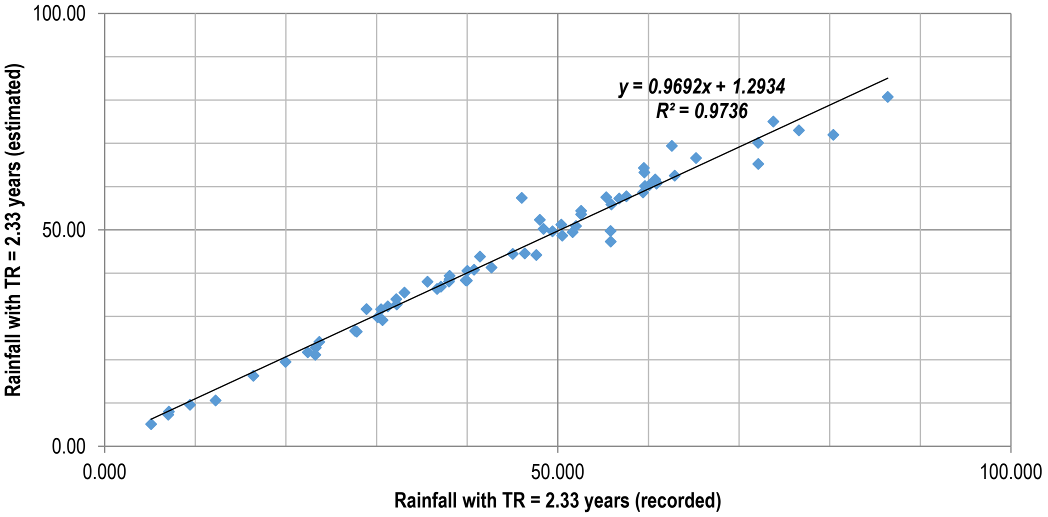


Figure 9. Plot for comparison of recorded and estimated MDR with 2.33 years of recurrence

4. CONCLUSIONS

The expressions shown in the tables are useful for estimating MDR because they have a reliability of 97.36% , which makes these can be used to determine design flows using synthetic methods, obtaining reliable and safe results, which ultimately facilitates the development of hydrological studies including in basins where no count with available records of rain gauges.

The use of test of homogeneity, stationary, independence and avoid the presence of outliers allowed improve records of maximum daily rainfall provided by institutions and institutes in the two countries, tests of independence and homogeneity yields errors of 10% maximum, for sure, successfully passed.

If only we have some variables for describing the similarity between sites, streams or catchments, then the use of scatter plot is necessary for displaying groupings but much more difficult with a larger number of variables. Here we learned and applied a good method of viewing patterns of similarity or dissimilarity across multiple dimensions.

It is important to highlight, with the results presented here, the importance of maximum daily rainfall data in different regions; here the need has been demonstrated that such records are valid in both quality and quantity. From this, it should be noted as the main recommendation of this case of study, the implementation of new stations for precipitation and continuity of those records over the Catamayo-Chira binational basin, affected by extreme events and suffering requirements of meteorological information to design hydraulic infraestructure.

ACKNOWLEDGMENTS

The author is grateful and gives an extensive appreciation for the support to the Ecuadorian Secretary of Water, The Chira Piura Special Irrigation Project and SENAMHI. These institutions supported this research development and provided meteorological information, concluding that this research will be a parapet for future studies and a rich and useful source for studies for the design of hydraulic infraestructure in Piura and Loja provinces.

REFERENCES

Andrews, D. F. (1972). Plots of high-dimensional data. Bell Telephone Laboratories, Murray Hill, New Jersey and Princeton University. Biometrics 28: 125 a 136. United States of America.

Catamayo-Chira Binational Project (2005). Catamayo-Chira, water and territorial characterization. Loja, Ecuador and Piura, Peru.

Escalante Sandoval, C., Reyes Chavez, L. (2005). Statistical technics in Hydrology, 2nd Edition. National Autonomous University of México, Engineering Department, 298p. México.

International Engineering Company - IECO (1968). Development Comprehensive Study of river Tumbes-Piura-Chira. Piura.

Rao, A.R., Hamed, K.H. (2000). Flood Frequency Analysis; CRC Press LLC; ISBN: 0-412-55280-9; 349 p.

Rodríguez P, Olmos L., Kuroiwa J., García M. (2012) Analysis and Behavior of Maximum daily rainfall in the Middle East Region of South America. *XXII National Congress of Hydraulics*. Acapulco, México.