ANALYSIS OF SYNTHETIC DAILY RAINFALLS IN THE BASIN OF THE STREAM CALLED "SAN ANTONIO". VALLEDUPAR – CESAR, COLOMBIA.

Héctor Alfonso Rodríguez Díaz[1], Andrés Humberto Otálora Carmona[2], Romeo Lenin Ramos Quintero[3]

[1]Associate Professor. Escuela Colombiana de Ingeniería Julio Garavito, Bogotá, Colombia. alfonso.rodriguez@escuelaing.edu.co.
[2]Instructor Professor. Escuela Colombiana de Ingeniería Julio Garavito, Bogotá, Colombia. andres.otalora@escuelaing.edu.co
[3]Engineer of Environmental Projects, Drummond Ltd, La Loma, Cesar. rramos@drummondltd.com

ABSTRACT

This paper presents analysis, calculations, results and recommendations about the use of a simplified model for the generation of synthetic rainfall series in the basin: "Arroyo San Antonio" for the exploitation area in the department of Cesar, Colombia. A daily data reanalysis method by Universidad de Chile (2015) was applied in this study.

Taking into account the study area hydrological peculiarities and aiming to minimize differences between synthetic and observed data patterns, the mentioned method was improved. These data generation methods reproduce patterns of each variable's magnitude while maintaining the properties of observed series. The main goal of generating synthetic data is to reproduce specific patterns in order to perform hydrological studies with these series as actually observed data series.

INTRODUCTION

This article describes study results for synthetic data series generation to basins located in southern Cesar department of Colombia. A daily data reanalysis method by Universidad de Chile (2015) was applied in this study. Taking into account the study

area hydrological peculiarities and aiming to minimize differences between synthetic and observed data patterns, the mentioned method was improved.
Stochastic methods for daily data series generation use observed series like statistical sample for the to-be-reproduced hydrological variable. Data generation methods reproduce data with specific patterns, which are later assumed to be future observed series for hydrological analysis. To use these methods, daily data series with a year-length records are required. Due to uncertainty in deterministic random generation, 3 alternative methods have been proposed in this study.

First synthetic generation alternative requires representative daily data for a year; in this case, a 365-record series is built with mean multiannual daily precipitation, that is, the average of daily precipitation for the years when information is available. Once this base series is defined, several series can be obtained by repeating the procedure above.

Second synthetic generation alternative uses every 365 records of years when information is available. Therefore, for each observed year of data, another synthetic data series can be obtained so that original data pattern is maintained, although mean and variance might change.

Third synthetic generation alternative is similar to the former but a scrubbing and optimization procedure is used to maintain a similar mean and variance to the original records.

HYDROCLIMATOLOGICAL INFORMATING COLLECTION

In order to gather representative hydroclimatological information for the study area, available data from the hydroclimatological data network was acquired and revised. Data was collected from ordinary, spatial, pluviographic and pluviometric hydro-climatic stations from the IDEAM (Institute of Hydrology, Meteorology and Environmental Studies – Colombia).

From the collected information, only those stations whose records were long enough were considered. Finally, 27 stations with daily precipitation data for the selected time lapse were chosen. Figures 1 and 2 show daily precipitation data for two of the most representative stations in the study area.

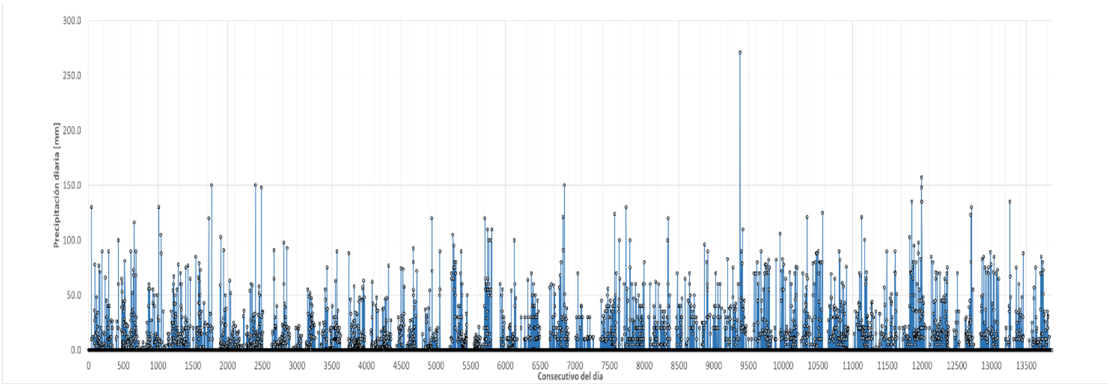

Figure 1. Daily precipitation from 1963 to 2010. La Loma station, observed data.

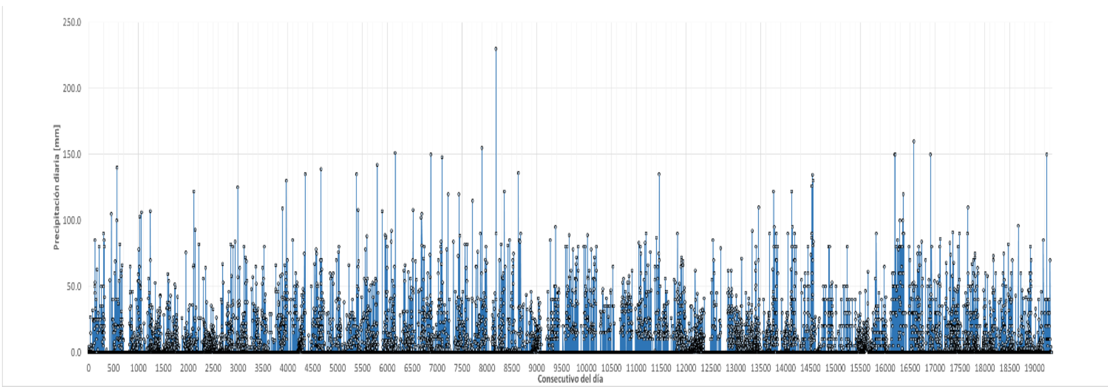

Figure 2. Daily precipitation from 1963 to 2010. La Jagua Station, observed data

LENGTH OF SERIES AND COMPLEMENTED DATA

After an initial refining of obtained pluviometric information, all data series were standardized, completed and extended. This was carried out using base stations with long, homogeneous data series that allowed data complement or correction to other stations. For this process, data mean as well as data trend daily were mostly maintained. To illustrate this, Figure 3 shows completed and extended data series for La Jagua station according to data on the base station.

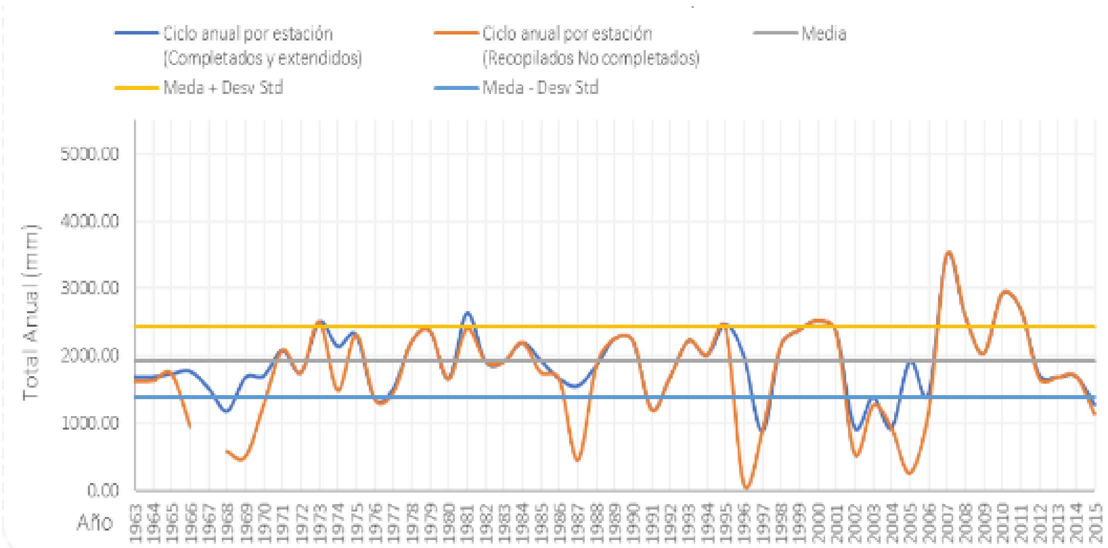

Figure 3. Comparison between hydrologic cycles for daily precipitation. Base station and La Jagua station.

GENERAL DESCRIPTION OF SIMPLIFIED SYNTHETIC RAINFALL GENERATION MODEL

As mentioned before, for daily rainfall generation the data reanalysis method from Universidad de Chile (2015) was applied. This method generates daily data based on monthly statistics for the observed values statistics, preserving data mean and variance.

This method applies the concept of hydrologic persistence, defined as the probability that after two or more consecutive rainy days, there is a high probability of another consecutive rainy day. The former concept is also valid for dry weather conditions. This methodology poses an advantage because it removes negative values; for generating the randomness parameters, Gamma and Normal distributions are used, whose values can be constrained to positive real numbers.

Due to the high amount of information processed and aiming to minimize computational time, a simple script in Wolfram Mathematica was developed. This script allows fast calculation for the described method in yearly data series. Good results that minimize difference between synthetic and observed data are easily achieved with this script.

SYNTHETIC RAINFALL DATA COMPUTATION METHODOLOGY

The theoretical basis of the synthetic rainfall data generation are explained below.

In this method, each day of the year is classified as Rainy (R) or Dry (D). If accumulated precipitation in a specific day is greater than or equal to 0.1 mm, then that day is considered rainy; otherwise, that day is considered Dry. It is necessary that two consecutive days are compared in order to know the possible dry and rainy days combination, and to classify them accordingly: Rainy-Rainy (RR), Dry-Dry (DD), Rainy-Dry (RD), Dry-Rainy (DR).

Monthly statistical parameters of the rainfall series need to be determined, as well as probability distribution of occurrence of Rainy-Dry and Rainy-Rainy combinations, formerly PRD and PRR. Following terms defining this probability distributions are:

$$PRD = \frac{NRD}{ND} \quad [1]$$

$$PRR = \frac{NRR}{NR} \quad [2]$$

Where,

NRD: Number of combinations of Rainy-Dry days in **i-**th month
NRR: Number of combinations of Rainy-Rainy days in **i-**th month
ND: Number of Dry days in **i-**th month
NR: Number of Rainy days in **i-**th month

These two discrete probability functions can be used to generate a synthetic rainfall series according to continuous probability functions Gamma and Normal.

Before using these probability functions, a random number sequence between zero (0) and one (1) should be created in such a way that its trend mimics Normal probability distribution. This procedure becomes easy as some of the features are already implemented in the script and some worksheets. Something to highlight is the fact that Wolfram Mathematica script changes the range for random numbers to minimize error between synthetic and observed series.

After generating normally distributed random numbers, following scheme is used to classify each number as Rainy (R) or Dry (D) according to mean and variance of observed data. Such classification is obtained by comparing as follows the random numbers with the occurrence probability of the combination of RD and RR, known as PRD and PRR:

$$\text{If } x_{i-1} = D \;\; and \;\; randomNumber > PRD \;\; then \;\; x_i = D \qquad [3]$$
$$\text{If } x_{i-1} = D \;\; and \;\; randomNumber < PRD \;\; then \;\; x_i = R \qquad [4]$$
$$\text{If } x_{i-1} = R \;\; and \;\; randomNumber > PRR \;\; then \;\; x_i = D \qquad [5]$$
$$\text{If } x_{i-1} = R \; and \;\; randomNumber < PRR \;\; then \;\; x_i = R \qquad [6]$$

The first random number must be classified by the research scientist, as it is the seed number for computations. Once all random numbers are classified as shown above, Gamma probability function might be used for calculating synthetic rainfall numbers:

$$\text{If } x_i = S \;\; then \; P_i = 0 \qquad [7]$$
$$\text{If } x_i = S \;\; then \; P_i = Gamma\ Inverse\ function\ for\ \boldsymbol{i}th\ month \qquad [8]$$

Where,

P_i: Synthetic rainfall datum for **i**-th month

Gamma probability function is a continuous that depends on occurrence probability of an event x and on Alpha (α) and Beta (β) parameters. Consequently, occurrence probability varying in the range [0, 1] corresponds to the random numbers above, so that:

$$\alpha_i = (\frac{\bar{x}_i}{\sigma_i})^2 \qquad [9]$$
$$\beta_i = \frac{{\sigma_i}^2}{\bar{x}_i} \qquad [10]$$

Where,

α_i Alpha parameter for **i**-th month
β_i Beta parameter for **i**-th month
$\bar{x}_i$ Mean of daily data for **i**-th month
σ_i Sample standard deviation of daily data for **i**-th month

So as to evaluate each alternative and its performance, synthetic series were produced with observed daily precipitation from La Jagua and La Loma stations. As mentioned before, two alternatives were evaluated: the first generation alternative uses a representative data series obtained as daily multiannual means. Second alternative generates a synthetic data series for each observed year of data. As an example, a comparison between first generation alternative and observed data is shown.

Figure 4 presents a comparison between a synthetic rainfall series and original rainfall series for La Loma Station, where the first generation alternative was used. Although daily values for precipitation not necessarily keep the daily base pattern, they do preserve annual statistic properties of observed data.

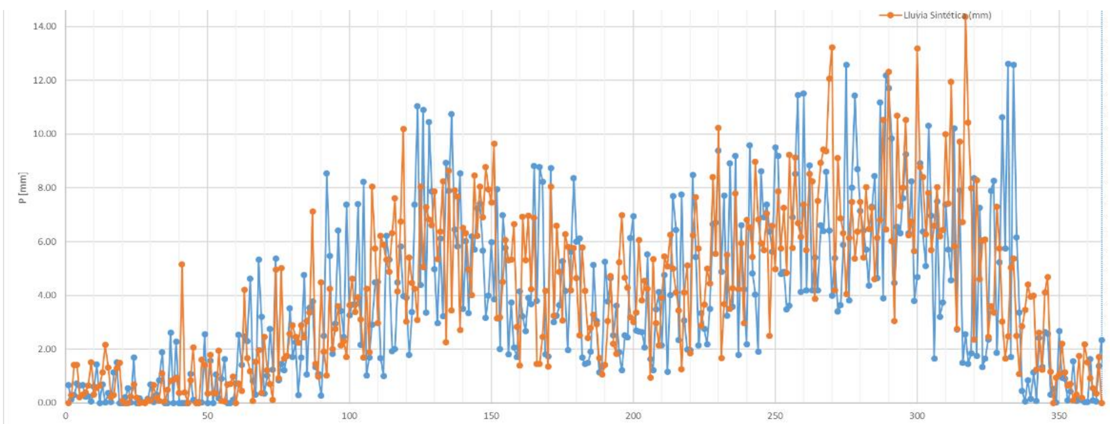

Figure 4. Comparison between synthetic and observed series for La Loma Station

RESULTS OF SYNTHETIC RAINFALL GENERATION FOR THE FIRST ALTERNATIVE

In the first alternative, 50 years of observed precipitation data were available for the study area. Therefore, a representative year for each station was established by using the daily multiannual mean of observed data.

Once this representative series was established, the reanalysis method mentioned above was applied by selecting those series whose differences in mean and variance with generating data series were minimal. Selection was carried over with Wolfram Mathematica and Microsoft Excel by creating a parallel work flow whose purpose was determining best series after 50 simulations for each year on the model.

Figures 5 and 6 show synthetic rainfall series generated with length of 50 years by the first alternative using observed daily multiannual precipitation series from La Loma and La Jagua stations. Precipitation on those figures show an annual cyclic pattern because the method preserves the observed data trend (multiannual averages).

Mean and variance are similar to base daily data series. It can be shown that extreme precipitation values are attenuated due to usage of averages on each day of the generated series, thus biasing the result.

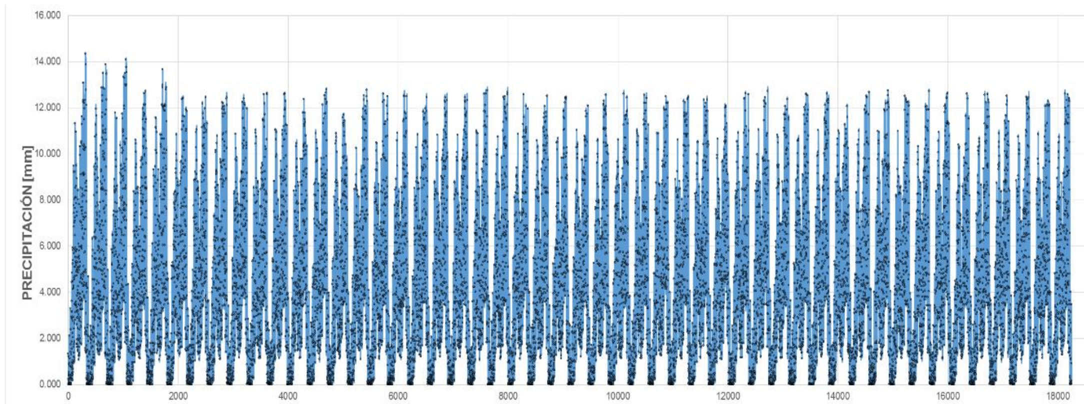

Figure 5. Synthetic rainfall series for 50 years using daily multiannual mean of observed series. La Loma Station

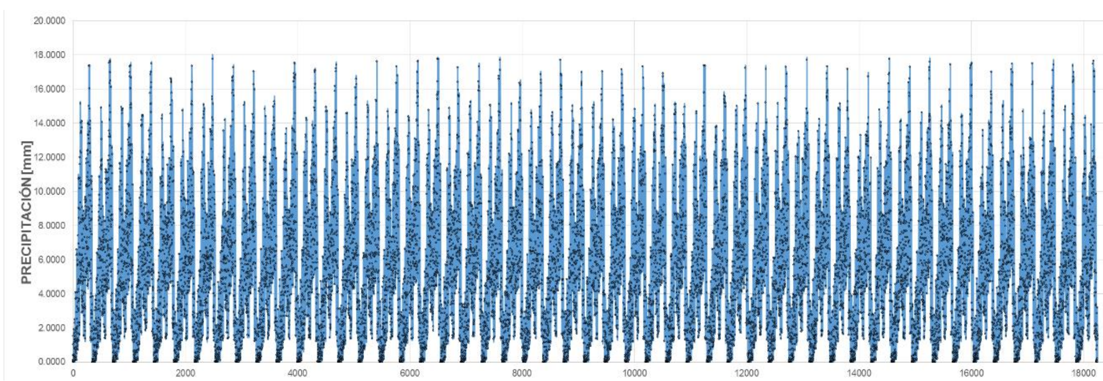

Figure 6. Synthetic rainfall series for 50 years using daily multiannual mean of observed series. La Jagua Station

By using this generated precipitation series on a rainfall-runoff model, specific rainfall data might be undervalued and consequently runoff discharges might be undervalued.

RESULTS OF SYNTHETIC RAINFALL GENERATION FOR THE SECOND ALTERNATIVE

Taking into consideration previous results for synthetic precipitation generation and the observed data behavior, it is concluded that using a single base series might not be recommended due to the attenuation of extreme daily values. As a result, a second alternative for synthetic rainfall generation was developed.

In this alternative, each observed year of data is used for generating a synthetic data series while preserving mean and variance of observed data. For each generated series, two procedures were executed in order to evaluate data sensibility. In the first procedure, synthetic series with a mean higher than the one from observed data. In the second procedure, a upper and lower limits have been defined for generated data in order to achieve a mean to the daily observed data. In this case, synthetic series preserves daily pattern of base series.

Variability on daily precipitation can be analyzed with procedures like the mentioned above, by recreating different conditions without changing statistical properties of the observed series. Moreover, several different rainfall series with similar mean can be generated as wanted by the user, although variance on synthetic data will change. Besides, by rearranging synthetic series, different precipitation trends can be achieved, while keeping the same mean and variance. Figures 7 and 8 show results of both procedures for La Loma station, while Figures 9 and 10 show the same results for La Jagua station.

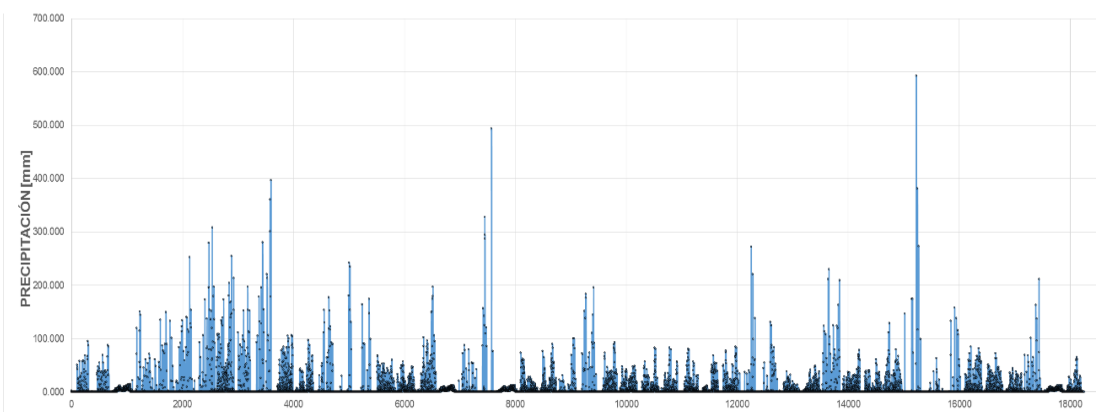

Figure 7. Synthetic rainfall series year by year, preserving annual mean for observed data. La Loma station. First procedure of second alternative

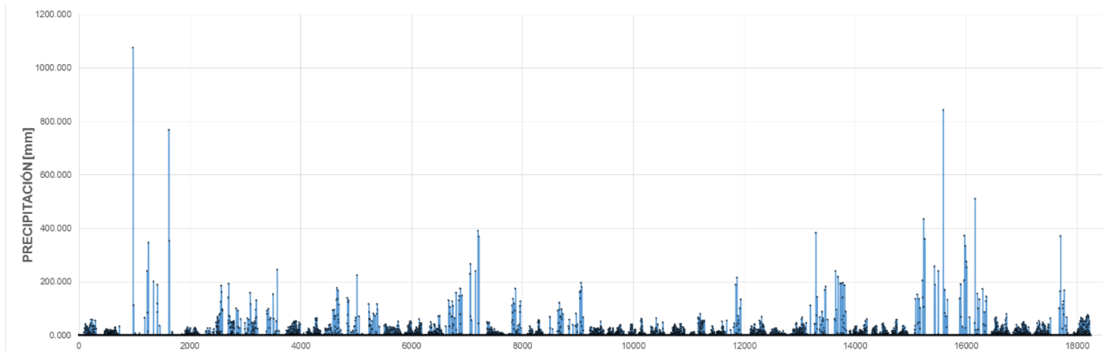

Figure 8. Synthetic rainfall series year by year, preserving annual mean for observed data. La Loma station. Second procedure of second alternative.

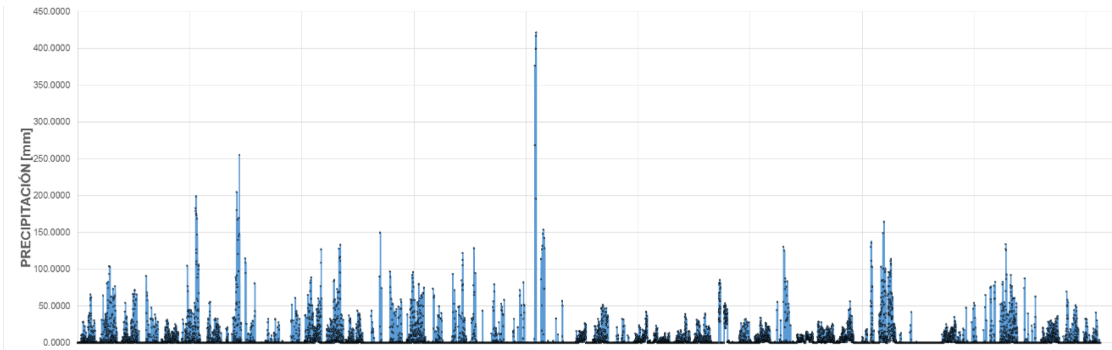

Figure 9. Synthetic rainfall series year by year, preserving annual mean for observed data. La Jagua station. First procedure of second alternative.

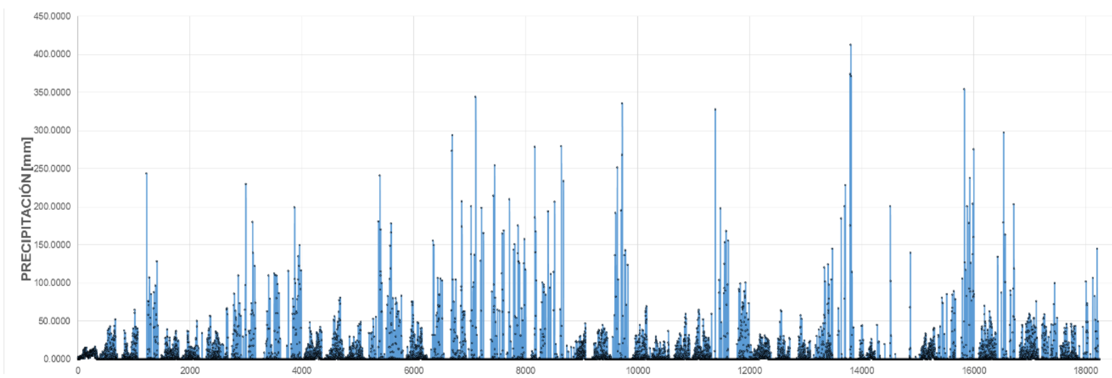

Figure 10. Synthetic rainfall series year by year, preserving annual mean for observed data. La Jagua station. Second procedure of second alternative.

From the former figures, differences in magnitude of daily data can be observed, suggesting that the same observed series can be used for generating several statistically different synthetic series. Such deterministic and random model properties might be used for creating different scenarios for the same rainfall event with similar magnitude. This opens up the possibility for calculating an envelope for runoff and other hydrological results that depend on rainfall. Key step of this procedure is to choose synthetic events that reproduce the observed series.

Besides the variability of results in generated series, base series strongly influence method outcome. It is possible to generate the synthetic data for each observed datum, or generate several synthetic data with an user-selected representative series.

CONCLUSIONES Y RECOMENDACIONES

The selected model can generate daily synthetic series by using statistical parameters of observed series. Depending on how this synthetic series is generated, extreme values of the observed series can attenuate, especially if daily multiannual averages are used for that purpose.

According to the performed procedures and alternatives, synthetic series should be generated with each observed year of data in order to preserve extreme values.

Evaluation of statistical parameters of synthetic date series is necessary, especially if extreme events occur during the observed data series.

Limits of classification for Rainy and Dry weather conditions can be established according to properties of hydrological cycle in study area.

Based on the results and analysis, every synthetic daily data series can be modified according to hydroclimatological peculiarities.

BIBLIOGRAPHY

Olave Cuadra, Patricio E. and Vargas Mesa, Ximena (2015). “Generación de montos diarios de precipitación a partir de series observadas rellenada con datos de reanálisis” *Tesis de Pregrado*. Santiago de Chile, Chile.

Escuela Colombiana de Ingeniería Julio Garavito. (2018). “Caracterización del área de influencia (componente hidrología e hidráulica) requerida para la modificación de la licencia ambiental del proyecto la loma para el periodo 2019-2048”. Centro de Estudios Hidráulicos. Bogotá, Colombia.

Félix Francés, Rafael García Bartual and et al. (2010). “Diseño hidrológico de alto período de retorno mediante generación sintética de eventos” Tesis de Pregrado. Instituto de Ingeniería del Agua y del Medio Ambiente (IIAMA).

UPC-GHS (2009). “Generación de series futuras de precipitación y temperatura”. Universidad Politécnica de Cataluña.

García Gutiérrez, Francisco Pablo (2010). “Modelación hidrológica estocástica: desarrollo de un modelo de generación sintética de series temporales”. Universidad Autónoma Gabriel Rene Moreno. Santa Cruz – Bolivia.

Yevchevich, V. (1972). “Stochastic Processes in Hydrology”. Water Resource Publications. Fort Collins, Colorado, USA.

García Gutiérrez, Francisco Pablo (2010). “Modelación hidrológica estocástica: desarrollo de un modelo de generación sintética de series temporales”. Universidad Autónoma Gabriel Rene Moreno. Santa Cruz – Bolivia.

Yevchevich, V. (1972). “Stochastic Processes in Hydrology”. Water Resource Publications. Fort Collins, Colorado, USA.