MEAN AIR TEMPERATURE-PRECIPITATION MODELS USING GENETIC PROGRAMMING AND ITS APPLICATION UNDER CLIMATE CHANGE SCENARIOS

Ing. Gonzalo Daniel Mejía Santana, Dra. Maritza Liliana Arganis Juárez, M. I. Eliseo Carrizosa Elizondo, Ing. Jacquelinne Mariles Gutiérrez Ing. José Carlos Ramírez García, M.I. Nikte Ocampo Guerrero, M.I. Margarita Preciado Jiménez, Dra. Jeanette Zambrano Nájera.

Instituto de Ingeniería, Universidad Nacional Autónoma de México. Circuito Escolar S/N, Edificio 5,
Ciudad Universitaria, Del. Coyoacán, México D.F., México. C.P. 04510

Abstract: The problem of obtaining equations that allow, not only to reproduce the history of monthly rainfall event as a function of air temperature, but also to predict behavior in the near future, will give decision-makers in the field of civil protection time for decision-making in terms of prevention and mitigation of damage to a meteorological phenomenon such as precipitation. Genetic programming was successfully applied to get such models. Later those models were applied to new data obtained from scenarios of climate change, in some stations models were able to reproduce the annual rainfall behavior pattern.

Introduction

The study of the way in which climatological variables have been or will change over time in relation to greenhouse gas emissions and Climate Change has been the reason for numerous investigations since the end of the 19th century, which have increased since the eighties of the 20 century (Allerup et al. (2000), Piani et al. (2010), Chistensen et al. (2008), Tuomenvirta (2001), Berget et al. (2005), Brunetti et al. (2005), Cahalan et al. (1996), Cho et al. (2001). Strack et al. (2008), De Lima et al. (2012)). Evolutionary computation, genetic algorithms, genetic programming and other bioinspired algorithms have been booming in their use in engineering problems since the last two decades of the 20st century.

It is known that the average temperature on the Earth's surface has increased during the last century, as well as the amount of precipitation that has occurred. Usually precipitation in the north (above 30 ° N) to 70s, and has since declined. Decline trends have also been observed in the tropics since the 1970s. As temperature and rain patterns vary from region to region, many countries have started their own records.

Most of the studies analyzed analyze annual conditions, however, changes in daily peaks in temperature and precipitation have greater social, economic and environmental impacts. This and a thousand reasons more demonstrates that it is necessary, a poor and qualitative study of these climatological variables, in diverse points of our planet, and to establish more precise methods of calculation that show good results.

The Intergovernmental Panel on Climate Change (IPCC) defines climate change as "a statistically significant change, either in average climatic conditions or in its variability, which is maintained for an extended period (typically decades or longer)". Climate change may be due to natural internal processes or external forcing, to lasting changes in the composition of the atmosphere or changes in land use resulting from human activities. The Framework Convention on Climate Change (UNFCCC), in Article 1, defines climate change as "change attributed directly or indirectly to human activity that alters the composition of the global atmosphere climate, and that is in addition to natural variability climate observed over comparable time periods. The UNFCCC thus makes a distinction between "climate change" attributable to human activities altering the atmospheric composition, and "climate variability" attributable to natural causes. (Magaña, 2006).

As a developing country, Mexico tends to be more vulnerable to many developed countries to climate change. IPCC projections and other groups of scientists dedicated to the analysis of the impacts of climate change suggest that even small increases in temperature, climate change could lead to serious negative impacts on various sectors, mainly those related to water resources. The climate change scenarios are "coherent, internally consistent and plausible description of a possible future state of the world". They are not forecasts, as each stage is an alternative of how to involve the future climate. A projection can serve as source material for a stage, but the scenarios usually require additional information; for example, conditions emissions of greenhouse gases or a base scenario. A set of scenarios is adopted to reflect, in the best way possible, the range of uncertainty in projections.

Socio-economic scenarios can be constructed as it has done in the IPCC Special Report on Emissions Scenarios. These scenarios were conducted to explore the future development of the global environment, with special emphasis on the production of greenhouse gases. The emission of these gases into the atmosphere depends largely on the level of development of countries in the future, its population and the use of oil as a main source of energy supply. To speak of scenarios, you must first know the terminology:

- Evolutionary line: narrative description of a scenario (or family of scenarios) that highlights its main characteristics, relationships between key driving forces and the dynamics of their evolution.

- Scenario: projections of future potential, based on a clear evolutionary line logic and quantified.

• Scenario Family: Scenarios that have a similar evolutionary line with respect to their demographic, social, economic and technological change characteristics. The series of SRES scenarios consists of four families: A1, A2, B1 and B2.

In the IPCC Special Report on Emission Scenarios four storylines (A1, A2, B1 and B2), where the driving forces described in emissions of greenhouse gases and aerosols and their evolution during the twenty-first century were developed both globally and in different regions. Each line represents a level evolutionary divergent development in demographic, social, economic and technological issues.

In simple terms, the four storylines combine two sets of divergent trends: a series develops variations between economic and environmental values; the other series explores the variations between increased globalization and regionalization. These storylines can be summarized as follows.

- High emissions A1B
- Media Releases - High A2
- Average emissions - Low B2
- Low emissions B1

Problem solving and relevance of research

The population growth in the medium and short term mean greater urbanization with the consequent change in land use in a country; civil protection measures to be implemented will be essential to alert and protect future populations to the occurrence of extraordinary weather events. In different countries it has units responsible for giving notice to the populations in these cases. Currently there are models that allow to forecast in the short term of the possible occurrence of a phenomenon, for example, precipitation or volcanic alert or arrival of a hurricane to a site, allowing safeguard the population. In particular knowledge of the behavior of precipitation historically occurred on a site and prognosis of it in the long run from variables relative simplicity of measurement (such as the temperature) is very useful to be able to take action preventive. Rainfall forecast with different climate change scenarios are useful for obtaining new forecasting models that include the history and predicted a first approximation of new series with longer than the historical record length.

Within the evolutionary computation genetic algorithms and genetic programming are tools that help to obtain parameters proposed models or obtain completely new models to shape which can reproduce the pattern of behavior of a dependent variable as a function of n independent variables.

Considering new data from climate change scenarios assume research brings a new component with respect to existing forecast models.

Study site.

The IMTA SEDEPECC (IMTA, 2016) platform uses different climate change scenarios, so their records were taken as the basis for this project; a drawback is that the platform does not provide a convenient way for each weather station of the database CNA accuracy, the platform only throws site data with an accuracy of 0.5 °, and established the coordinates is limited to one type of scenario A1B for a period from January 2010 to January 2060, with data for the Normal or averages for each year, data from temperature and intensity of rainfall is collected.

In the Table 1 the actual and approximate location regarding the platform SEDEPECC data, which information to work this step down is presented.

Table 1. Location of sites near stations analyzed according to the SEDEPECC platform.

			LATITUDE, °		**LENGTH, °**	
STATION	**NAME**	**STATE**	**Real**	**P/PGalileo**	**Real**	**P/PGalileo**
25172	San JoaquIn	Sinaloa	25.6678	25.5	-108.02	-108
22043	Coyotillos	Querétaro	20.5992	20.5	-100.2097	-100
7343	Cuauhtémoc	Chiapas	16.7617	17	-92.9228	-93
26022	Colonia Morelos	Sonora	30.825	31	-109.2217	-109
31027	Santa Elena	Yucatán	20.3275	20	-89.6394	-89.5
30041	Chicontepec de Tejada (SMN)	Veracruz	20.9933	21	-98.1639	-98
16228	Acahuato	Michoacán	19.1478	19	-102.3325	-102
12142	Acapulco de Juárez (SMN)	Guerrero	16.8664	17	-99.9056	-100
3039	Ojo de Agua	Baja California Sur	26.3242	26	-111.9847	-112
5147	Ejido Primero de Mayo	Coahuila	27.2292	27	-101.2275	-101

In the figure 1, is shown the location of weather stations and separation CNA Site studying the IMTA SEDEPECC.

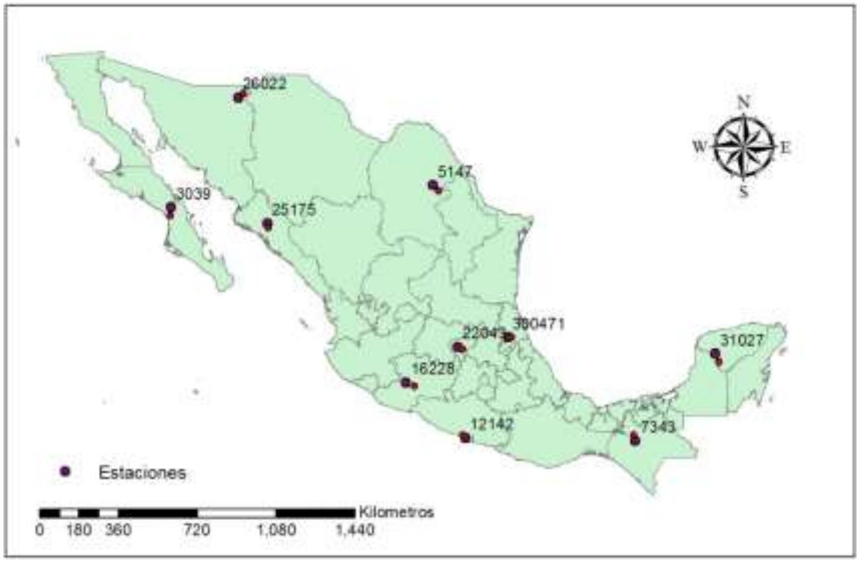

Figure 1. Location of weather stations CNA and site modeling SEDEPECC IMTA

Methodology

10 weather stations that have daily records of rainfall and air temperature were selected; monthly data were obtained from these records. In order to obtain this monthly data, the daily precipitation indexes were accumulated and simultaneously the maximum temperatures and minimum daily temperatures were averaged, according to their corresponding month.

Several correlation analyzes were performed between rainfall and maximum temperature historical data, as well as between rainfall and mean temperature, which resulted in a higher correlation in the latter combination.

Because of the arrangement and distribution of the data, the maximum value representative of each month of rainfall was considered, as well as the mean air temperature values corresponding to this year would complement the data of the month under analysis.

But as in any system there are information losses due to technical failures, either due to some extreme event, fall of the system or maintenance of the same station, for which it was necessary to resort to adjacent information, To the information of adjacent or nearby stations, with the necessary care that the complementary stations did not exceed 20 km of separation and in conjunction with a correlation analysis, a complete record was integrated. For the missing data, the climatological aspects of precipitation, maximum and average temperature were monthly averaged, then the months with empty data were identified, then, according to the identified month, the climatic characteristic that had a higher correlation was searched with the base station, complementing the missing month with the characteristics of the same station with the highest correlation.

Once the information of the twelve months has been concentrated, the database is exported to a format that a genetic programming (GP) algorithm codified in MATLAB© recognizes and performs the appropriate calculations. Subsequently said program gave a solution file with a prefix notation which also comes with operators own the program, the task of converting it to infix notation was a purely careful work.

The genetic programming algorithm includes the establishment of the independent variables and the dependent variable in the problem, operators and constant vector to be considered for the construction of the models to be tested must also be defined. It should provide a probability of exchange or cross (crossing) of the best individuals (set of selected operations) and a probability of mutation must be given. A number of generations (iterations) is proposed to finish the optimization process. In this study objective function consisted in minimizing the mean square error between the measured and the calculated rainfall data with the models tested by the GP algorithm.

The GP algorithm starts with the random generation of an initial population of n individuals (each individual corresponds to a mathematical model consisting of different operators, variables and constants), individuals are then evaluated in the objective function and the best ones are selected (selection can be performed by obtaining a relative frequency of the result obtained by each individual in the objective function divided by the average value given by all tested individuals), individuals with higher relative frequency can be used more than once to be exchanged or crossover, and mutation may also create new individuals and the individuals with lower performance are eliminated and no longer enter the exchange process and / or mutation; so that the new population is again size n. The new individuals are again tested on their performance, selected and the best ones creates new individuals who pass to the next generation, this process is repeated until the number of generations or iterations is reached and the best individual in the last generation will be the one with higher performance and represents the optimal mathematical model found.

In this case the set of arithmetic operators TS = [+,-,*] was considered, a vector of constants obtained randomly, as independent and dependent variables the air temperature T and the rainfall (hp) were set. Populations of 300 individuals (models) of 25 nodes (consisting of operators, variables and constants), a crosses a probability of 0.9 and mutation probability of 0.05 were assumed; finally, 10,000 generations to finish the process were established. The objective function was to minimize the mean square error between rainfall data measured and calculated by each genetic programming model.

Another way to evaluate the obtained results is by drawing the relationship between the monthly rainfall (hp) values (representative year) and the hp values calculated by the PG as a function of temperature against an identity function.

For the next stage, monthly data of the representative year with similar characteristics (in their magnitude and by considering the time of year) were grouped to obtain new GP submodels.

Later the IMTA SEDEPECC platform was applied for this study by considering the nearest sites to the known stations with such coordinates were selected and it was limited only to the climate change scenario type A1B for a period from January 2010 to January 2060, with data of temperature and intensity of rainfall collected for the normal or averages for each year.

Results

The equations obtained previously with historical data were tested with temperature data extracted from the SEDEPECC platform under the climate change scenario, like new data in an attempt to investigate the behavior of the PG models.

So GP models were used to calculate rainfall based on the temperature of each stage, taking this analysis for a period of 50 years from January 2010 to January 2060.

Then the results are presented for each of the stations:

- State: Sinaloa
  Name: San JoaquIn
  Municipality: Sinaloa
  Key: 25172

In Figure 2 data SEDEPECC, according to the analysis, the data obtained greater correspondence with cast by the SEDEPECC were calculated by Horizontal dissection which is shown graphically observed Figure 3 on the same graph to expose the above.

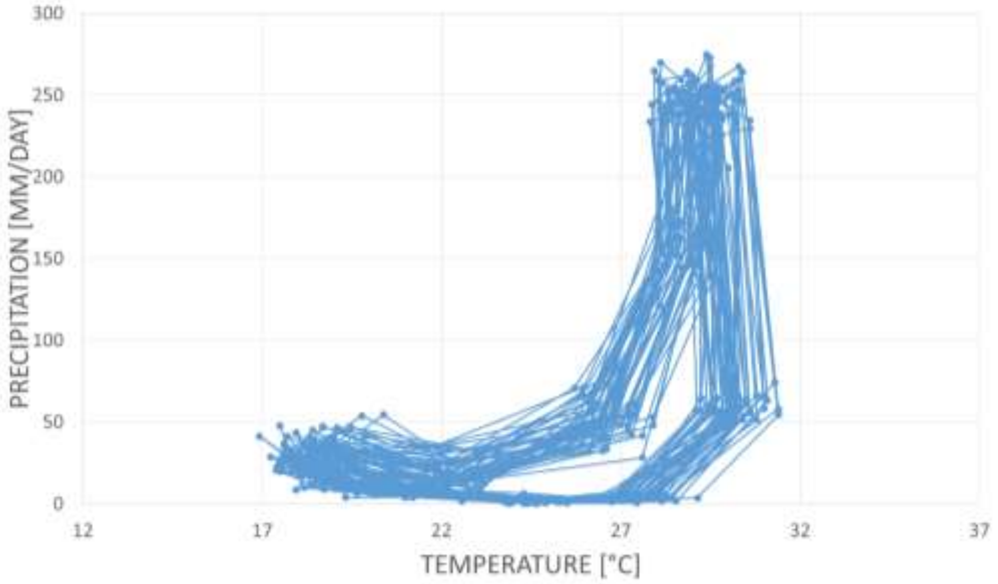

Figure 2. Rainfall against temperature, data SEDEPECC

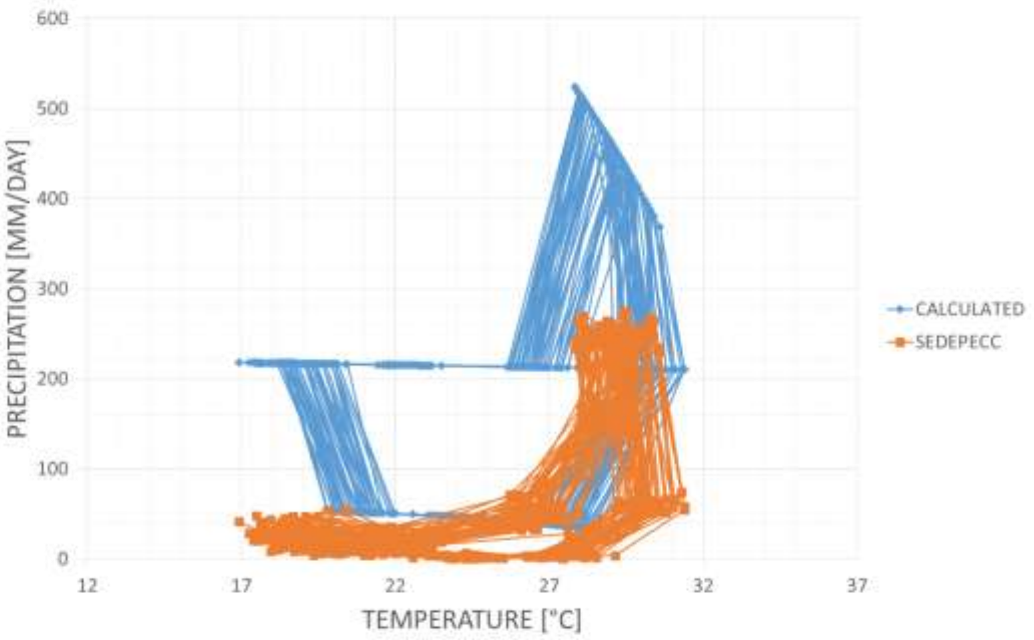

Figure 3. Rainfall against temperature, with adjustment formula overlaying the results recorded in the SEDEPEEC platform.

From the previous form the following 9 stations are presented

- State: Querétaro
  Name: Coyotillos
  Municipality: El Marques
  Key: 22043

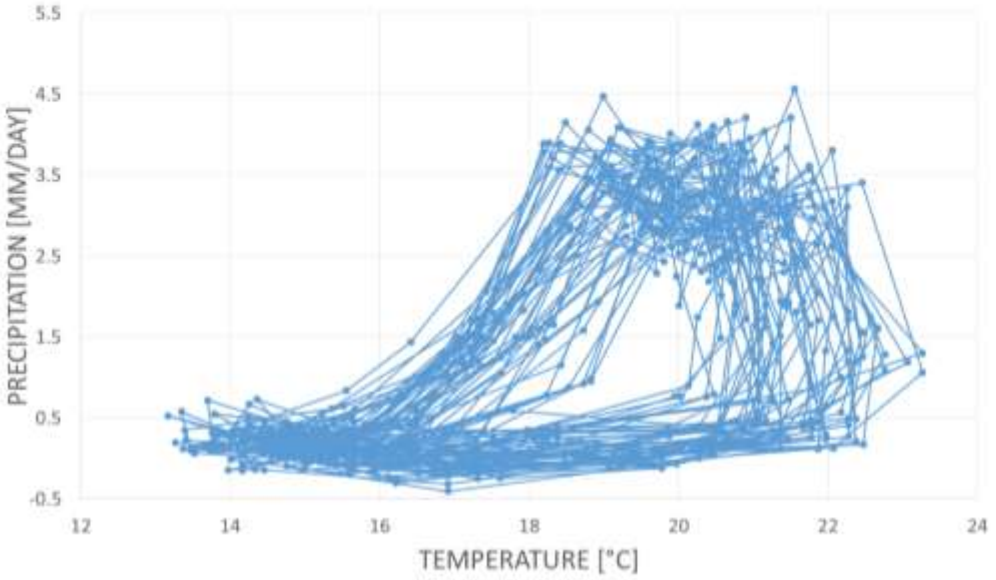

Figure 4.Rainfall against temperature, data SEDEPECC

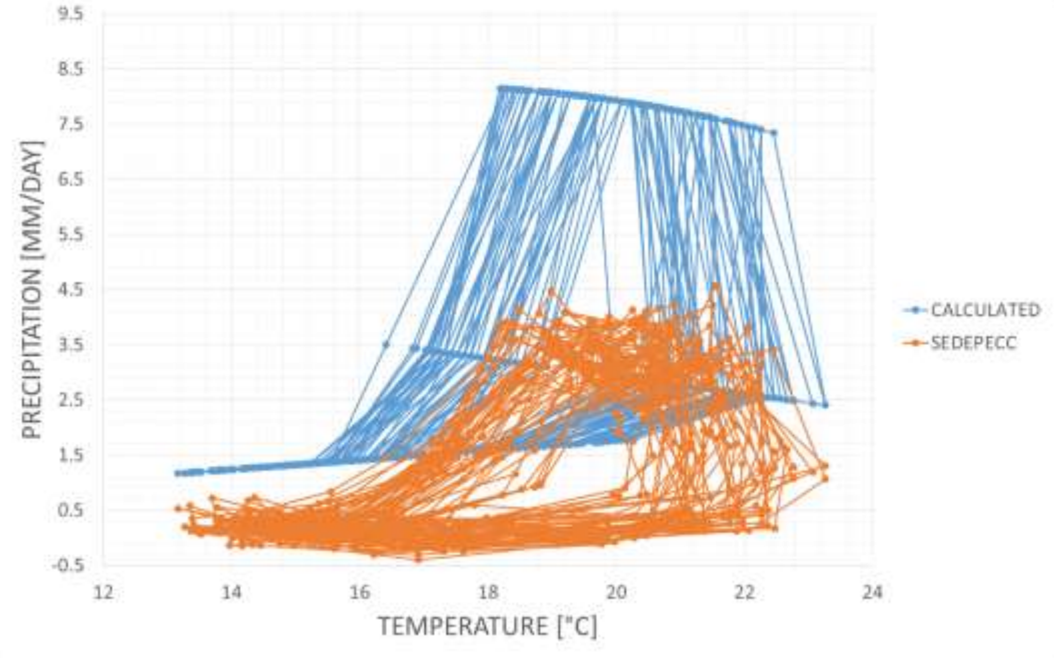

Figure 5. Rainfall against temperature, with adjustment formula overlaying the results recorded in the SEDEPEEC platform.

- State: Chiapas
  Name: Cuauhtémoc
  Municipality: Ixtapa
  Key: 07343

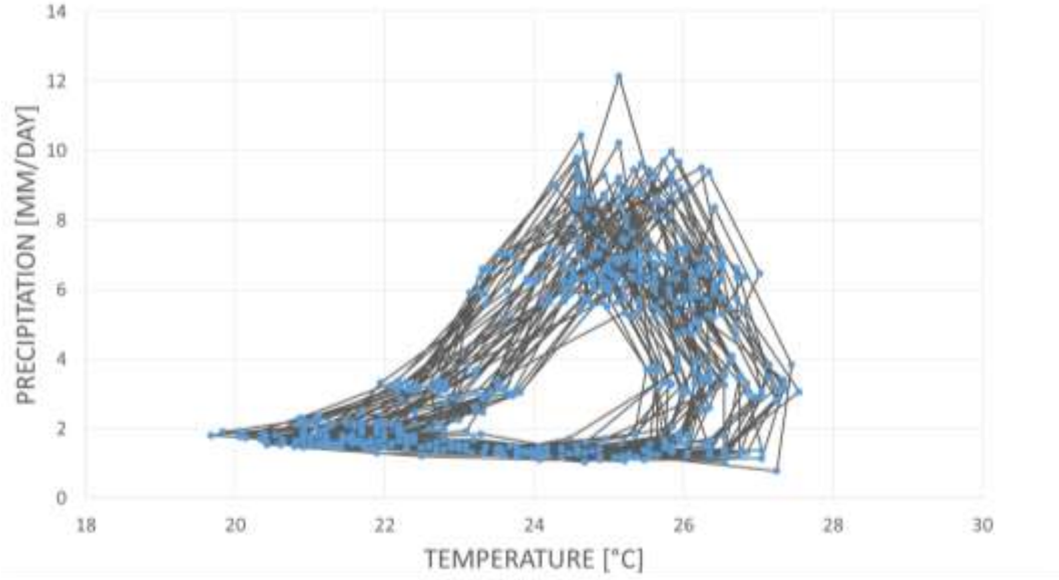

Figure 6.Rainfall against temperature, data SEDEPECC

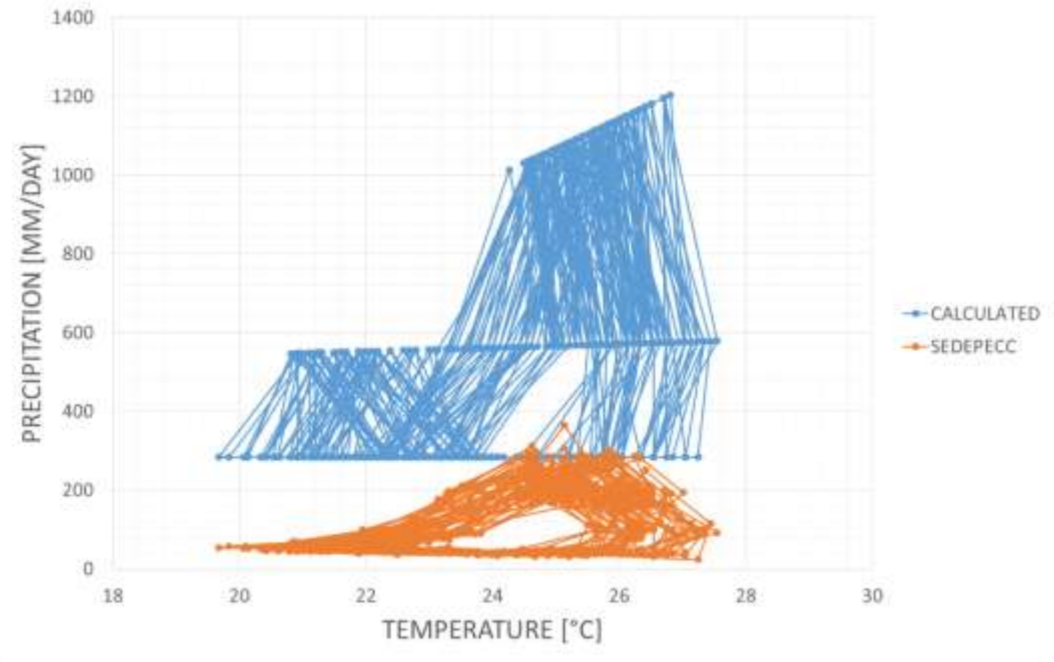

Figure 7. Rainfall against temperature, with adjustment formula overlaying the results recorded in the SEDEPEEC platform.

- State: Sonora
  Name: Colonia Morelos
  Municipality: Agua Prieta
  Key: 26022

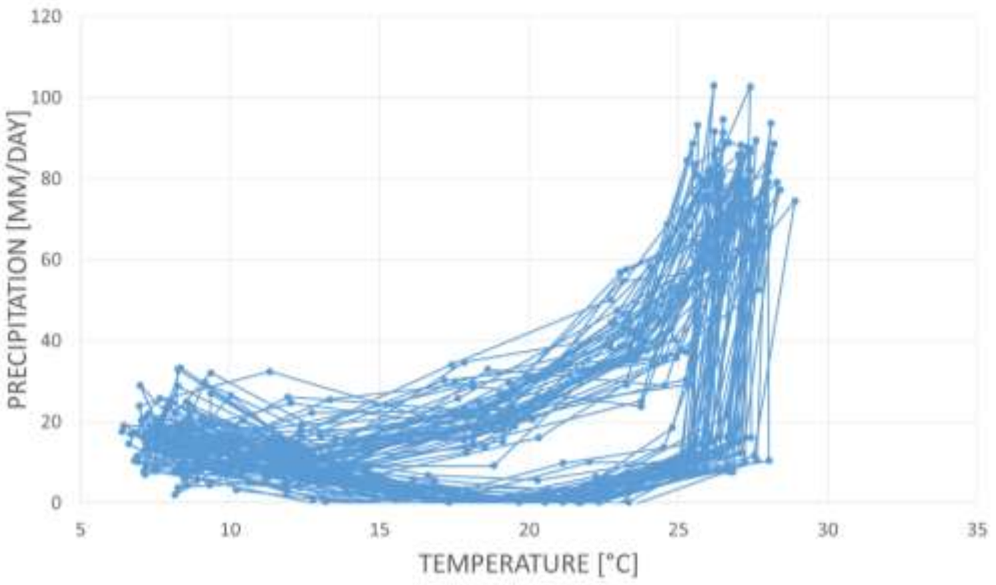

Figure 8.Rainfall against temperature, data SEDEPECC

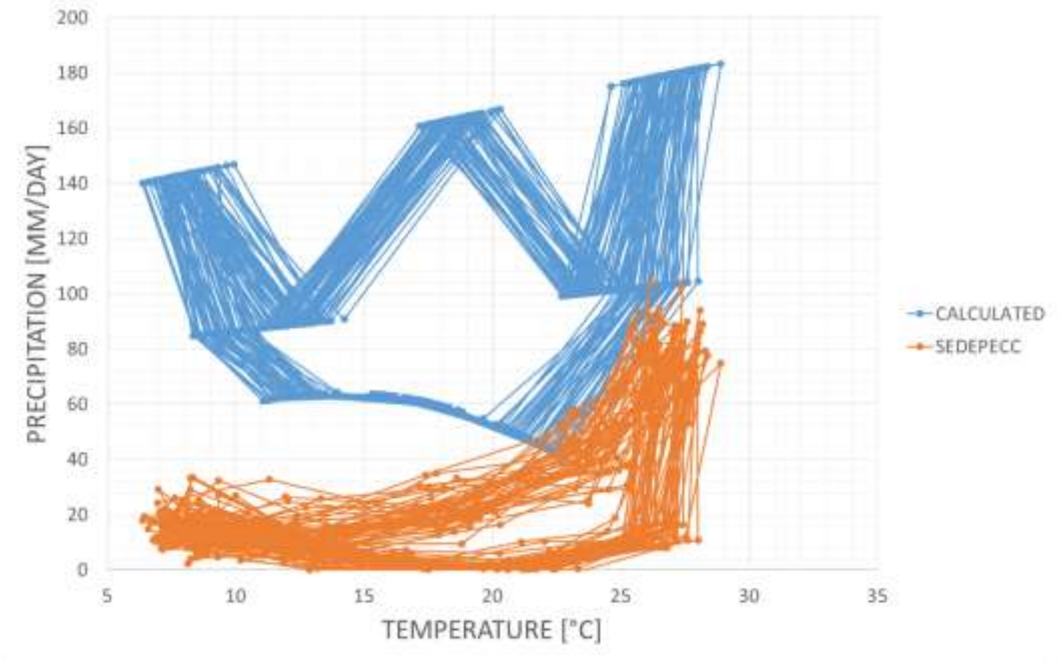

Figure 9. Rainfall against temperature, with adjustment formula overlaying the results recorded in the SEDEPEEC platform.

- State: Yucatán
  Name: Santa Elena
  Municipality: Santa Elena
  Key: 31027

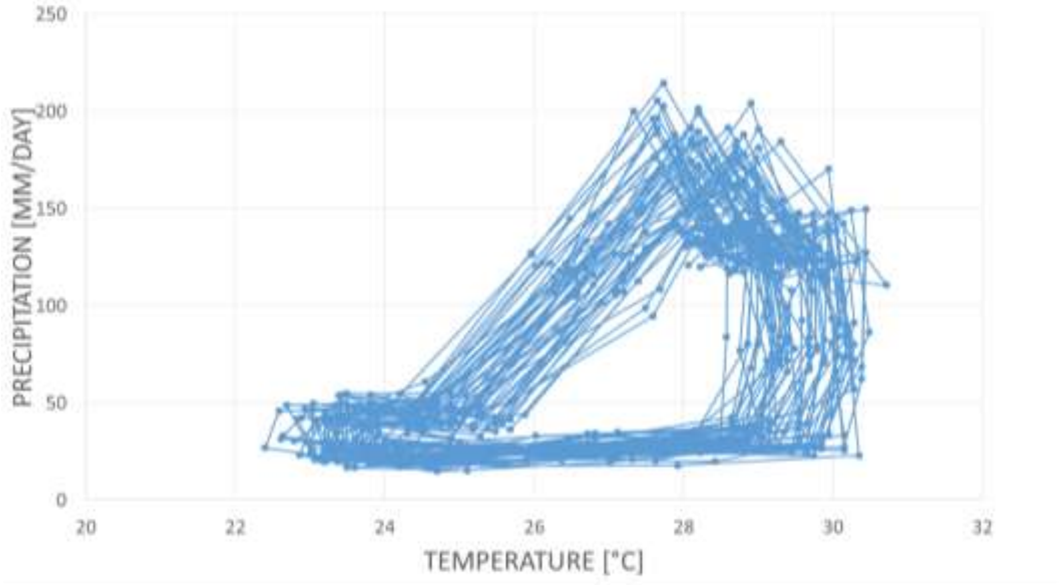

Figure 10.Rainfall against temperature, data SEDEPECC

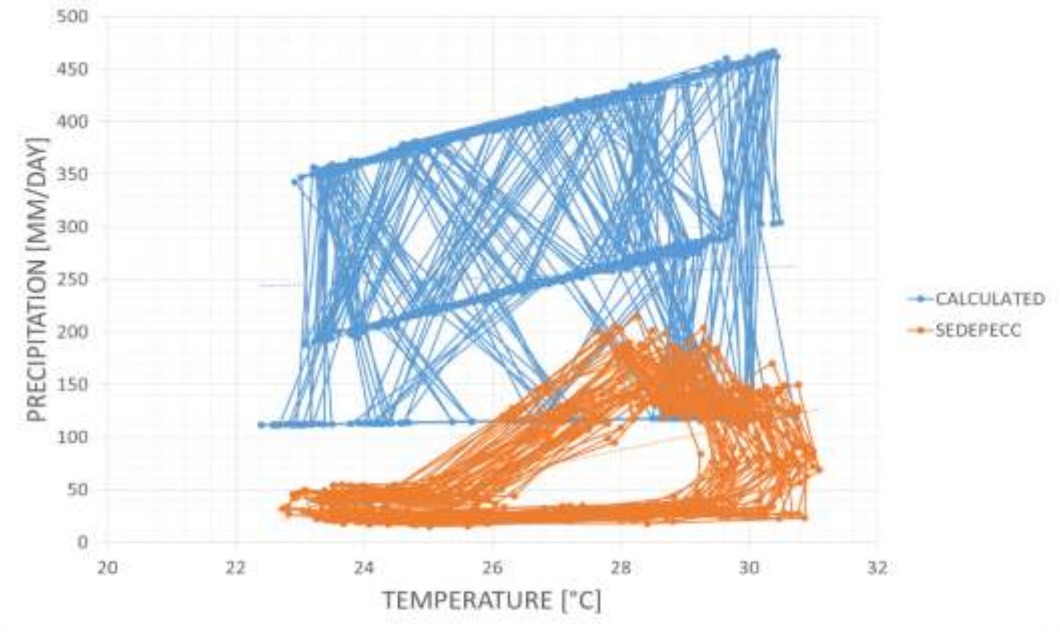

Figure 11. Rainfall against temperature, with adjustment formula overlaying the results recorded in the SEDEPEEC platform.

- State: Veracruz de Ignacio de la Llave
  Name: Chicontepec de Tejeda
  Municipality: Chicontepec de Tejeda
  Key: 30041

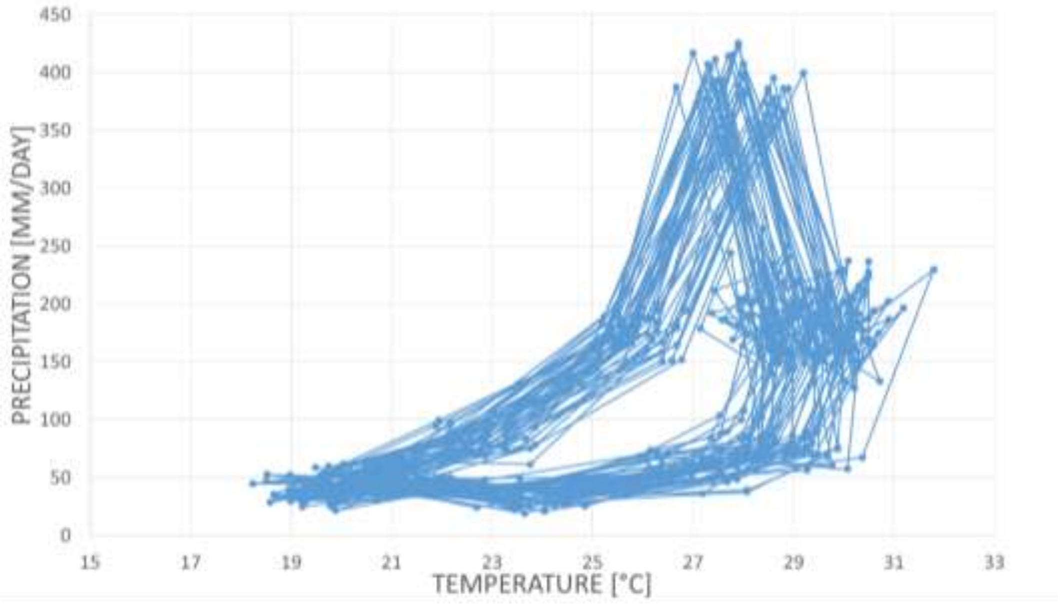

Figure 12.Rainfall against temperature, data SEDEPECC

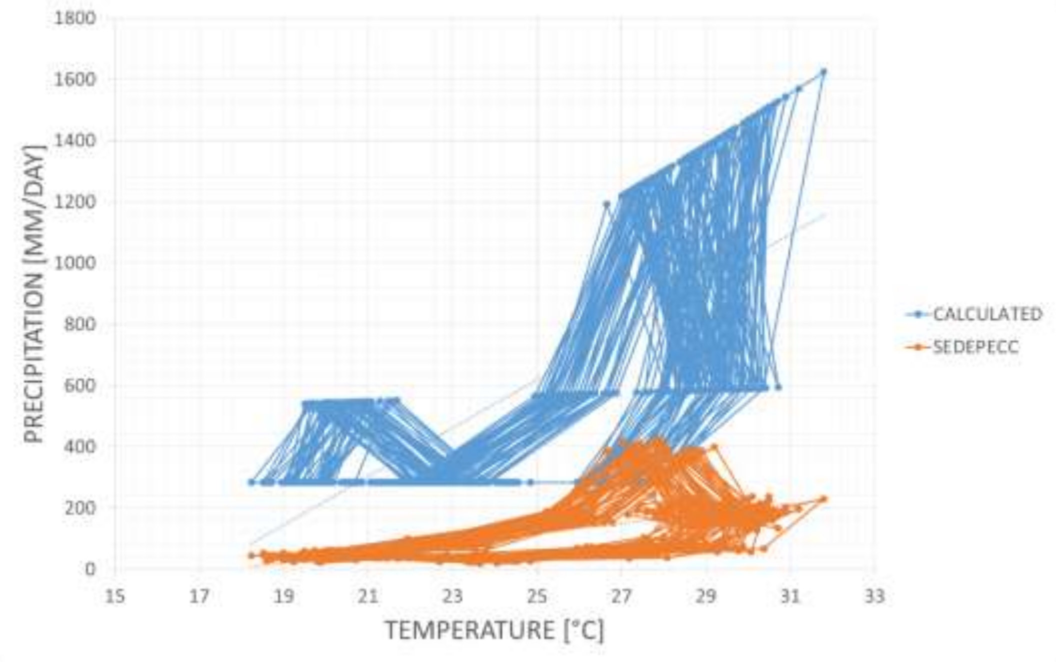

Figure 13. Rainfall against temperature, with adjustment formula overlaying the results recorded in the SEDEPEEC platform.

State: Michoacán de Ocampo
Name: Acahuato
Municipality: Apatzingán
Key: 16228

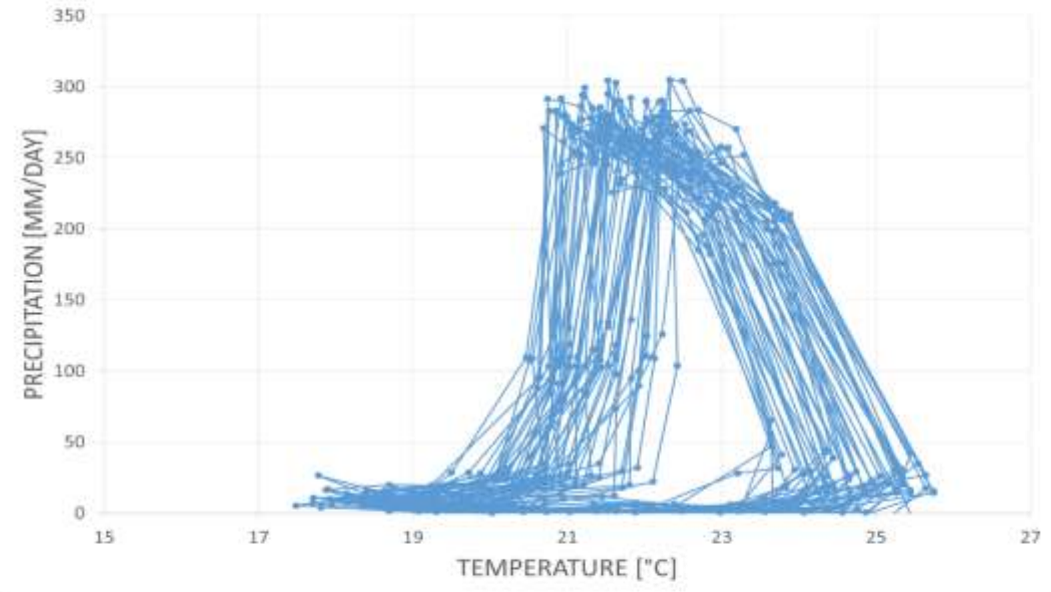

Figure 14.Rainfall against temperature, data SEDEPECC

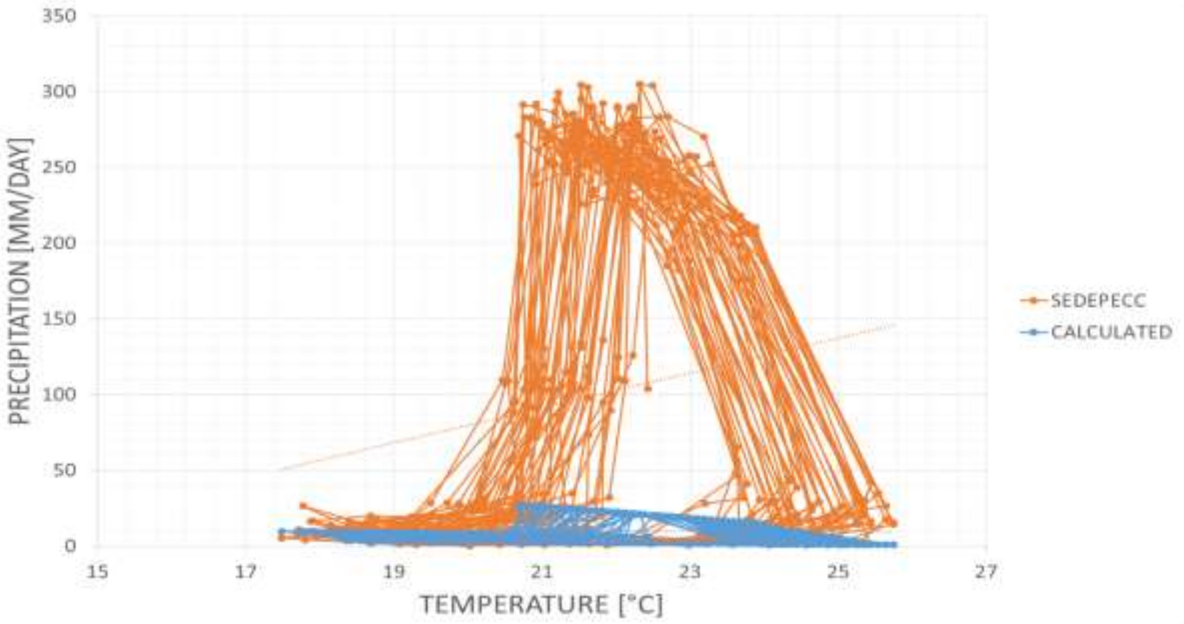

Figure 15. Rainfall against temperature, with adjustment formula overlaying the results recorded in the SEDEPEEC platform.

- State: Guerrero
  Name: Acapulco de Juárez
  Municipality: Acapulco de Juárez
  Key: 12142

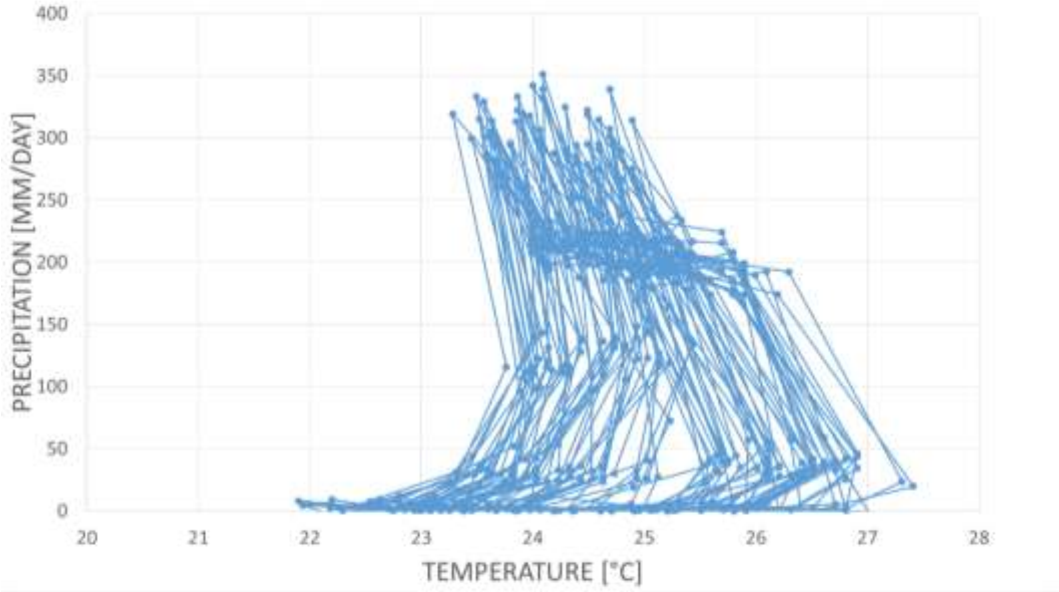

Figure 16.Rainfall against temperature, data SEDEPECC

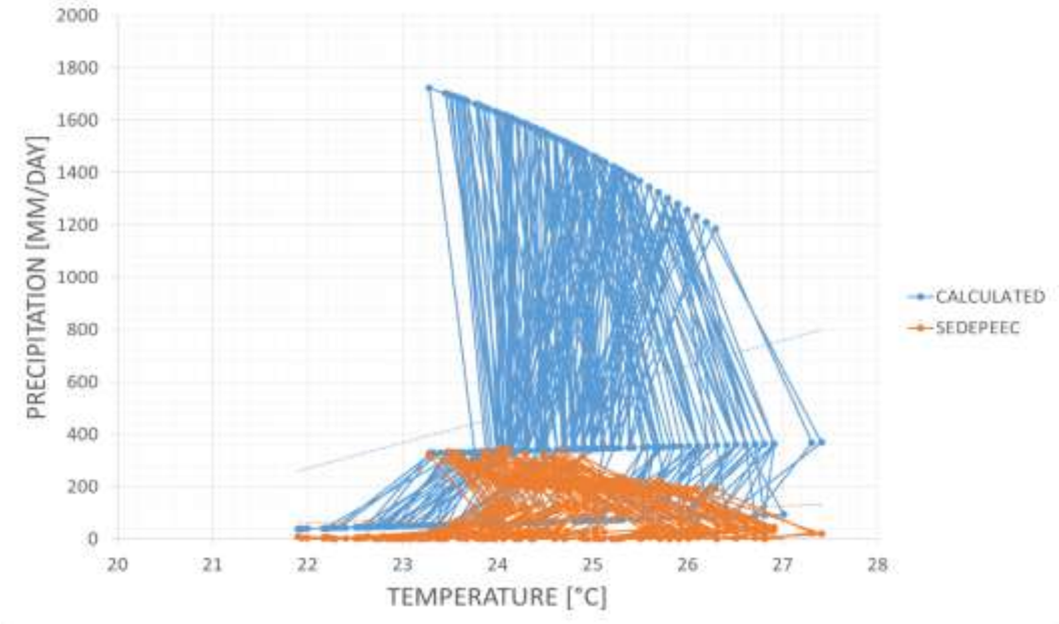

Figure 17. Rainfall against temperature, with adjustment formula overlaying the results recorded in the SEDEPEEC platform.

State: Baja California Sur
Name: Ojo de Agua
Municipality: Comondu
Key: 3039

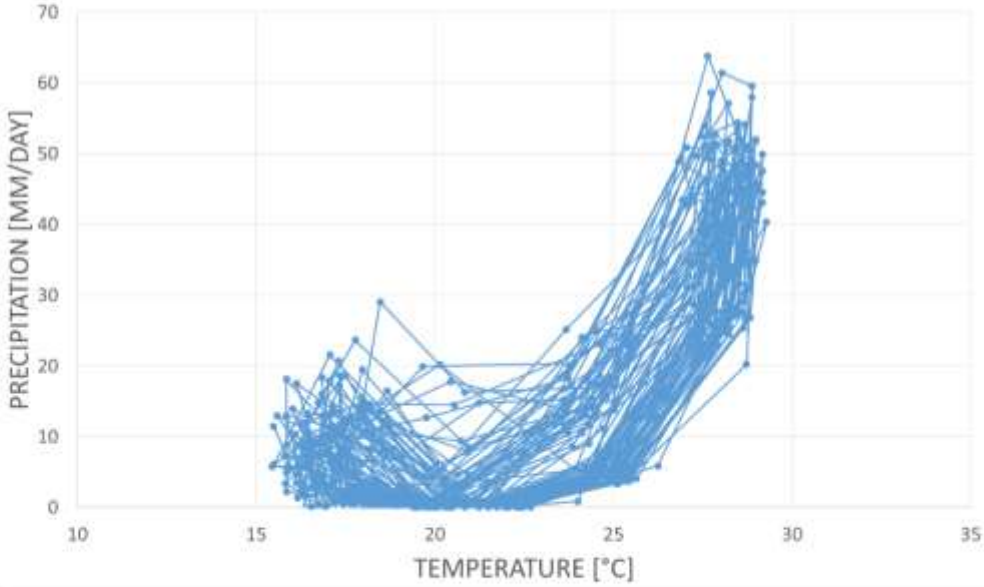

Figure 18.Rainfall against temperature, data SEDEPECC

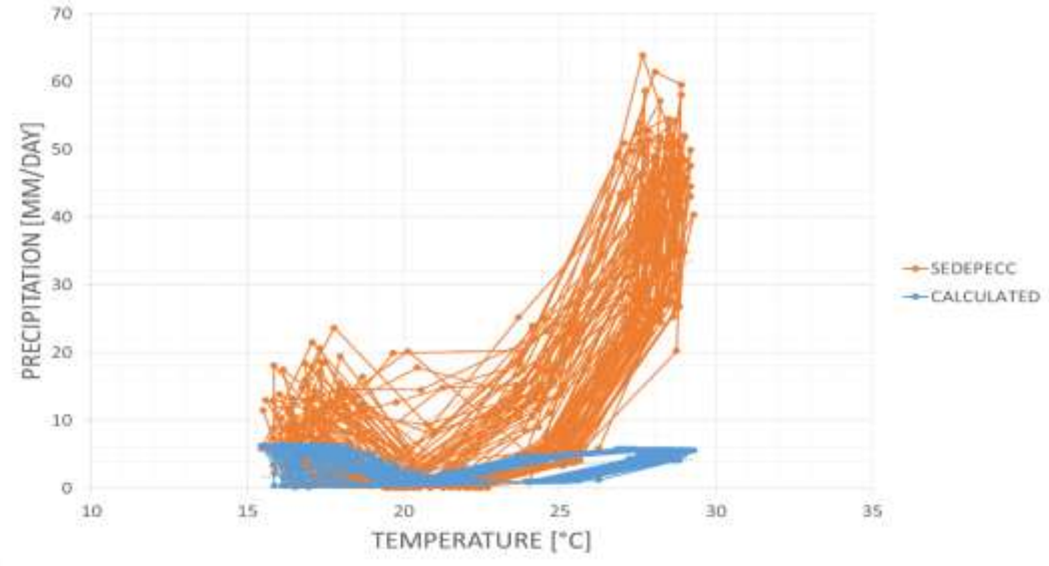

Figure 19. Rainfall against temperature, with adjustment formula overlaying the results recorded in the SEDEPEEC platform.

State: Coahuila de Zaragoza
Name: Ejido Primero de Mayo
Municipality: Escobedo
Key: 5147

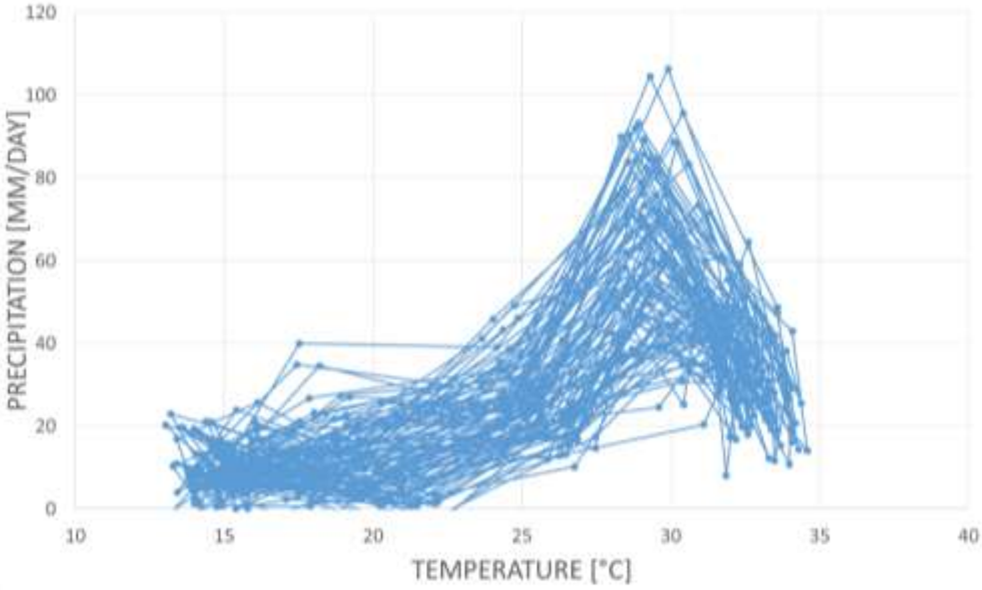

Figure 20.Rainfall against temperature, data SEDEPECC

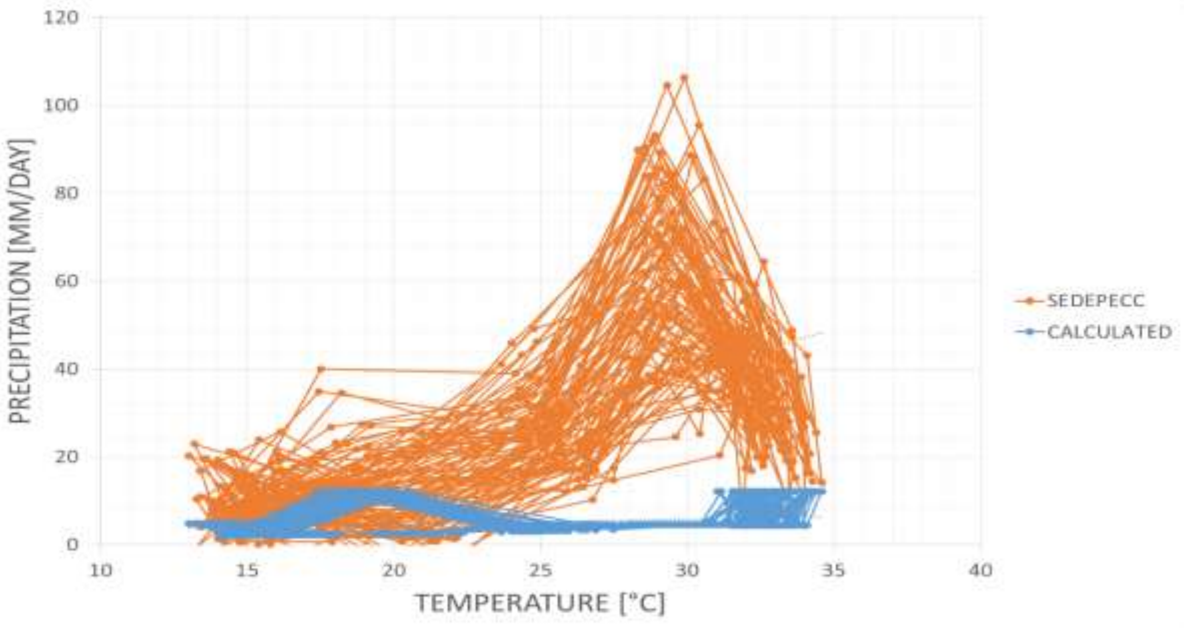

Figure 21. Rainfall against temperature, with adjustment formula overlaying the results recorded in the SEDEPEEC platform.

According to the above figures, where the graph of the data collected by the SEDEPECC platform and the graph of the values calculated through PG shown, an analysis and comparison of variances obtained through the formula is performed

$$R^2 = \frac{Variance\ \ y - Variance_{\ error\ y}}{Variance\ \ y} \quad (1)$$

Conclusions and recommendations

Ten climatological stations were selected that have daily precipitation and air temperature records; the patterns of monthly precipitation behavior were identified as a function of temperature year-on-year at historically recorded values. Historical adjustment models were obtained with genetic programming; These models were adjusted to a method of dissection both vertical and horizontal, coupled with a higher correlation to the horizontal dissection method. These models were subsequently used with data generated with climate change to observe the feasibility of their application to future measurements.

In this work new models obtained with GP were provided to approach meteorological phenomena based on monthly rainfall and air temperature in certain regions of Mexico, which can be useful to take appropriate measures to mitigate the damage or effects that could occur in nearly future.

And although there are few climatological stations analyzed, distributed throughout the territory of Mexico, we can appreciate the changes and trends that are suggested in a common period, the idea of this research is to establish solid bases to continue the study of a greater number of stations and for a much longer period of time.

According to the results, the station that best adapted to the modeling based on the behavior of rainfall intensity against the historical average temperature is Acahuato in Michoacan, Mexico with Key 16228 since obtained the highest R^2 result of establishing the best pattern through the following equations (Table 2).

Table 2 . Summary of equations used for station 16228.

	STATION	**16228**	
Month	DATA SET	*EQUATION*	*R^2*
ALL	EQUATION REPRESENTATIVE YEAR	$hp=-0.0000017T^5+0.0134T^2+T-9.4449$	*0.15*
2,3,4,5,12	DISSECTION EQUATION H1	$hp=10.1959-0.3657T$	*0.99*
1,10,11	DISSECTION EQUATION H2	$hp=-0.055T^2+1.4554T+1.0515$	
6,7,8,9	DISSECTION EQUATION H3	$hp=104.246-3.7341T$	
1,2,9,10,11,1,2	DISSECTION EQUATION V1	$hp=-0.00000005T^6+0.000001T^5+0.0212T^2-0.0196T$	*0.58*
3,7,8	DISSECTION EQUATION V2	$hp=-0.4174T^3+9.2732T^2+8.86481T+1.8319$	
4,5,6	DISSECTION EQUATION V3	$hp=154.517-5.9036T$	

Figure 22 shows the comparison between the data measured in a representative year and the data calculated by the PG models, likewise shows the data grouped horizontally and vertically, using an identity line to establish a better comparison parameter.

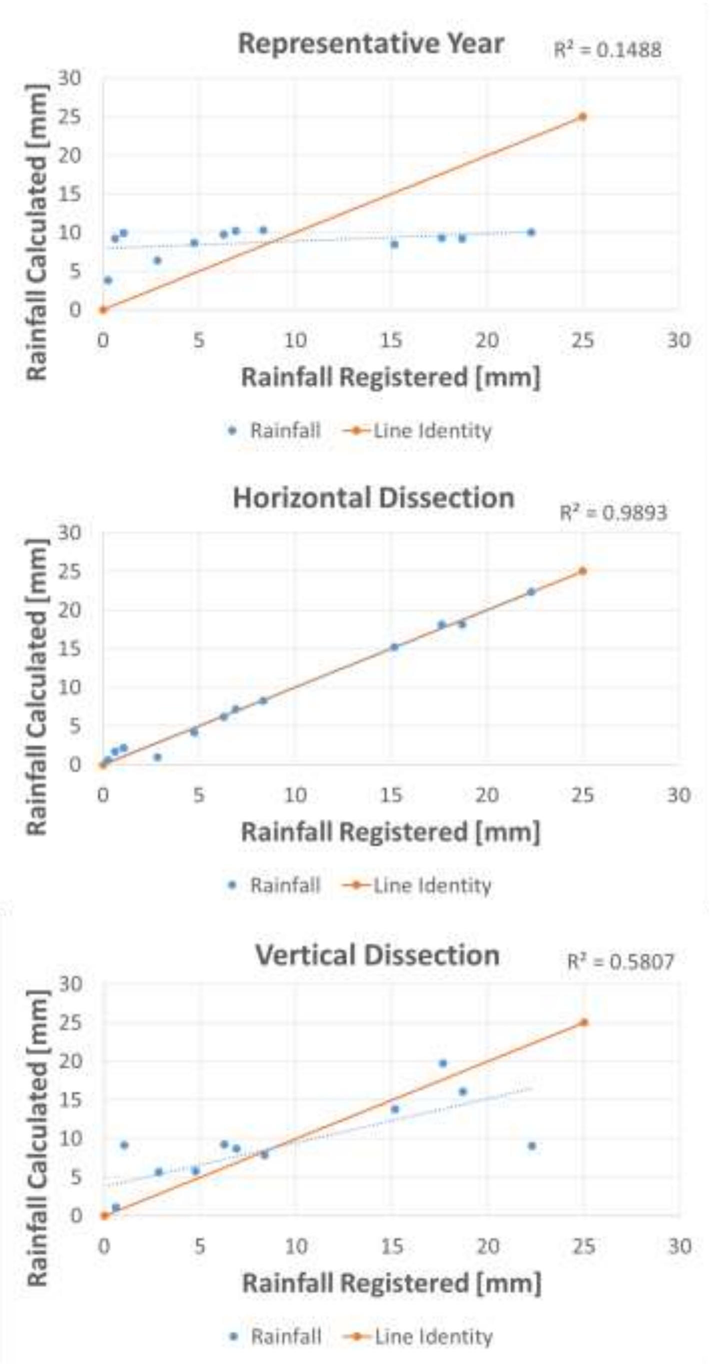

Figure 22. Graphs of comparison between the measured rain and the calculated rain, according to their corresponding grouping

Figure 23 show graphically, the relationship between the record in the SEDEPEC the IMTA and calculated with the best model (horizontal dissection)

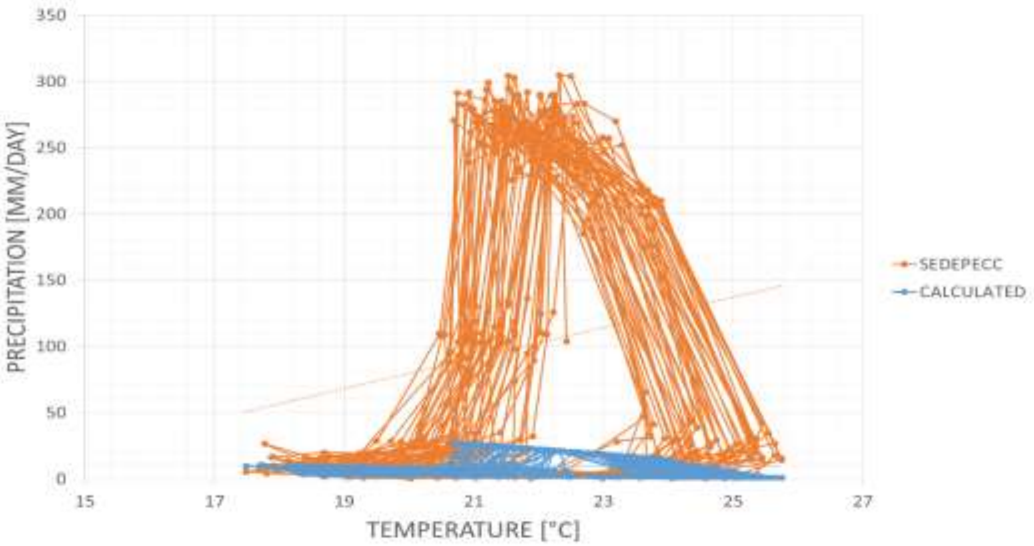

Figure 23. Rainfall against temperature, with adjustment formula overlaying the results recorded in the SEDEPEEC platform.

According the Figures given above, an overview of how rainfall and temperature will rise or will decrease as the case can be noticed, creating long-term scenarios, which can establish a hydraulic work or prevention policies that mitigate the effects caused by these changes in these climatic characteristics.

References

1. Allerup, P., Madsen H. y Vejen F. (2000) Correction of precipitation based on off-site weather information. Atmospheric Research. Volume 53, Issue 4, pages 231-250.
2. Berget, M., schlegel, T., kirchhofer W. (2005) Homogeneous temperature and precipitation series of Switzerland from 1864 to 2000. International Journal of Climatology. Volume 25, Issue 1, pages 65–80
3. Brunetti, M., Maugeri, M., Monti, F., Nanni, T. (2005) Temperature and precipitation variability in italy in the last two centuries from instrumental homogensed time series. International Journal of Climatology. Volume 26, Issue 3, pages 345–381
4. Cahalan, R F, Lawrence, W.E., Man-Li, W. (1996) Empirical orthogonal functions of monthly precipitation and temperature over the United States and homogeneous stochastics models, Journal of Geophysical Research. Volume 101, Issue D21, pages 26309–26318

5. Christensen, J. H., Boberg, F., Christensen, O.B., Lucas-Picher P. (2008) On the need for bias correction of regional climate change projections of temperature and precipitation. Geophysical Research Letters. Volume 35, Issue 20, Pages 1-6.
6. Cho, J., Komatsu, H., Pokhrel, Y. (2011) The effects of annual precipitation and mean air temperature on annual runoff in the global forest regions. Climatic Change. Volume 108, pages 401–410
7. De Lima, M.I.P., Espírito, S.F, Ramos. A., De Lima, J.L.M.P. (2012) Recent changes in daily precipitation and surface air temperature extreme in mainland Portugal, in the period 1941-2007. Atmospheric Research. Volume 127, pages195–209
8. IMTA.2016. SEDEPECC: http://galileo.imta.mx/aplisedepecc.php?entrar=1
9. Magaña Rueda V.. 2006 Climate Change: A View from Mexico
10. Piani, C., Weedon, Gp, Best, M., Gomes, Sm, Viterbo, P., Hagemann, S., Haerter, Jo (2010) , Statistical bias correction of overall daily precipitation and temperature simulated for the application of hydrological models. Journal of Hydrology. Volume 395, Issues 3–4, 15 December 2010, pages. 199–215
11. Strack, Je, Pielke, R., A., Steyaert, L.T., Knox, R.G. (2008) Sensitivity of June and near-surface temperatures in the eastern precipitation United States to historical land cover since European settlement Changes. Water Resources Reseach. Volume 44, Issue 11, pages 1-13.
12. Tuomenvirta H. (2001) Homogeneity Adjustments of temperature and precipitation series-Finnish and Nordic dat. International Journal of Climatology. Volume 21, Issue 4 30, pages 495–506