Article

Models to predict probabilistic precipitation in Tabasco, Mexico generated with published information

Eugenio Romantchik Kriuchkova[1§]
Ramón Arteaga-Ramírez[1§]
Rocío Cervantes Osornio[2]
Mario A. Vázquez-Peña[1]

[1]Departamento de Irrigación- Universidad Autónoma Chapingo. [2]Instituto Nacional de Investigaciones Forestales Agrícolas y Pecuarias.

[§]Autor para correspondencia: arteagarr@gmail.com.

Abstract

The weather stations usually present data lost in their records, which complicates probabilistic studies in this case of precipitation. But there is published information obtained for this purpose, so the objective was to generate models to predict probabilistic precipitation in the state of Tabasco with published information. There was information published graphically of 19 stations in the state of Tabasco, of these the average precipitation was taken and the probabilistic precipitation was generated at levels of 80, 60, 40 and 20%, the simple linear model was used and four models were generated to estimate the probabilistic precipitation at the indicated levels based on the average rainfall with data from 17 stations, the other two were used in the validation of the models. To define the predictive goodness of these, the square root of the mean square of the error (RCCME) was used. The four generated models presented good adjustment, since their coefficients of determination were 0.959, 0.985, 0.991 and 0.97, in the probability levels of 80, 60, 40 and 20% respectively. The values of the RCCME varied from 4.6 to 27.7 mm which indicates that the models are good predictors.

Keywords: precipitation estimation, linear model, square root of the mean error square and crop zoning.

Reception date: june 2018
Acceptance date: july 2018

Introduction

The influence of meteorological elements on the primary activities of mankind was conceived since the dawn of humanity, nowadays it is not surprising that this influence is necessary to address and understand it, given that a large part of the surface for agricultural production in Mexico is carried out in temporary conditions (rainfed). Lazcano (2006) reports that 22 million hectares are planted annually, of which six million have irrigation water and the remaining 16 depend on the precipitation that is received directly in the plots.

In Mexico, due to its geographical position and its steep orography, an irregular pluvial distribution, both spatial and temporal, originates. So it must face a permanent fight against deficiencies or excesses of water, depending on the region in question, so it is necessary to analyze the spatial and temporal distribution of precipitation, as noted by Lozada and Cesar (2003). Both forms are very useful in the knowledge of water availability. Loomis and Connor (2002) consider that rainfed agriculture in wetlands could be considered fortunate to freely dispose of a good such as water. However, water supply is rarely the ideal, varying from excesses at various times to temporary deficiencies in others.

But note that climate resources have not been used in those regions where the climate is humid or semi-humid, that is, where unique conditions for rainfed agriculture are theoretically offered, since rainfall is abundant and combined with high temperatures in all year round and in addition there are no low temperatures that damage crops (Bassols, 1998).

From an agroclimatological point of view, it is interesting to know about precipitation: its total annual quantity, its distribution through the months, its frequency and intensity (quantity and duration), and the effect of precipitation depends on the amount of the sheet of rain accumulated in the year, also of its temporary distribution, and this is very unequal during the year, as well as its interannual and intrazonal variability (Smith, 2000).

In general, when analyzing the variability of precipitation and the estimation of probabilities in its application to agriculture, it is a matter of knowing how often the soil will receive a certain amount of precipitation and how often that soil will receive an amount lower or higher than she. García-Benavides (1979) indicates that the answer is specified as a fraction or percentage of probability, for example: 0.8 or 80% probability of exceedance or as a frequency period, one year out of five, one in four, etc., also proposes that these levels of probability are derived from economic considerations according to which it is accepted that the production of a crop is economically acceptable.

To analyze the variability of precipitation, it is necessary to have data measured for many years and in different places in a region, state or country, but unfortunately, as indicated by Campos (2007), there are many weather stations whose records are incomplete, or missing one or several months in a row in one or more years, in this case the stations of Tabasco are not the exception. The main objective was to generate models to calculate the monthly probabilistic precipitation based on the average monthly rainfall in the state of Tabasco, with published information.

Materials and methods

We used information presented by García (1977) in his publication of climates for the state of Chiapas and Tabasco. In Table 1 we have the identification numbers of the Tabasco state stations used, their total precipitation for different levels of exceedance probability and the annual average value and in Figure 1 we can see their location in the state, their name and identification number.

Table 1. Weather stations that were used in the study and their total precipitation (mm) at different levels of exceedance probability. García (1977).

Station	Total precipitation (mm) at the indicated probability levels (%)				Average annual rainfall (mm)
	80	60	40	20	
001	620.0	1 029.0	1 471.0	2 328.0	1 335.4
002	930.0	1 305.5	1 793.0	2 472.0	1 675.5
003	1 308.0	1 843.0	2 353.0	3 095.0	2 150.1
004	1 071.0	1 533.0	2 109.0	2 887.0	1 900.3
006	1 019.0	1 571.0	2 177.0	3 172.0	1 985.0
007	1 783.0	2 509.0	3 231.0	4 196.0	2 930.0
008	1 041.0	1 505.0	2 033.0	2 760.0	1 835.0
009	1 184.0	1 702.0	2 242.0	3 045.0	2 043.5
010	983.0	1 441.0	1 967.0	2 746.0	1 784.6
011	1 238.0	1 865.0	2 508.0	3 374.0	2 246.6
012	1 325.0	1 868.0	2 390.0	3 173.0	2 189.3
015	1 103.6	1 562.6	2 078.8	2 671.1	1 982.0
016	1 001.9	1 431.4	1 939.8	2 611.7	1 857.7
017	853.7	1 213.4	1 630.3	2 205.6	1 580.1
018	1 610.0	2 350.0	3 195.0	4 135.0	2 797.8
019	2 145.0	2 880.0	4 000.0	4 975.0	3 470.2
020	1 165.0	1 690.0	2 257.0	2 855.0	2 004.5
021	876.0	1 268.0	1 656.0	2 315.0	1 529.1
022	1 001.0	1 489.0	1 969.0	2 887.0	1 791.9

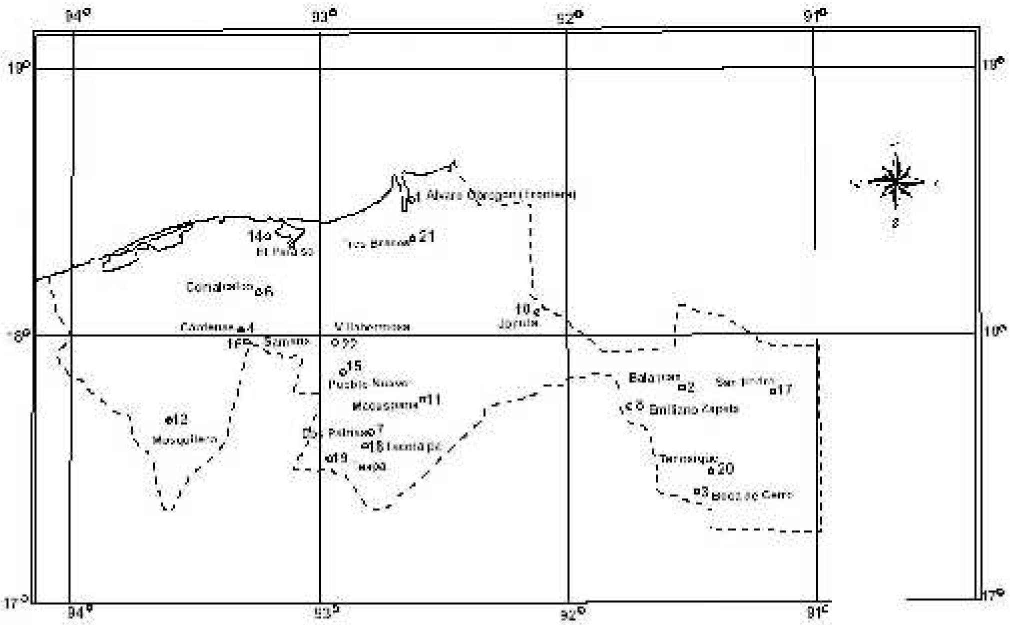

Figure 1. Stations used in the study with your name and identification number. García (1977).

Of the 19 stations (Table 1), 17 were used in the adjustment or generation of the probabilistic models and two for the validation of these.

The generation of probabilistic rain was made with the help of 12 figures (one per month) per station (17) that presents in its publication García (1977) for the state of Tabasco. The procedure performed was: of the figures mentioned (Figure 2, October is presented), the amount of rainfall expected each month is obtained graphically by means of a proportional ratio according to the scale used in each figure, for the levels of exceedance probability selected in advance, in this case; 20, 40, 60 and 80%. In addition, the figures show the following data: minimum precipitation (X1), maximum precipitation (X2), coefficient of variation (CV), average monthly precipitation (XM), standard deviation (σ) and probability of occurrence (PM) of the average monthly rainfall (Figure 2).

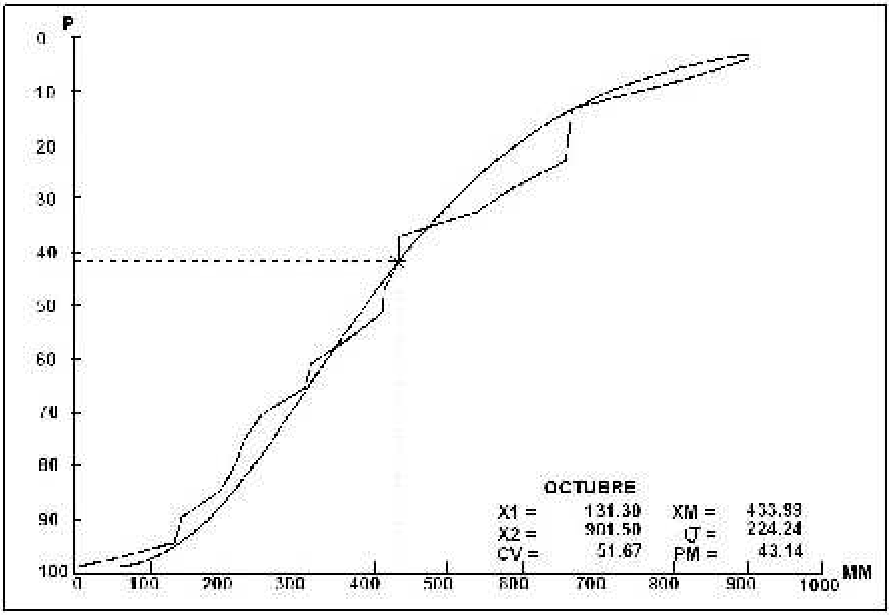

Figure 2. Empirical and theoretical probabilistic distribution for the month of october. García (1977).

With the data obtained in the previous step, four dispersion diagrams were made (one for each probability level), where the independent variable is the average monthly precipitation and the dependent variable the amount of monthly precipitation at a probability level.

With the trends presented by the dispersion diagrams, the corresponding model was adjusted, the simple linear was used and according to Said and Zárate (2000) to know the adjustment of the model it is necessary to determine the coefficient of determination (r^2).

Finally, the validation of the obtained models consists of estimating with the linear models obtained and with the measured data of average monthly rainfall in stations 008 and 015 the probabilistic monthly rain data for each probability level, these estimated data were compared with the data obtained from the figures of these two stations presented by García (1977), which are considered as observed or measured values. Two comparisons were made, the first was with the simple linear regression method, if the estimated data (E) are similar to the measured ones (M), the linear model would be M= E, so, the criteria in evaluating how good estimator is a model are: if the value of 'a' is close to zero and that of 'b' to one (Donatelli *et al*., 2004; Allen *et al.*, 2006). The second method to define how good predictors of data are the models obtained, was the square root of the mean square error (RCCME), since studies conducted by George *et al*. (2000); Cai *et al*. (2007); Tojo *et al*. (2007); Kang *et al*. (2009) indicate that this index is used to evaluate the goodness of fit of a model and is calculated with the following relationship:

$$R \qquad = \left(\frac{\sum_{i=1}^{N} (E_i - M_i)^2}{N} \right)$$

Where: E_i estimated value; M_i measured value; and N= number of observations.

Results

As shown in Figures 3 to 6, where the dispersion diagrams are presented between the average monthly precipitation of all months (12) and all seasons (17) with the amount of precipitation of the probability levels that were considered in this study, all trends are linear.

In Table 2, the values of the ordinate to the origin (a) and the slope (b) of each one of the obtained models are presented, as well as its r^2. As indicated by the data of stations 008 and 015 that were selected at random, they were not considered in the calculation of the coefficients of the models obtained (Table 2) in each probability level, so that these models were validated.

Table 2. Coefficients and statistics of the models obtained between the average monthly precipitation amount and the probabilistic one.

Probability (%)	A	b	r^2
20	35.2	1.286	0.970
40	- 1.6	1.113	0.991
60	- 17.1	0.921	0.985
80	-26.2	0.723	0.959

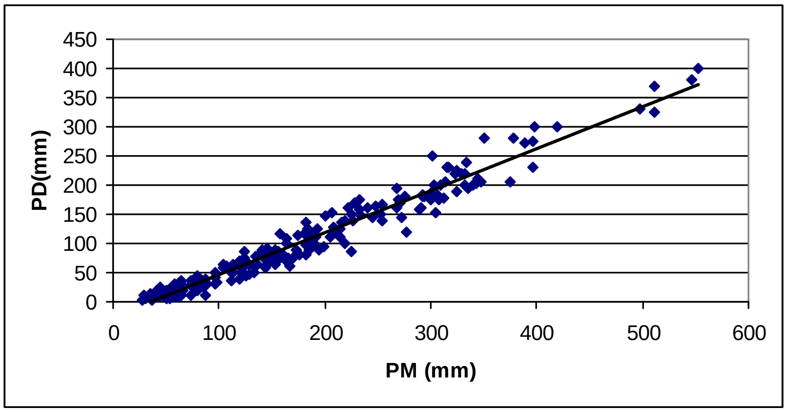

Figure 3. Dispersion diagram between the average monthly precipitation (PM) of all months (12) and seasons (17) and probabilistic precipitation (PD) at 80%.

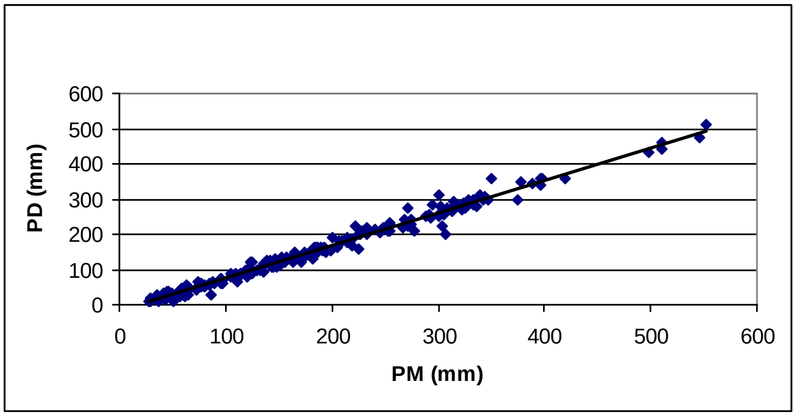

Figure 4. Dispersion diagram between the average monthly precipitation (PM) of all months (12) and stations (17) and probabilistic precipitation (PD) at 60%.

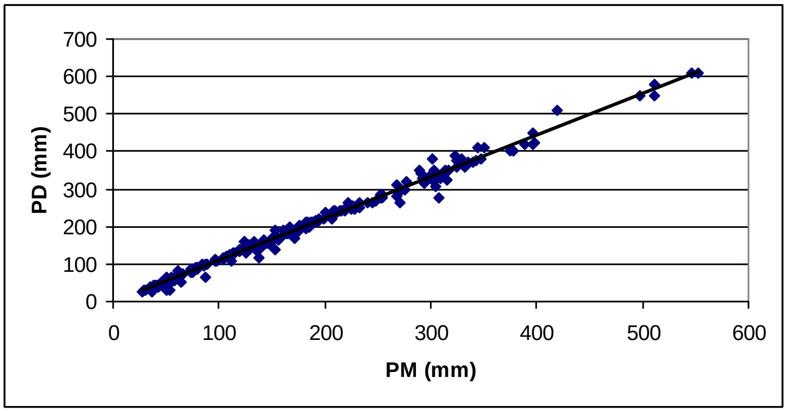

Figure 5. Dispersion diagram between the average monthly precipitation (PM) of all months (12) and stations (17) and probabilistic precipitation (PD) at 40%.

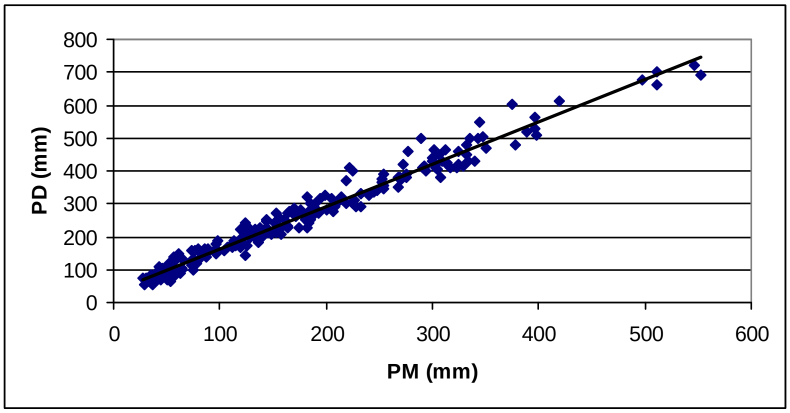

Figure 6. Dispersion diagram between the average monthly precipitation (PM) of all months (12) and seasons (17) and probabilistic precipitation (PD) at 20%.

In the Table 3 shows the values of 'a' and 'b' and their coefficient of determination at each level of probability, of stations 008 and 015, when comparing the estimated values of probabilistic rainfall amount with the obtained models (Table 2) from their average rainfall with the data obtained (measured) from the figures presented by García (1977), the estimated data are very similar to those measured. In Table 4 we have the values obtained from the square root of the mean square of the error per station and level of probability.

Table 3. Validation of the models with the data of station 008 and 015.

Station	Probability (%)	A	B	r^2
008	20	7.3	0.960	0.990
	40	5.9	0.970	0.999
	60	- 3.4	1.040	0.997
	80	2.0	1.005	0.992
015	20	- 28.0	1.012	0.990

	40	-10.8	1.010	0.994
	60	-5.1	1.002	0.990
	80	0.3	0.984	0.997

Table 4. Values of the square root of the mean square of the error (mm).

Station	Probability (%) 20	40	60	80
008	13.6	4.6	6.6	7.1
015	27.7	11.8	9.7	9.8

In the Figure 7 shows the precipitation data estimated with the models, against the measure in the probability levels 20 and 40%, in station 008 it is observed that for both probability levels a very good estimation is presented since all the data are very close to the theoretical line 1:1.

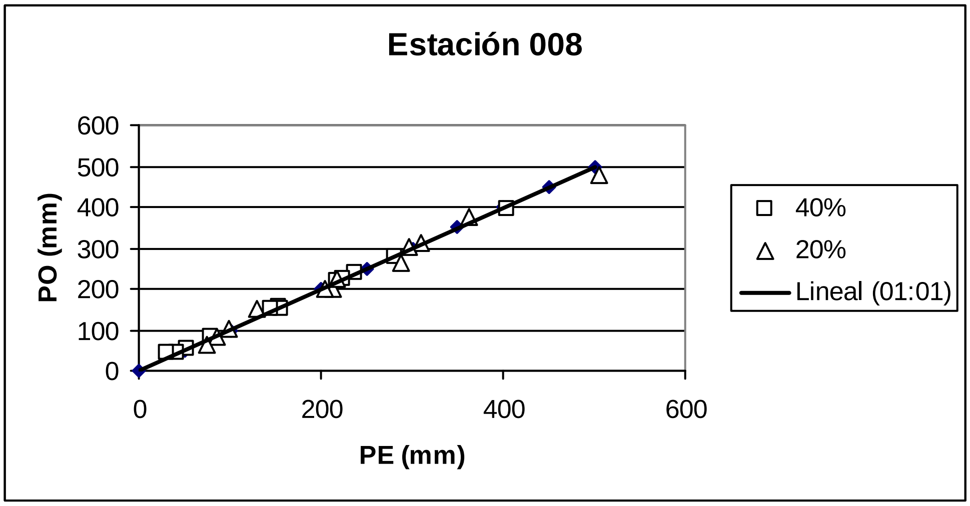

Figure 7. Comparison of average monthly rainfall values (measured, PO) versus estimated (PE).

In Figure 8 we have the data of station 015 for 20 and 80%, the first was the one that did not present a very good estimate, it is observed that some data are close to the theoretical line (1:1) and the estimated values underestimate those observed, which is verified with the data presented in Table 3 (a= -28 and b= 1.012). In the second, all points are on the theoretical line 1:1 (a= 0.3 and b= 0.984).

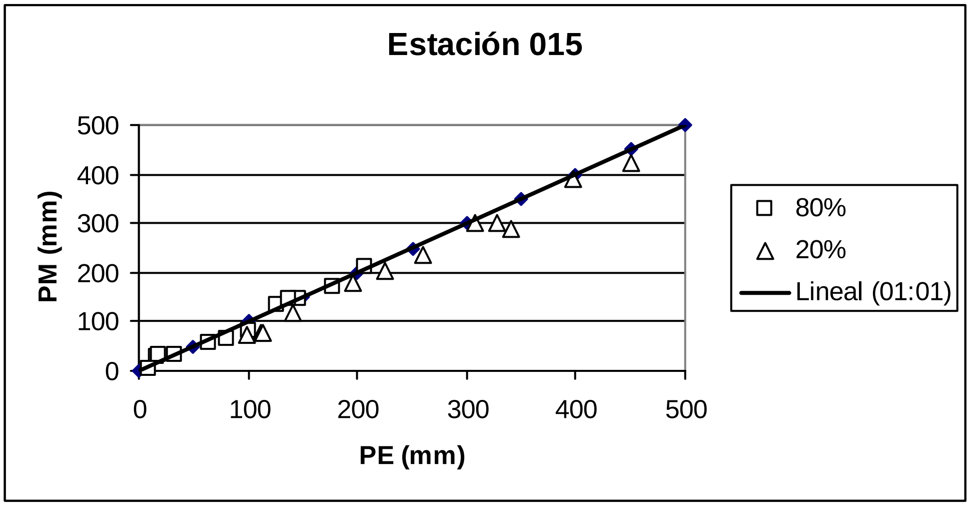

Figure 8. Comparison of average monthly rainfall values (measured, PO) versus estimated (PE).

The average annual precipitation (Table 1), varies from 1 335.4 to 3 470.0, in 80% probability of exceedance is from 620 to 2 145 and in 20% from 2 328 to 4 975, these values correspond to station 001 (Border) and 019 (Teapa) respectively.

In the Figures 9 and 10 show the temporal variation of the average precipitation and the different levels of probability considered, in stations 001 and 019 respectively, with which in a specific way (for each station) the variation of the water potential is defined with what is counted and in that month.

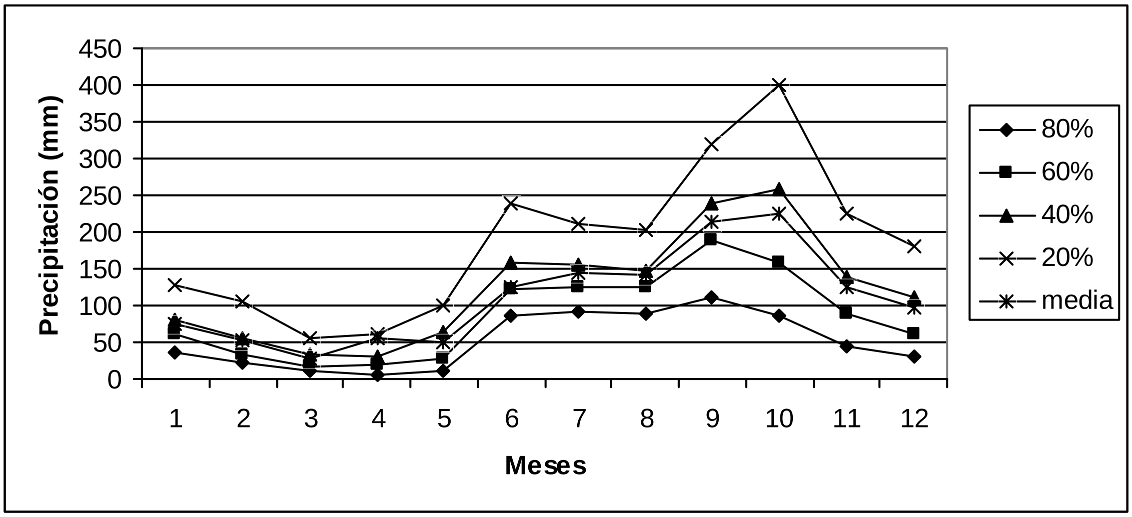

Figure 9. Temporal variation of the monthly precipitation in station 001, for each of the exceedance probability levels indicated.

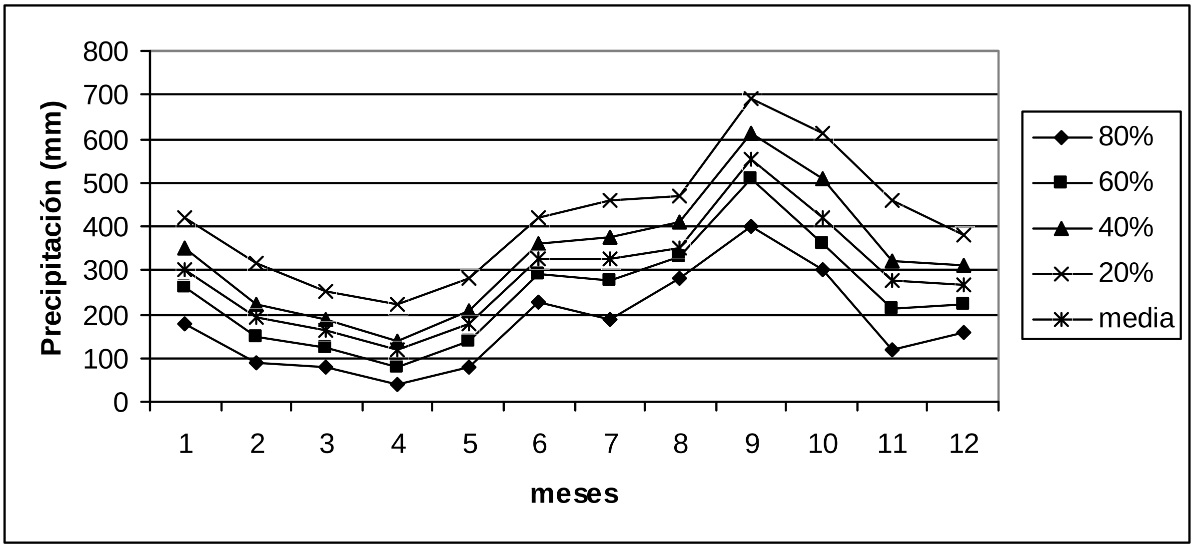

Figure 10. Temporal variation of the monthly precipitation at station 019, for each of the exceedance probability levels indicated.

In the Figure 11 shows only the plane of average precipitation isohyets for the month of July taken from García (1977), in which the isohyets of 150, 200, 250, 300 and 400 mm are seen, these were transformed to new values of probabilistic isohyets with the obtained models, in Table 5 these are given, with which, four planes one for each level of probability were generated, with which there is the spatial and temporal variability of the precipitation.

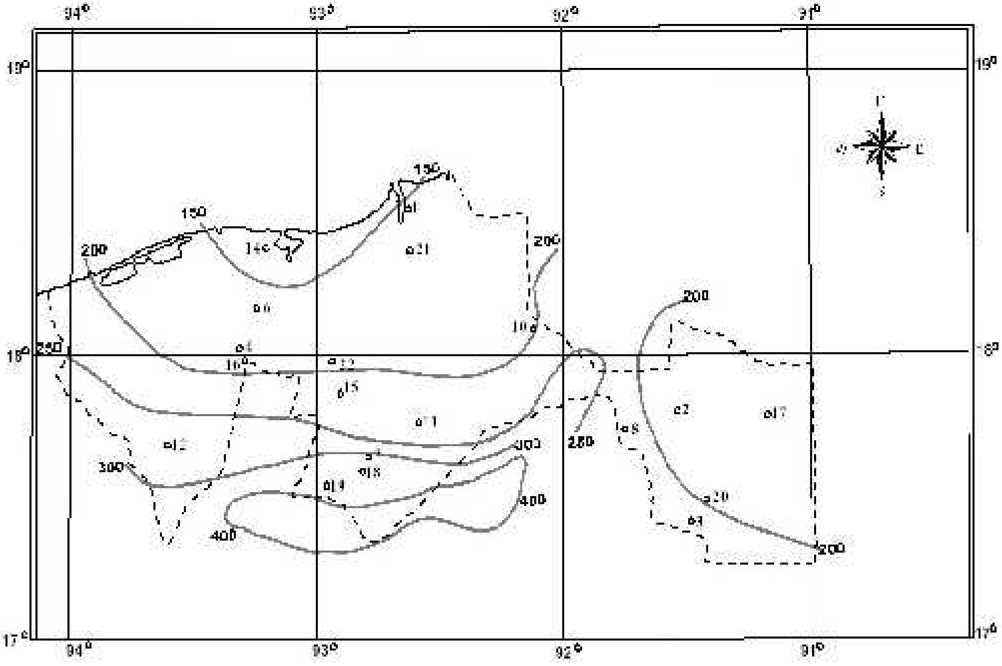

Figure 11. Average precipitation of the month of july in mm. García (1977).

Table 5. July isohyets with their new values obtained with the models generated (mm).

Isohyet of July	New values of the isohyets (mm) for the Probability (%) indicated			
	20	40	60	80
150	228.1	165.4	121.1	82.3
200	292.4	221.0	167.1	118.4
250	356.7	276.7	213.2	154.6
300	421.0	332.3	259.2	190.7
400	549.6	443.6	351.3	263.0

Discussion

All linear models presented a good fit similar to what Hargreaves (1975 and 1977) determined. The best was the probability level at 40%, for its coefficient of determination and all are highly significant. It should be noted that the amount of probabilistic precipitation estimated 40%, is very similar to the amount of average monthly precipitation, which agrees with García (1977), this indicates that it has a frequency of occurrence of two out of every five years in the region in study. The values of 'a' and 'b' of the model obtained for 80% of probability agree with those reported by Oldeman (1987) for a probability of 75%, in eight regions of the world: the values of 'a' vary from -14 (Suriname) to -32 (Malaysia) and those of 'b' from 0.85 (Amazon Region, Brazil) to 0.77 (Suriname), for the relationships at levels of 60, 40 and 20% have not been reported for other places.

With the first criterion for the validation of the models, that of applying the linear regression method, in both stations, the value of 'b' is very close to the unit and when considering what Allen *et al*. (1998), that if the value is between: 0.7< b <1.2 the estimation of the data in each of the probability levels is good. The value of 'a' presents variations that depends on the magnitude of the information that was used, they are slightly larger at station 015 for 60, 40 and 20% is the exception at 80% (Table 3). The best model in the estimation of the data with this criterion corresponds to the 80% probability level and was followed by 60, 40 and 20% (Table 3), in the two stations when considering the values of 'a' and 'b', the estimation of the data with the models was better in station 008.

With the second criterion that is the value of the RCCME (Table 4), it was confirmed that in station 008 the obtained models predicted the data better than in station 015. In addition, it is corroborated that the 20% probability model is the one that presents the highest values (13.6 and 27.7 mm) in both stations. With respect to the order of priority in the estimation of the data in station 008, the 40% model is the best model followed by models 60 and 80% that present very similar values. In station 015 the best ones are 60 and 80% that have very similar values and still 40%.

The second criterion was that which was considered to define the priority of the models, since with the first one there is subjectivity on the part of the person who is interpreting the values of 'a' and 'b', in this work it was the value from 'a', since those of 'b' in all the validation models are very close to one.

If you compare the rainfall totals by probability level of stations 008 and 015 that are presented in Table 1 with the values in Table 4, in fact all models are good estimators, since for the largest value of the RCCME, represents 1% of the total rainfall.

In all the stations under study, the wettest month is presented in September, in 14 of the 19 stations which agrees with that determined by García (2003) and in the other five is in october, the station that has the maximum amount is 019. The driest month is april in 16 seasons, in the other three it is march and the season with the lowest rainfall is 001.

The temporal variability of precipitation defines two periods (Figures 9 and 10), which have the highest values of precipitation (June to January), it is important to define which crops are the most likely to prosper in the state. In the season where precipitation is lower (February to May), extensive agriculture is proposed, without forgetting the natural vocation of the region (agricultural, livestock or forestry). When considering the average total precipitation, it is defined that the climate of the state is humid, since 18 stations have average values greater than or equal to 1 500 mm according to the classification proposed by Oldeman (1987) and agrees with that found by Osias *et al*. (2012).

The maximum amount of precipitation (20%) of station 001 (Figure 9) is similar to the minimum (80%) of 019 (Figure 10), the difference is that for the first level it is expected that this amount of precipitation will be present in one of every five years, in the second in four out of five, based on this the water resource planning has to be totally different in each one of them. The above, helps determine the behavior of precipitation to define the areas where the maximum amounts are, as well as the minimum and if it is adequate or not, either by excess or deficiencies.

At the area level, in this case the state of Tabasco, it is defined that in most of the region, excess precipitation is a limiting factor in agricultural production and that the construction of drainage works would be necessary to improve it. In these areas the important thing is to define with what amount of precipitation and level of probability the design of these will be done, since the precipitation presents great variability in total quantity, year after year, within a year and by regions. In addition, in the areas where the main activity is extensive livestock farming, it also presents its problems. The above, agrees with what García-Benavides (1979) indicates, that the spatial distribution allows to identify geographic zones with certain characteristics of humidity in a given period and the distribution in time helps to know the variation of rainfall during the period of culture, these are adapted in both distributions.

In the zoning of crops, the spatial distribution of precipitation for perennial and annual crops is considered, and for the latter the time of establishment of the crop is related to the temporal distribution. Phillips *et al*. (1992) consider that the evaluation of the amount of precipitation in time and space is required in a number of applications in agriculture and natural resource management, these include the management of water resources in different areas, hydrological modeling, modeling forest, modeling soil moisture in crop production, planning works for different purposes (eviction of excesses, water table control, etc.) and irrigation schedules.

An important aspect to be highlighted are the average annual precipitated sheets (1 335 to 3 470 mm), which generate excess surface water throughout the state, their damage varies in scale, whether it is a farm, an agrarian settlement or a large agricultural area, also depends on the location of the farms inside a basin. These are indicative of poor surface drainage that have an unfavorable influence on the development and cost of crops, on agricultural practices; hinder access, mechanization and the application of inputs. Also these excesses generate adequate conditions in the proliferation of pests and diseases, as well as the rapid decomposition of the fruits that are in contact with the soil. Therefore, the probabilistic analysis of the occurrence of precipitation amounts in time and space is essential in the planning, design and operation of the agricultural, livestock or forestry activities of a region.

Conclusions

The graphic information for the twelve months of each season (17) was represented by means of four models that are easy to use and with which probabilistic precipitation was generated from average monthly precipitation. The four models generated in the probability levels of 80, 20, 60 and 40% presented good adjustment. The results of the validation guide to decide whether a model is used or not, in this case the four models determined to estimate probabilistic precipitation (80, 60, 40 and 20%) with average rainfall in the state of Tabasco and because the difference between the estimated and measured data was minimal, indicating that the models are good and are widely recommended. In stations of the state that have data of average precipitation, with the obtained models, reliable probabilistic precipitation data is generated or estimated. The temporal variation presented in this work is not in a particular year, since the amount of probabilistic precipitation summarizes the information of the entire data series of a station. The 12 planes of medium isohyets (one for each month) were transformed into 48 planes, where the spatial variation of the amount of precipitation of each probability level is observed in addition they show the temporal variation of this.

Cited literature

Allen, R. G.; Pereira, L. S.; Raes, D. y Smith, M. 2006. Evapotranspiración del cultivo. Guías para la determinación de los requerimientos de agua de los cultivos. Estudio FAO. Riego y Drenaje 56. Roma, Italia. 298 p.

Bassols, B. A. 1998. Recursos naturales de México. Teoría, conocimiento y uso. 8ª edición. Editorial Nuestro Tiempo. México, DF. 93-128 pp.

Caí, J.; Liu, Y.; Lei, T. and Santos, P. L. 2007. Estimating reference evapotranspiration with the FAO Penman-Monteith equation using daily weather forecast messages. Agric. Forest Meteorol. 14:22-35.

Campos, A. D. F. 2007. Estimación y aprovechamiento del escurrimiento. Universidad Autónoma de San Luís Potosí, SLP, México. 333 p.

Donatelli, M.; Acutis, M.; Bellocchi, G. and Fila, G. 2004. New indices to quantify patterns of residuals produced by model estimates Agron. J. 96:631-645.

García, B. J. 1979. Estructura metodológica para la caracterización agroecológica de áreas por procedimientos cuantitativos de análisis y su posterior zonificación. Tesis de Doctor en Ciencias. Especialidad suelos. Colegio de Postgraduados. Chapingo México.

García, E. 1977. Precipitación y probabilidad de lluvia en la República Mexicana y su evaluación. Climas. Chiapas y Tabasco. Comisión de Estudios del Territorio Nacional (CETENAL). Secretaría de la Presidencia. Instituto de Geografía (IG). Universidad Nacional Autónoma de México (UNAM), México, DF.

García, E. 2003. Distribución de la precipitación en la República Mexicana. Investigaciones Geográficas 50:67-76.

George, B.; Shende, S. and Raghuwanshi, N. 2000. Development and testing of an irrigation scheduling model. Agric. Water Manag. 46:121-136.

Hargreaves, G. H. 1975. Water requeriments manual for irrigate crops and rainfed agricultura. EMBRAPA and Utha State University. Publication No. 75-D158. 40 p.

Hargreaves, G. H. 1977. World water for agricultural. Climate, precipitation probabilities and adequacies for rainfed agriculture. Utah State University, Logan, USA. 161 p.

Kang, S.; Evett, S. R.; Robinson, C. A. and Payne, W. A. 2009. Simulation of winter wheat evapotranspiration in Texas and Henan using three models of differing complexity. Agric. Water Manag. 96:167-178

Lazcano, M. I. 2006. La producción de biocombustibles en México. Bioetanol. La producción situación. Instituto Mexicano de los Biocombustibles AC. http://www.agroetanol-iocombustible.com.mx/art_bioetanol_1.html.

Loomis R. S. y Connor, D. J. 2002. Ecología de cultivos productividad y manejo en sistemas agrarios. Ediciones Mundi-Prensa Madrid, España. 591 p.

Lozada, G. B. I. y César, S. P.2003. Diferencias entre las deficiencias y excedentes hídricos estimados a partir del balance hídrico climático normal y secuencial de las localidades de Bramon, Venezuela, y Piracicaba, Brasil. Rev. Bras. Agrometeorol. 11(1):195-198.

Oldeman, L. R. 1987. Characterization of main experimental sites and subsites and questions of instrumentation. *In*: Bunting, A. H. (Ed.). Proceedings of Conference on Agricultural Environments. Characterization, Classification and Mapping. Rome, Italy. 101-112 pp..

Phillips, D. L.; Dolph, J. and Marks, D. 1992. A comparison of geostatical procedures for spatial análisis of precipitation in mountainous terrain. Agric. Forest Meteorol. 58:119-141.

Ruiz, A. O.; Arteaga, R. R.; Vázquez, P. M. A.; Ontiveros, C. R. E. y López, L. R. 2012. Balance hídrico y clasificación climática del estado de Tabasco, México. Universidad y Ciencia, 28(1):1-14.

Said, I. G. y Zarate, L. G. P. 2000. Métodos estadísticos. Un enfoque interdisciplinario. Trillas. 463-531 pp.

Smith, M. 2000. The application of climate data for planning and management of sustainable rainfed and irrigate crop production. Agric. For. Meteorol. 103:99-108.

Tojo, S. C. M. César, S. P. and Hoogenboom, G. 2007. Application of the CSM-CERES Maize model for planting date evaluation and yield forecasting for maize grown off-season in a subtropical environment. Eur. J. Agron. 27:165-177.