International Conference on Hydroinformatics

2014

Rainfall Interpolation Models Obtained By Means Of Evolutionary Computing

Maritza L. Arganis

Margarita Preciado

Katya Rodriguez

Alfredo Ocon

Ramón Domínguez Mora

RAINFALL INTERPOLATION MODELS OBTAINED USING EVOLUTIONARY COMPUTING

PRECIADO, MARGARITA[1], ARGANIS MARITZA[2], OCÓN ALFREDO[1], DOMÍNGUEZ RAMON[1], RODRÍGUEZ KATYA[3]
(1): Instituto Mexicano de Tecnología del Agua, Jiutepec, Mor., México
(2 and 3): Instituto de Ingeniería2, IIMAS3, Universidad Nacional Autónoma de México, México, Edificio 5 Cub. 403 Avenida Universidad 3000 Coyoacán 04510 D.F., México

ABSTRACT

Complete missing records have a great importance in hydrology problems. In this paper we applied genetic algorithms and genetic programming models for precipitation in Cutzamala River Basin, Mexico. Such models depend on the geographical coordinates and altitude of climatologic stations currently operating. The obtained models were applied to interpolate 24 hours rainfalls for three analyzed storms. The genetic programming model (GP) outperformed the genetic algorithm (GA) in the estimation of 24 hours precipitation estimated on site, with respect to the measured data.

INTRODUCTION

Information of rainfall in a watershed is very important for hydrological studies; thus, flows that can occur at intermediate points and to the basin exit can be estimated. Furthermore, precipitation may be related to other climatological data for climate analysis behavior, also there are lot rainfall interpolation techniques, like highlighting the Kriging type (Bargaoui y Chebbi [2]). In this paper precipitation patterns were obtained depending on the latitude, longitude and altitude of weather stations in Cutzamala Mexico river Basin, using both genetic algorithms and genetic programming, which are tools of evolutionary computation used in several engineering problems since the mid-eighties of the past century (Rodríguez et al. [8], Preciado et al. [7] Arganis et al.[1].

METHODOLOGY

Genetic Algorithms

Inspired in Darwin's evolution theory, these algorithms were first purposed by Holland [5]; they are robust and in few steps (generations) a near-optimal solution can be reached. This method considers an initial population of *Nind* individuals randomly created by taking into account the search interval for each unknown variable, each individual represents a solution whose fitness is verified by means of an objective function. The best individuals are selected and the genetic operators, crossover and mutation, are applied in order to get a new population of *Nind* individuals that represents the next generation. The process is repeated until a defined number of generations is reached. The best individual in the last generation represents the best (optimal or near-optimal) solution to the problem. In this study nonlinear models were assumed to get their parameters with genetic algorithm (GA).

Genetic Programming

The genetic programming (GP) algorithm (Cramer [4], Koza [6], Banzhaf et al. [3]) traditionally involves the random generation of an initial population of trees formed from a functions set and variables according to the problem to be solved. The objective function is defined in order to evaluate the performance of each individual, subsequently, as in the case of genetic algorithms (Goldberg, 1989), individuals with the best performance are randomly selected, and crossover, reproduction and mutation operators are applied to generate new individuals which represent the next generation. In this case, each individual represents a mathematical model. In this study the set of operators and variables considered were precipitation (*hp*) as the dependent variable and the independent variables were: latitude *x*, longitude *y*, altitude *z*. The operators considered were both arithmetical: +,-,*, / and transcendental: sin, cos, exp.

Objective function

The objective function was defined as the minimization of the mean square error between the measured (hp) and the calculated ($\widehat{hp}$) rainfall, where *n* is the number of gathered data.

$$OF = \min\left(\frac{1}{n}\sum_{i=1}^{n}(hp_i - \widehat{hp_i})\right) \quad (1)$$

APPLICATION AND RESULTS

Study site

Cutzamala river basin is located in the center of Mexico (Figure 1), it represents an important source of water supply for the Mexican Republic capital.

Three biggest storm events were considered and they were registered by the stations shown in Table 1: Storm 1 on July 29th 2006, storm 2 on August 23th 2006 and, storm 3 on September 2th 2006.

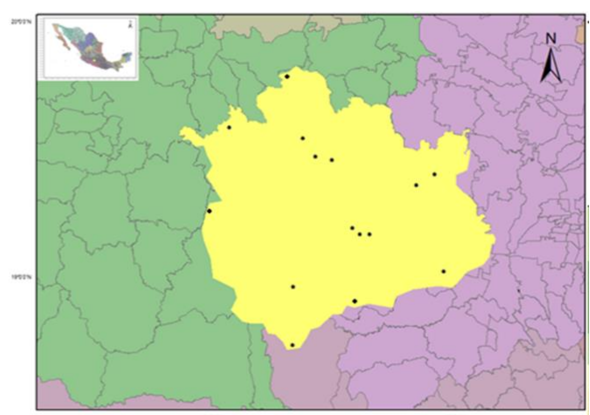

Figure 1. Cutzamala River Basin with the considered climatologic stations

Table 1. Climatologic stations. Cutzamala Basin, México

Number	**Name**	**State**	**x** **m**	**y** **m**	**z** **m**
12083	Tehuehuetla,s.m.totolapa	Gro.	351628.128	1970494.05	1250
12166	San miguel totolapan,	Gro.	353718.908	2009212.46	280
12141	Tlapehuala, tlapehuala	Gro.	341401.253	2016613.73	275
12019	Ciudad altamirano	Gro.	323824.938	2026063.37	250
12163	Cutzamala de pinzon	Gro.	338129.608	2042539.92	265
12036	El gallo, cutzamala de p	Gro.	324250.674	2072106.18	400
15046	Presa colorines, (cfe)	Mex.	375556.949	2112312.71	1680
15371	Ixtlahuaca, ixtlahuaca	Méx.	415756.25	2091834.41	2174
15353	Buenavista (estancia v.)	Méx.	391272.188	2101146.58	2576
16122	Susupuato de guerrero,	Mich.	351012.706	2121681.65	1560
15140	P.chilesdo,v.de allende	Méx.	378999.261	2139625.99	2395.95
15265	Camp. Berros, san jose	Méx.	465338.326	2138370.8	2150
16514	Jaripeo, la punta	Mich.	342540.305	2151306.61	1300
16002	Agostitlan, cd. Hidalgo	Mich.	330342.522	2160605.03	2380
16136	Tzitzio, tzitzio	Mich.	298919.735	2166464.93	1850
16235	Huajumbaro, cd. Hidalgo	Mich.	318319.412	2175557.13	2285

For analysis purposes, the climatologic station “15046 Colorines” was selected for this study and it was removed from the models estimations, in order to perform an interpolation of the rainfall data at that point for validation. The results for each analyzed storm are presented following.

Storm 1
Genetic Algorithm
The obtained model was:

$$\widehat{hp} = 3003.0213x^{-71.6422} + 7229.5663y^{-1.3642} + 20.8289z^{-0.4395} \qquad (2)$$

In Figure 2, the objective function value obtained with the optimal solution is presented. In Figure 3, it appears the comparison between the measured and the calculated data obtained in this case against the identity function and its linear correlation. It can be noticed a correlation coefficient value of 0.3933. Figure 3 also shows the result in the station that it was removed; in this case there was an overestimation of 8.79 mm on its value.

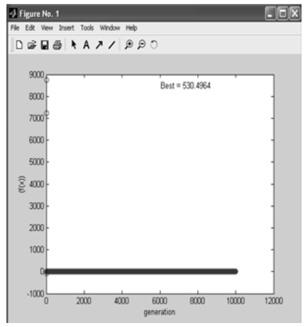

Figure 2. GA Storm 1.Value of the objective function for the optimal solution

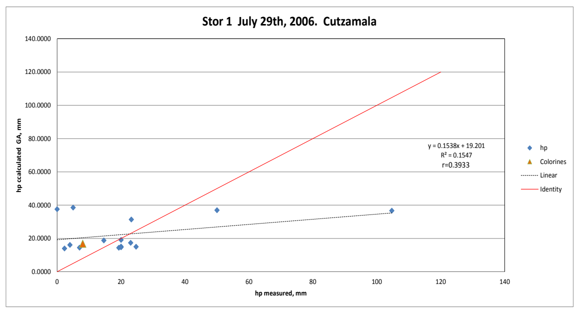

Figure 3. GA Storm 1. Comparison of measured, calculated and interpolated data

Genetic programming

By means of the use of genetic programming, the model with best performance was:

$$\widehat{hp} = \exp\left\{\exp\left[\sin\left(\sin\left(\sin\left(\frac{z}{0.713368} - y\right)\right)\right) + \cos\left[y - \left[\exp\left(\sin\left(\sin\left(\frac{0.868975}{0.564769} - y\right)\right)\right) - 0.717702\right]\right]\right] - \cos(0.611312\cos(z))\right\} \qquad (3)$$

The objective function had a value of 21.7159 with this model (Figure 4). The comparison between the measured and calculated data with GP model against the identity function is presented on Figure 5. The correlation coefficient in this case was 0.7842, which represents a better approach than the one obtained with GA. The calculated data in Colorines station (Figure 4) using the GP model was also overestimated but with a difference of 4.93 mm almost a half of the result obtained with GA. From equation 3, it is remarkable the dependence of the rainfall with the altitude and the latitude y.

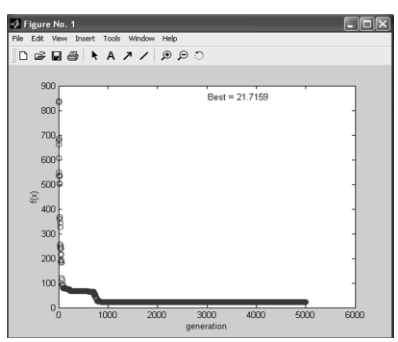

Figure 4. Storm 1.Value of the objective function for the best GP individual

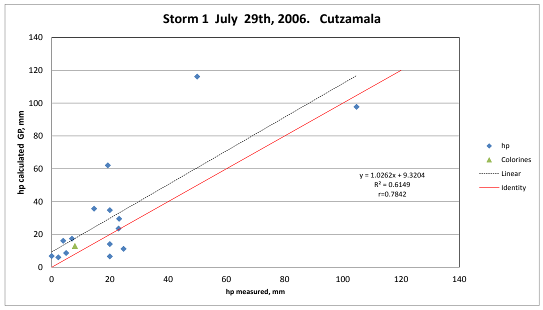

Figure 5. GP Storm 1. Comparison of measured, calculated and interpolated data

Storm 2

Genetic algorithm

For the storm 2, the best obtained model with GA was:

$$\widehat{hp} = 4420.603x^{-1.67241430} + 9986.74580y^{-1.05594040} + 10.348827z^{-0.43946654} \quad (4)$$

The objective function values obtained with the best individual are shown on Figure 6. The comparison of the measured an calculated data with the GA model against the identity function (Figure 7) shows a correlation coefficiente of 0.1581, and the differences between the measured and calculated data for the Colorines Station recorded a difference of -13.46 mm, that means an understimation.

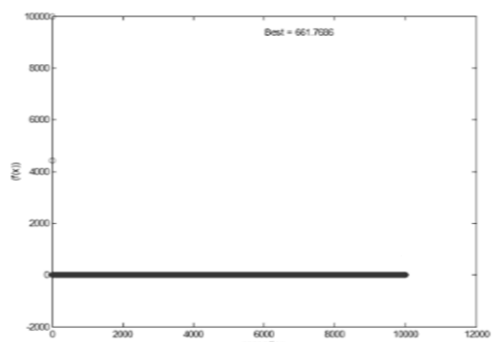

Figure 6. Value of the objective function for the best individual (GA storm 2)

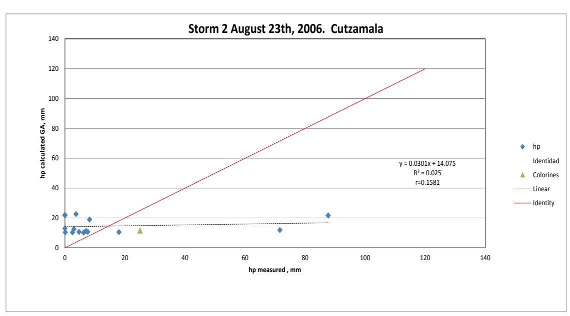

Figure 7. Comparison of measured, calculated and interpolated data

Genetic programming

The best performance model obtained with GP was:

$$\widehat{hp}=\frac{y-\frac{\left[\frac{\cos(\cos x)}{\cos(x-z)*\left[\frac{\left|\frac{\cos(y)}{0.03603254y}\right|}{\exp(\cos(\exp(x)))}\right]}\right]}{(y\cos(z)-2z)}}{z} \qquad (5)$$

The behavior in the objective function is shown in Figure 8 for the best individual. In this case, a higher correlation coefficient of 0.9734 was found between measured and calculated data (Figure 9), but, the approach for the removed data in “Colorines Station” resulted in a difference of 30.60 mm, there was and underestimation, so the model did not make a good estimation in this missing data.

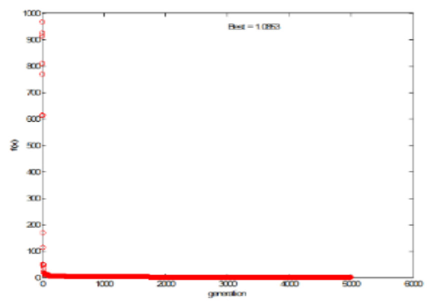

Figure 8. Value of the objective function for the best individual. GP Storm 2

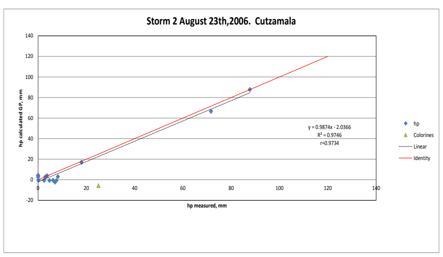

Figure 9. Storm 2 GP. Comparison of measured, calculated and an interpolated data

Storm 3

Genetic Algorithm

$$\widehat{hp} = 1746.6475x^{-1.67241430} + 9999.6918y^{-1.0559404} + 15.897092z^{-0.13122959} \quad (6)$$

The objective function values obtained are shown on Figure 10. The comparison of the measured and calculated data with the GA model against the identity function (Figure 11) shows a low correlation coefficiente of 0.0096; whereas, the differences between the measured and calculated data for the Colornes Station recorded a difference of 9.94 mm, that means an overestimation.

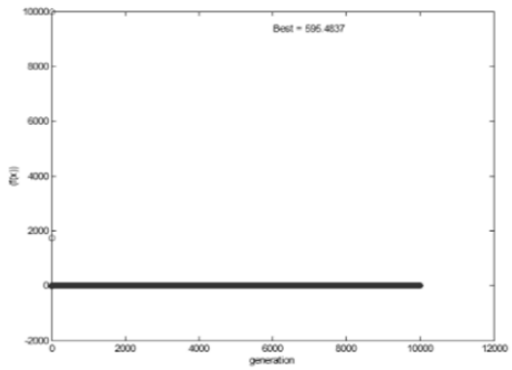

Figure 10. GA Storm 3: Value of the objective function for the best individual

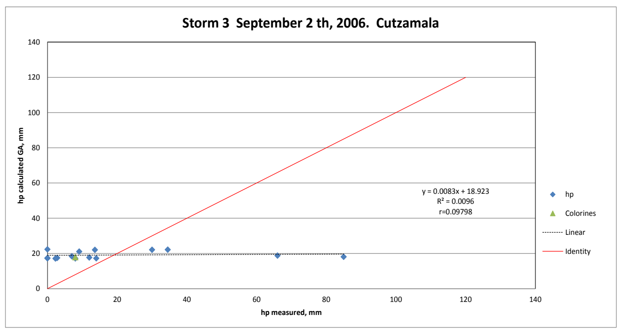

Figure 11. Comparison of measured and calculated data as well as an interpolated data. GA Storm 3

Genetic programming

$$\widehat{hp} = \cos\,(y - z) * 18.4507537 - \left[\left(\left(0.077851 - \frac{1.09482135}{z}\right) - \exp\{\exp(\mathrm{sen}(z - \exp(z)) * (0.267723 - z))\right) * \exp(\cos(\cos(z - 0.671812)))\}\right] \quad (7)$$

The behavior of the objective for the best individual is shown in Figure 12. In this case, the same correlation coefficient of 0.5626 was found between measured and calculated data (Figure 13), but, the approach for the removed data in “Colorines Station” resulted in a difference of 12.09 mm, there was an overestimation, so the differences were bigger than those obtained with genetic algorithm.

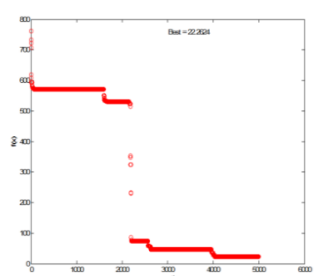

Figure 12. Value of the objective function for the best individual. GP Storm 3

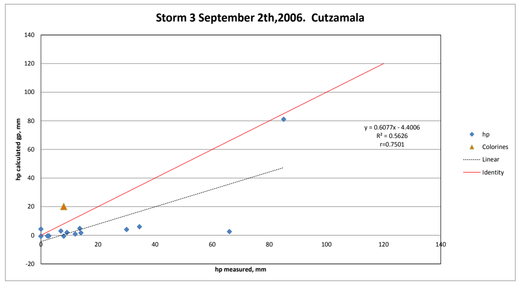

Figure 13. Comparison of measured and calculated data as well as an interpolated data. GP Storm 3

We can notice than in equation 6 the resulted rainfall dependent only with the *z* (altitude) and *y* (latitude) variables.

Conclusions

The obtained GP models showed a weak dependence from the longitude *x* of the rainfall in the Storms 1 and 3. GA models had more difficulties to reproduce the rainfall events whereas GP models improved the correlation coefficients but any of them could get a good approach for the recorded data in “Colorines Station”. Several nonlinear models must be proposed to get their parameters with GA in order to get better estimations for this phenomenon.

References

[1] Arganis, J. M. L., R. Val S., J. Prats R., K. Rodríguez V., R. Domínguez M., J. Dolz R., "Genetic programming and standardization in water temperature modelling," *Advances in Civil Engineering*. Hindawi Publishing Corporation. Article ID 353960, (2009) pp 10.

[2] Bargaoui, Z.K and A. Chebbi., Comparison of two Kriging interpolation methods applied to spatiotemporal rainfall. *Journal of Hydrology*, (2009), 365 56–73

[3] Banzhaf, W., P. Nordin, R. E. Keller, and F. D. Francone, *Genetic Programming: An Introduction: On the Automatic Evolution of Computer Programs and its Applications*, Morgan Kaufmann Publishers,USA, (1998).

[4] Cramer, N.L.,”*A Representation for the Adaptive Generation of Simple Sequential Programs”*. In Proc. of Int. Conf. on Genetic Algorithms and the Applications (Grefenstette, J.J., editor), (1985), pp. 183-187

[5] Holland, J.H., “*Adaptation in Natural and Artificial System”*, The MIT Press, USA, (1975).

[6] Koza, J. R., Hierarchical Genetic Algorithms Operating on Populations of Computer Programs. In Proc. (1989)

[7] Preciado, J. M., M. Arganis J., A. Ocón., “*Aproximación de la función de distribución empírica bivariada de las avenidas históricas máximas de ingreso a una presa usando programación genética”*. Memorias del XXII. Congreso Nacional de Hidráulica. Del 7 al 9 de Noviembre . (2012).

[8] Rodríguez V. K., M. L. Arganis, J., C. Cruickshank V., R. Domínguez M.,“Rainfall-runoff modelling using genetic programming”. *Journal of Hydroinformatics*. Volumen 14 No. 1, (2012). pp 108-121.