DETERMINATION OF DIGITAL ELEVATION MODEL IN A PLAIN OF MEXICO USING GENETIC PROGRAMMING AND ESTIMATION OF FLOOD VOLUMES.

Mariles Gutiérrez Jacquelinne, Arganis Juárez Maritza Liliana, De Luna Cruz Faustino, Preciado Jiménez Margarita, Carrizosa Elizondo Eliseo.

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 equations of the digital elevation model surface area of a plot of Villa de Ocuiltzapotlán, Tabasco, using the genetic programming algorithm were estimated. The determination of the digital elevation model made possible to compare the results between the measured and calculated level curves. The algorithm considered simple and transcendental arithmetic operators, in addition to using one, five and even ten rules of correspondence. Additionally, with the help of both surfaces, areas prone to flooding caused by extraordinary events were identified, evaluating and comparing the results between both surfaces showing good agreement.

Introduction

The problem of having approximate digital terrain models through mathematical models is relevant in engineering, since the availability of a model of some surface allows to simulate processes and to experiment in the modeled surface independently of the real surface, avoiding risks and also whit the possibility of repeating indefinitely the experimentation.

Since the end of the last century, interest in the study of Digital Elevation Models (DEM) has been increasing due in large part due to the ease of working with the study surface in laboratory tests that may report a behavior similar to that of the real surface in case of extraordinary events or that can predict the behavior by the same events.

However, the complexity of the topography makes the models obtained based on their mathematical representation have no more than a symbolic meaning, so in practice, the dimensions of a zone are given by subregions obtained from exclusively applicable equations For that zone.

As a consequence, the studies carried out have focused on finding the way in which these models, rather than being obtained from a single mathematical function, represent the study surface more faithfully.

In the field of hydrology the problem of surface approximation has always been present, so that since the last century has work has be done to obtain digital models of surfaces using various techniques.

Previously, in Mendoza R. et al. (1996) it was necessary to obtain an explicit mathematical expression to later derive it and to construct a model of curvilinear coordinates. The problem was solved with a development of the Taylor series and least squares fit with the DFP (developed by Davidon, Fletcher and Powell) optimization method, the results obtained for that study were mostly satisfactory, however, it was found that in some points the estimation model threw errors not so acceptable, since they were able to observe magnitudes greater than 1 m which were more evident in areas within the site of study in which they were valleys or crests, which were surrounded by data of quota were quite different.

In order to improve the results obtained in Mendoza (2002), it was decided to use evolutionary computation (EC), specifically Genetic Programming (GP) applied in an area of the Colorado River, this decision significantly improved the errors obtained previously recaude the estimation was less than 10 cm.

Subsequent to this and trying to raise the difficulty of the region previously estimated, in Mendoza et. Al., (2007) studied another region within the same Colorado River basin but with an even greater degree of complexity, ie, where the topography represented a greater challenge to apply the PG, the results obtained were good, since there were errors smaller than 2.5 m where the points could not have a better fit to the terrain curvatures.

Recently, during the Fifth Regional Symposium on Hydraulics of Rivers carried out in Argentina, Rodríguez et. Al., (2011) presented a numerical model of the bifurcation of the Mezcalapa river located in Tabasco using genetic programming and from measurements obtained in the field by means of a topographic survey; this work was carried out with the objective of studying the processes of erosion and sedimentation of channels by the operation of a control structure located on the Carrizal river, almost 1 km downstream of the bifurcation, which was created from the Need to reduce the material damages and economic and human losses caused by the atypical rains presented in the region, the purpose was to obtain an alternative way to study and quantify these processes in the vicinity of the bifurcation, making a numerical model that would allow To visualize the surface of the channels in that zone.

Although the study had an abrupt topography where valleys peaks predominated, the results of the model are mostly good, with average errors up to 1.05 m, due to the fact that in areas where the information did not help to make a good estimate, the model softened the peaks or valleys.

Study site.

The study site of Villa de Ocuiltzapotlán is located in Tabasco state (Figure 1), located geographically southeast of the Mexican Republic, from the coastal plain of the Gulf of

Mexico to the mountains of northern Chiapas, can be defined Geographically between 17 ° 15 'and 18 ° 39' north altitude and 91 ° 00 'and 94 ° 07' west longitude.

The limits of the state are natural and artificial, since to the north it limits with the Gulf of Mexico and Campeche, to the south with Chiapas and Guatemala, to the west with Veracruz and to the east with Campeche and Guatemala.

The average annual temperature in the state is 27 ° C, the average maximum temperature is 36 ° C and the minimum average is 18.5 ° C. The average state rainfall is 2 550 mm annually. Although rains occur all year, they are more abundant in the months of June to October.

Figure 1. Tabasco State, México location. Source: INEGI

Tabasco is divided into 17 municipalities, one of them is the municipality of Centro, where the city is located and this is one of the 13 regional development centers (RRDC) of the municipality. Villa Ocuiltzapotlán (figure 2) is defined by the historian Rosendo Taracena Padrón as "Tierra de Zapotes", located approximately 18 km north of the state capital, Villahermosa, and forms part of the Metropolitan Area of Villahermosa, presents the appearance of a vast plain severed by stretches by some low hills and is at an average height of 10 meters above sea level, the predominant climate is extremely hot and humid, the average annual temperature is 26 ° C and it presents constant rains, its precipitation is of 2,750 Mm. Spring is dry in this region and in summer rains with are more intense, those are the torrential rains known as squalls, which covers autumn and winter.

The Villa Ocuiltzapotlán is irrigated by the lagoon Paso Segundo, lagoon of the Coco, and the rivers Gonzales (today river Azolvado), Garduza and Medellín. Figure 2 shows the topographic map of the site obtained through the electronic portal of the National Institute of Statistics and Geography (INEGI).

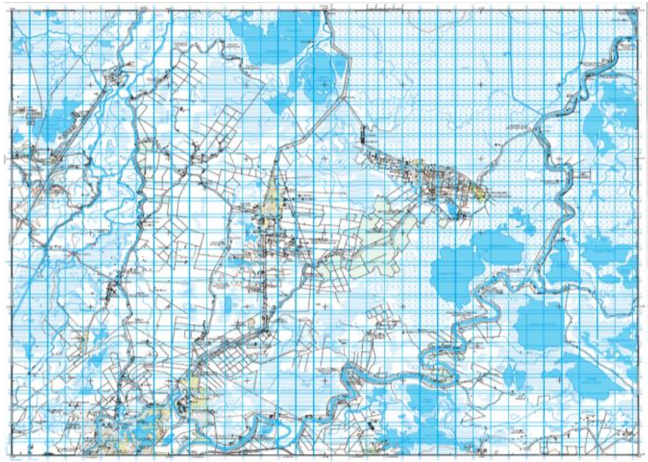

Figura 2. Topographical chart of Villa de Ocuiltzapotlán, Mexico. Source: INEGI

Methodology

Genetic Programming (GP) emerged as an evolution of Genetic Algorithms (GAs). It consists in the automatic evolution of programs using ideas based on natural selection (Darwin), it is itself a tool of evolutionary computation that allows to obtain mathematical models from measured data. By means of a genetic programming algorithm, mathematical models of one or more variables can be determined from previously known data as a method of optimizing an objective function similar to how a simple genetic algorithm works (Arganis et al., 2014).

The considered "father" of the GP is John R. Koza, and he was who assigned the term "Genetic Programming" in 1992 (Koza, 1992). This method belongs to the generic set of evolutionary computation methods, the best known examples are genetic algorithms (Holland, 1975).

The classic search paradigm that looks more like the GP is the "beam search" (Tackett, 1994), although it has certain peculiarities of its own. Like beam search, the GP maintains a finite population of possible good solutions or solution candidates (called individuals).

Like any search algorithm, the GP searches in a space of possible solutions represented in a certain way, it has search operators and a heuristic function that guides it, called fitness function. This function evaluates the goodness of each of the individuals in the population. Typically, the fitness function returns a numerical value or also a value vector, if multi-objective optimization is being done. This value is used to determine the probability of selecting each individual from the population, which is normally proportional to the value returned by the adequacy function.

The approximation functions are represented by trees, where each node is a data or an operator. The maximum number of nodes in this tree is a design variable. The number of branches that arise from each node depends on the number of operands that the operator requires in that node.

The search operators of the GP are the so-called genetic operators: reproduction, crossing and mutation, although generally only the first two are used.

The reproduction operator simply creates an individual exactly the same as the one passed as an argument using functions and terminals. The reproduction operator, or also called a crossover, chooses two individuals, randomly selects a node in each of the progenitor individuals and exchanges the two corresponding subtrees. The mutation operator selects a node in the parent tree, chops off the subtree that depends on that node, and replaces them with a randomly generated subtree.

In general terms, the methodology for generating topographic surfaces using Genetic Programming works in such a way that one or more mathematical functions are applied to the topographic location data (X and Y coordinates) to produce an estimate of the elevation.

The process begins by choosing a fixed number of individuals (data) to constitute the initial population of the algorithm. This number is less than or equal to the total number of topographic data. The size of the population has to be as little as possible by the resolution time to find the best solution to the problem.

The method works through iterations that are applied to all the individuals of the population, where the different levels are estimated applying the indicated operations found by each one of the generations, when a new generation starts the performance of each individual is evaluated, in this way the best individuals preserved will be those with which an estimated quota near the real quota is obtained, the rest of the individuals are applied again the genetic operators with the chosen probabilities. The method continues to iterate until some given error criterion is satisfied or until the maximum number of iterations is reached.

In each iteration of any global method we try to reduce the approximation error for all the points of study, while in GP we take a set of random points of the study region and choose as candidates those that minimize the performance function local. By preserving best-performing points, introducing new random points, and using cross-over and mutation operators over tuning functions, GP ensures that you can explore the search region globally.

Considered parameters

• The algorithm of the program developed by researchers of the Institute of Research in Applied Mathematics and Systems, UNAM, was codified in the MATLAB interpreter, which helped to obtain the necessary data of the corresponding models.

• To apply the GP technique, four arithmetic functions were considered: addition, subtraction, multiplication and division. And two transcendental functions: cosine and sine; Additionally a vector of random constants obtained by the algorithm was considered.

• The maximum number of nodes (number estimated between operators and operands) was 25.

• Generations, data and waiting time for resolution by the GP limited the work, so it decided to perform 1000 generations.

• The notation with which the GP is developed is the prefix or Polish, which consists of placing the operator before the operands, thus avoiding the use of parentheses in mathematical expressions. For example, to express the sum of two quantities A and B with the Polish notation or prefix is: + A B (Morris, 1994).

Estimation of flood volumes

The estimation of floodable volumes will be based on the comparison of floodable volumes between the real and the modeled surface, from a Grid of dimensions, in order to obtain a graph of comparison between them.

Results

The calculated dimensions were obtained by different models applied to topographic data of the Digital Elevation Models (DEM) provided by the National Institute of Statistics and Geography (figure 3).

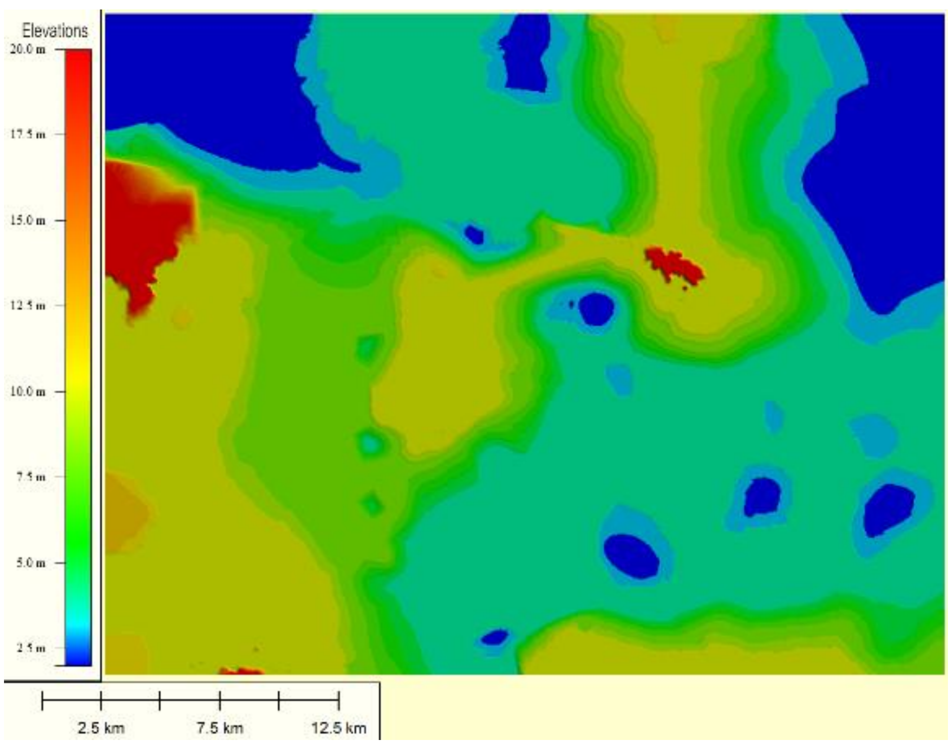

Figure 3. Ocuiltzapotlán, México DEM

In the first instance we worked with the 392,122 points that reproduce the surface of the region, using the four arithmetic functions and obtaining, then, a first and only rule of correspondence for all the measured data with which the estimated elevation data were obtained.

The best results showed that the mean error was 9.96 m, so that the calculated levels reported values between 3.99 m and 9.55 m, these values can be seen reflected in table 1, as well as the surface obtained in figure 4.

The next mathematical model was obtained according to the first report of the GP:

$$z = \left(\frac{y * 0.514}{x - A}\right) + \left(\frac{-2.185 - x - y}{x}\right) \qquad (1)$$

$$A = -0.121 * (x - 2y + 9.558)$$

Table 1. Summary of results

X [m]	Y [m]	Z [m]	Zcalculated [m]	Error [m]
535275	1990225	8	3.99	4.00
499975	2017925	2	9.55	-7.55

The surface was generated from the function of equation (1) that was obtained from the GP and compared to the real surface; the result of the generated surface is shown in figure 4, where the values obtained are not in any way similar to the real surface. This was due to the use of all the data and that a correspondence rule was obtained that was not able to reproduce the model satisfactorily and therefore did not obtain a resemblance to the original model.

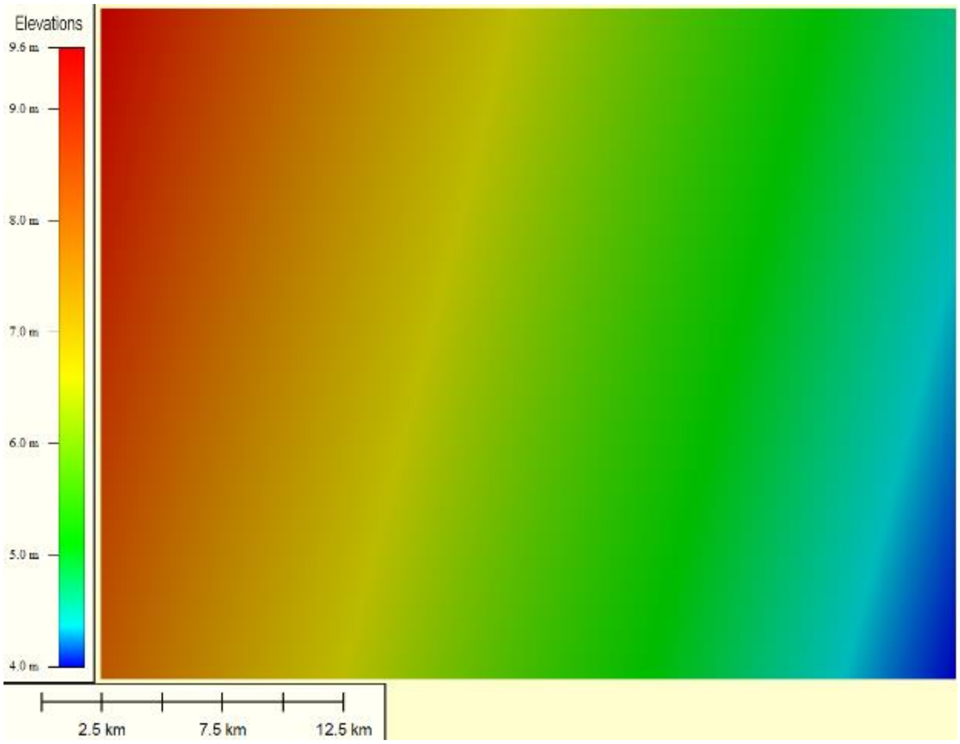

Figure 4. Area calculated from a rule of correspondence

After that, the number of correspondence rules was increased to five (Table 2), to obtain small areas where each of the equations obtained with the algorithm were applicable, besides that only arithmetic functions were used, it was observed that the error obtained was lower, however the surface generated from the results was not adequate (figure 5).

Tabla 2. Rules of correspondence

Rule of correspondence	Error intervals	
1	-7.55	-3.10
2	-3.10	1.35
3	1.35	5.80
4	5.80	10.25
5	10.25	14.70

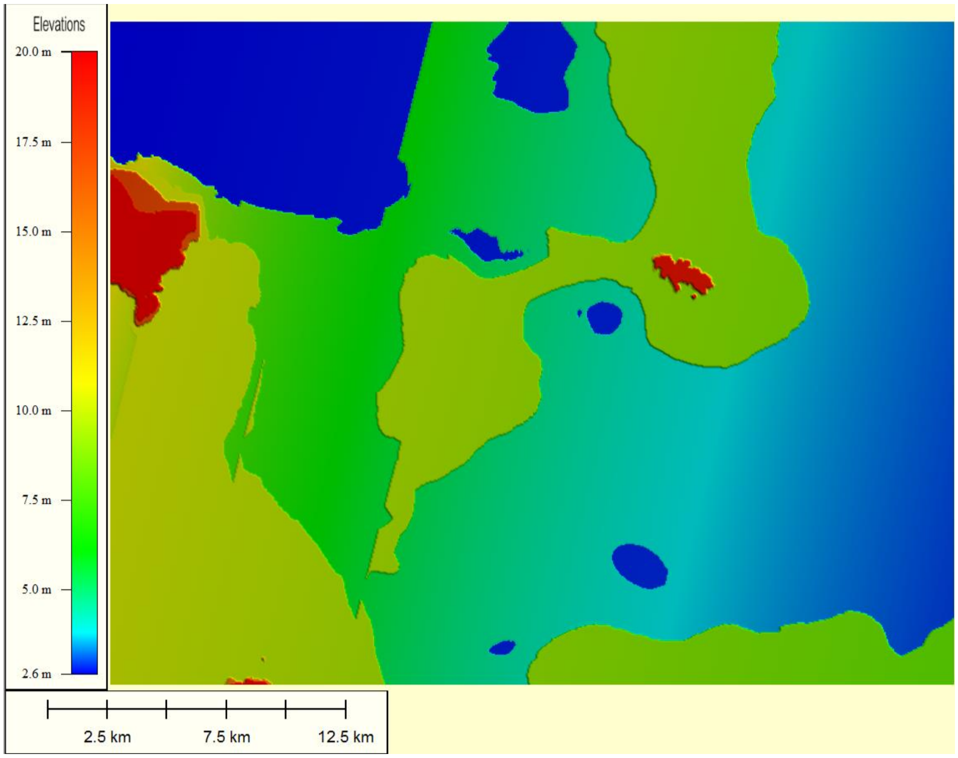

Figure 5. Surface calculated from 5 rules of correspondence.

The shape of the surface suggested that the sinusoidal and cosine trigonometric functions could be a new starting point, so that with the same number of matching rules, new equations were obtained, but this time using only 3 arithmetic functions (addition, subtraction and multiplication) and the two transcendental ones, the result corrected the generated surface, since it was observed how the reproduction was more similar to the real one, nevertheless the error did not present greater improvement , and in some cases, the algorithm used reported constants (figure 6).

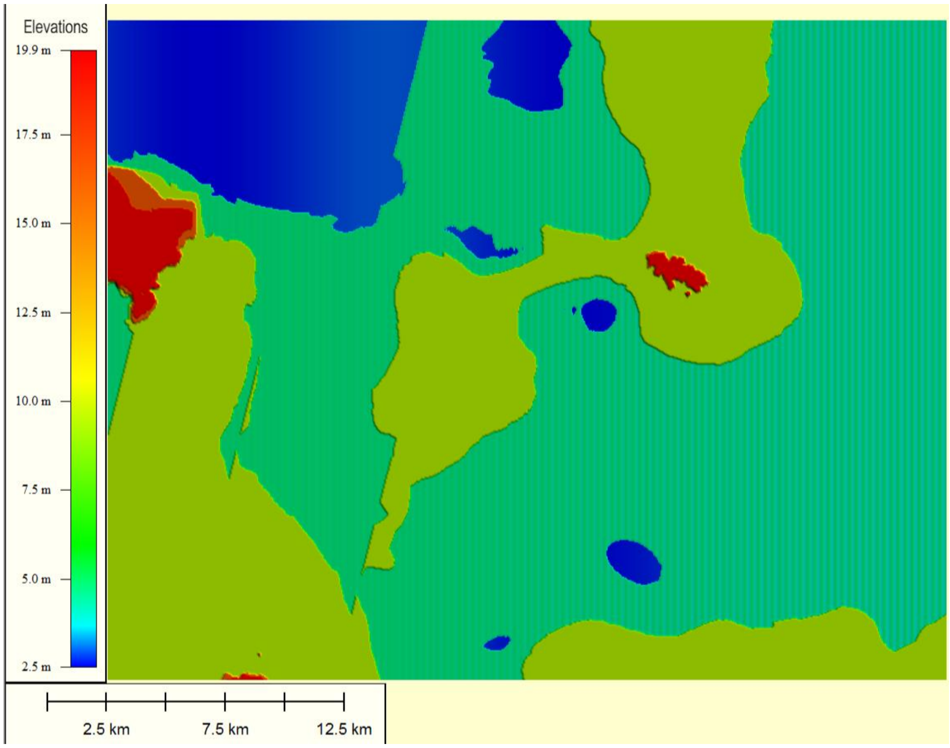

Figure 6. Surface calculated from 5 rules of correspondence and trascendental functions.

While observe the error decreased because of the division of the domain with respect to the obtained errors, the domain was divided into ten correspondence rules (Table 3) to minimize errors and to obtain a better modeled surface.

Table 3. Rules of correspondence

Rules of correspondence	Error intervals	
1	-7.55	-5.3
2	-5.33	-3.10
3	-3.10	-0.88
4	-0.88	1.35
5	1.35	3.58
6	3.58	5.80
7	5.80	8.03
8	8.03	10.25
9	10.25	12.48
10	12.48	14.70

The modeling surface was obtained using three arithmetic functions (addition, subtraction and multiplication) and the two transcendental ones (sine and cosine) for each of the rules of correspondence. In most cases, the best results the algorithm used reported constants, however the average errors obtained were much lower than those obtained previously.

Figure 7 below, it shows the close relationship between the measured data and the calculated data, which would ensure a better modeling surface due to the high correlation coefficient obtained between the data.

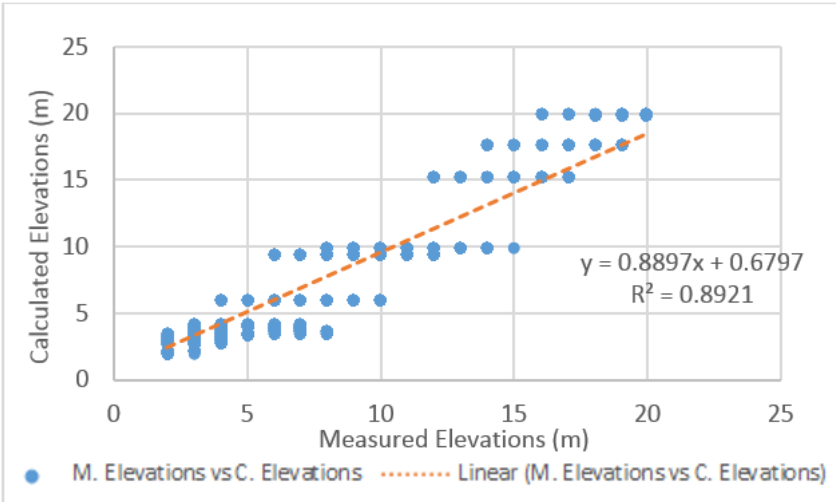

Figure 7. Measured elevations vs calculated elevations

The modeled surface in Figure 8, shows where the true resemblance between the real surface and the modeled surface of the study region is well appreciated.

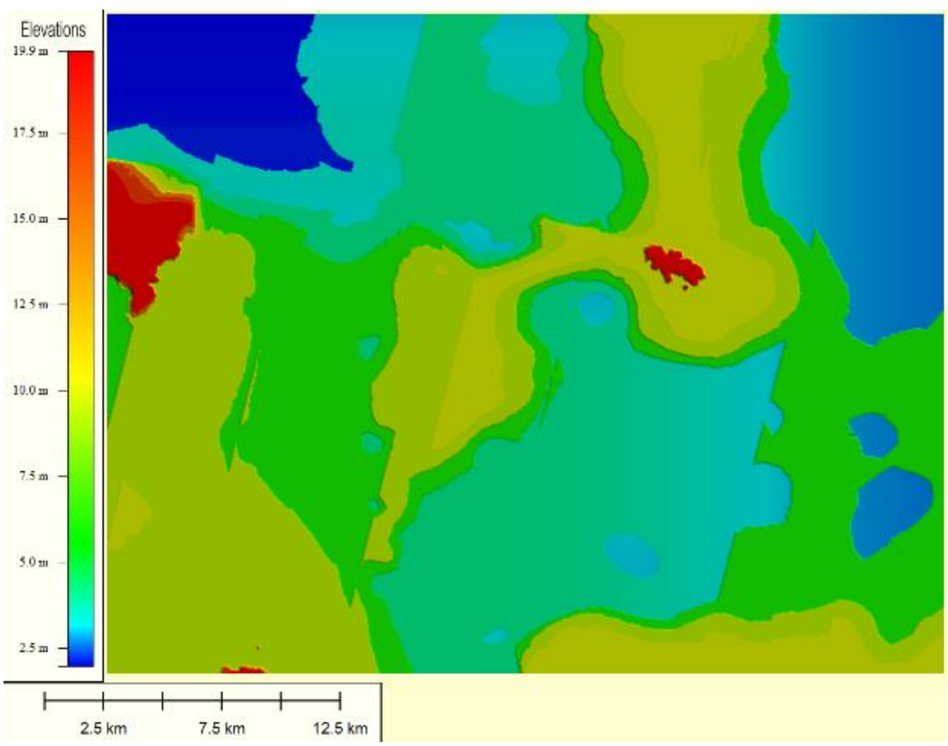

Figure 8. Area calculated from 10 rules of correspondence.

XVI
World Water Congress
International Water Resources Association (IWRA)
Cancun, Quintana Roo. Mexico. 29 May- 3 June, 2017

The details of the correspondence rules can be consulted with the main author at the Institute of Engineering, UNAM.

Flood plans estimation.

From the identification of the GP algorithm that reported a better surface resemblance to the actual surface, it was decided to confirm the reliability of data to be used later in works necessary for the hydrological field mainly and within which they have related studies To areas of flood.

This identification was made by obtaining volumes of flood in both models, from which the following graph (Figure 9) was obtained, in which it is observed how the volumes are similar and therefore the reliability of data also.

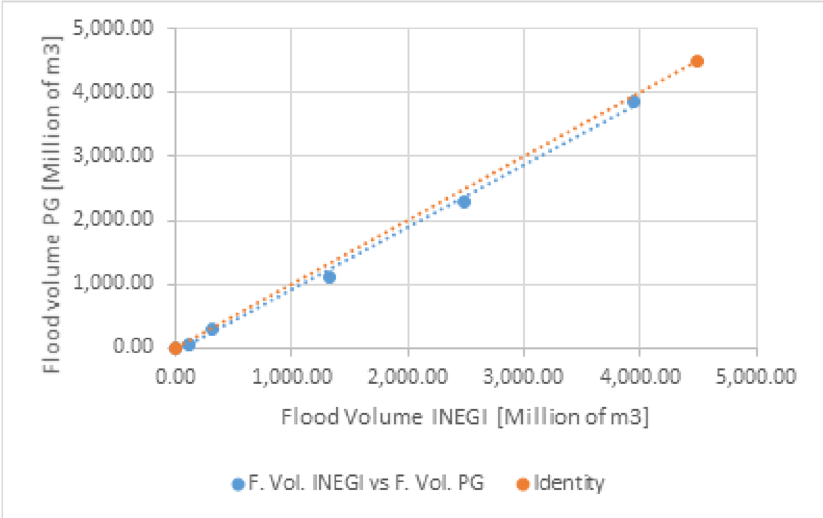

Figure 9. Comparison of flood volumes

Conclusions and recommendations

The application of Evolutionary Computing (EC) and more specifically the subject of Genetic Algorithms (GA) and Genetic Programming (GP) is a useful tool to estimate the topographical dimensions of the study area, which is a plain.

Working with all data allowed to obtain a single rule of correspondence, which made it difficult to find a model able to reproduce satisfactorily the entire portion of surface analyzed.

The restriction of the data to be used for each of the sections in which the study site is divided is fundamental in the resolution and search of a digital model of elevations, due to the large number of points to be considered to represent the surface Analyzed.

The different mathematical expressions obtained are easily programmable for evaluation in any programming language.

Because different rules of correspondence were used, it is observed that those where the algorithm uses transcendental functions for the representation of regions within the zone the similarity with the real surface is still much closer to comparison of when the subregions are modeled from of constants obtained with the same algorithm. In spite of this, the maximum errors obtained with GP were values of 10 cm.

With this surface and with the original it was possible to obtain the comparison of elevation dimensions that helped to obtain a difference of elevations that is useful to locate the areas of flooding from this difference and that will serve in later studies to have the certainty of the handling of a digital elevation model with respect to the actual surface.

Given the final results it is advisable that in later works, where in addition to continue studying the goodness of the program with this type of plains or surfaces with a more rugged topography, divide the study area into smaller regions, as in this investigation, allowed us to approach a more real value and obtain errors smaller than 1.2 m of difference with respect to the original surface. It is also advisable to make sure that the highest points of the real surface are part of it and not buildings because not recognizing them involves handling false information from the site.

References

1. Aler M, R., (1999) Programación genética de heurísticas para planificación, Tesis doctoral, Universidad Politécnica de Madrid, Facultad de informática, Madrid.
2. Arganis J. M. L., Preciado J. M. E., Herrera A. J. L., Rodríguez V. K. (2014) Función Bivariada de avenidas del conjunto Temascal – Cerro de Oro con Programación Genética, XXIII Congreso Nacional de Hidráulica, Jalisco, México.
3. Fallas, J., (2007) Modelos Digitales de elevación: Teoría, métodos de interpolación y aplicaciones.
4. Felicísimo, A. M. (1994) Modelos Digitales del Terreno, introducción y aplicaciones en las ciencias ambientales, Oviedo, España.
5. Holland, John H. (1975) Adaptation in Natural and Artificial Systems. The University of Michigan Press.
6. Instituto Nacional de Estadística y Geografía (INEGI) (2017). Website: http://www.inegi.org.mx/
7. Instituto Nacional para el Federalismo y el Desarrollo Municipal (2017) http://www.gob.mx/inafed
8. Koza, John R. (1992) Genetic Programming On the Programming of Computers by Means of Natural Selection. MIT Press, Cambridge, MA, USA.
9. Mendoza, R., Alarcón, P. y Berezowsky, M. (1996) Cálculo del campo de velocidades en cuerpos de agua con modelo matemático bidimensional en coordenadas curvilíneas adaptables. Proyecto CONACYT. Instituto de Ingeniería UNAM.
10. Mendoza, R. (2002) Aplicación de la computación evolutiva en la estimación de cotas topográficas. Tesis de maestría. UNAM.

11. Mendoza, R., Rodríguez, K. y Álvarez-Icaza, L. (2007) Generación de superficies mediante programación genética, ingeniería Hidráulica en México. UNAM.
12. Morris Mano, M., (1994) Arquitectura de computadoras, Tercera Edición, Universidad Estatal de California en Los Ángeles, Prentice Hall Hispanoamericana S.A., 565 páginas, México.
13. Rodríguez Vázquez K., Mendoza Ramírez R. y Jiménez Castañeda A., (2011) Modelo numérico de la bifurcación del Río Mezcalapa usando Programación Genética, Quinto Simposio Regional sobre Hidráulica de Ríos, Argentina.
14. Sarrías, Francisco Alonso, (2005) Sistemas de Información Geográfica, Capítulo 7, Universidad de Murcia.
15. Tackett, Walter Alden, (1994) Greedy recombination and genetic search on the space of computer programs. In L. Darell Whitley and Michael D. Vose, editors, Foundations of Genetic Algorithms 3, pages 271-297, Estes Park, Colorado, USA.
16. Walsh, R., (1960) Optimization, Mac-Graw Hill, 200 p.
17. Zúñiga, Ramón. Introducción al uso de MATLAB. Posgrado en Ciencias de la Tierra, UNAM.