DE GRUYTER OPEN

Mathematical Modelling in Civil Engineering
Vol. 13-No. 4: 27-38 - 2017
Doi:10.1515/mmce-2017-0012

GENETIC PROGRAMMING TECHNIQUE APPLIED FOR FLASH-FLOOD MODELLING USING RADAR RAINFALL ESTIMATES

Cristian DINU - Department of Hydrotechnic Engineering Technical University of Civil Engineering Bucharest, Romania, e-mail: kristian_dinu@yahoo.com

Radu DROBOT - Department of Hydrotechnic Engineering Technical University of Civil Engineering Bucharest, Romania, e-mail: drobot@utcb.ro

Claudiu PRICOP - Water Basin Administration Prut-Bârlad, Romania,e-mail:claudiu.pricop@dap.ro

Tudor Viorel BLIDARU - Water Basin Administration Prut-Bârlad, Romania,

Abstract: The rainfall-runoff transformation is a highly complex dynamic process and the development of fast and robust modelling instruments has always been one of the most important topics for hydrology. Over time, a significant number of hydrological models have been developed with a clear trend towards a process-based approach. The downside of these types of models is the significant amount of data required for building the model and for the calibration process: in practice, the collection of all necessary data for such models proves to be a difficult task. In order to cope with this issue, various data-driven modelling techniques have been introduced for hydrological modelling as an alternative to more traditional approaches, on the basis of their capacity of mapping out complex relationships from observation data. Having the capacity to generate meaningful mathematical structures as results, genetic programming (GP) presents a high potential for rainfall-runoff modelling as a data-driven method. Using ground and radar rainfall observation, the aim of this study is to investigate the GP technique capability for modelling the rainfall-runoff process, taking into consideration a flash-flood event.

Keywords: rainfall-runoff modelling, genetic programming, data-driven models, flash-flood, radar rainfall estimates

1. Introduction

Used for various tasks, both in offline (e.g. scenario analysis) and online mode (e.g. forecasting), hydrologic and hydraulic models have become key instruments in river catchment management, serving as support for the planning and decision-making process. The evolution in the development of rainfall-runoff (R-R) models went through different stages and according to the accepted hydrological classification there are three main categories of the R-R models: distributed physically based, lumped conceptual and empirical models. As a basic differentiation between the R-R models, the first two types of models usually imply the use of full or partial mathematical description of the physical processes which govern the hydrologic cycle (e.g. Système Hydrologique Européen - SHE model, Nedbør-Afstrømnings-Model, abbreviated as NAM model), whereas the empirical models make use of mathematical expressions (equations) derived from analysing the time series without any knowledge of the hydrological process (e.g. ARIMA models, linear regression models). The developments of various techniques from fields such as machine learning made possible, in the last two decades, the broadening of the hydrological modelling capabilities by empirical methods. Due to their abilities to identify relationships between input and output variables of a system (e.g. meteorological and hydrological data respectively), the collection of the new empirical methods employed for hydrological applications is known as data-driven modelling [1].

One of the most use data-driven techniques developed in machine learning and applied for hydrological modelling are the artificial neural networks (ANNs). These black-box models have the ability to map out existing relationships between a set of input and output variable and have

been suggested as efficient instruments for R-R modelling [2], [3] or [4]. The main issue with ANN is given by its black-box nature from which no insight or interpretation can be extracted regarding the underlying mechanisms of the analysed process.

Another technique that presents high potential in hydrological modelling is genetic programming (GP) [5], which can be catalogued as a population-based metaheuristic search algorithm. Using search mechanisms inspired from biological evolution and genetics is the latest addition to the family of others evolutionary computation (EC) techniques such as evolution strategies (ES), evolutionary programming (EP) and genetic algorithms (GA). Also referred to as symbolic regression, GP technique has the capacity to generate mathematical expressions suited to link input and output data. By generating functional mathematical structures, GP method has an advantage when compared to black-box or traditional regression models, in the sense that GP solutions can possible retain some physical meaning of the analysed process. Being a relatively recent technique its application in R-R modelling [6], [7], [8] occured on a smaller scale compared with ANNs.

Beside the progress of empirical modelling techniques, the developments of observation methods such as remote sensing or radar technology are providing new ways of determining the useful parameters for meteorological and hydrological application. The weather radar has become one of the most important instruments in identifying and forecasting rainfall position and intensity over an area. Due to the good spatial representation of rainfall, radar products are of great importance for rainfall-runoff modelling, particularly in catchments with a low density of ground-based monitoring network. In addition, if the analysed catchment has a fast response that can generate flash-flood events, radar products can deliver more reliable information about rainfall spatial distribution and intensity than the monitoring network.

In this paper a study focused on using the GP algorithm for developing mathematical structures able to simulate the R-R process with an event-based approach is presented. The main objective is to investigate the GP capabilities to generate, via symbolic regression, expressions that can be employed to simulate a flash-flood event measured in Bahluet watershed at Targu Frumos gauging station using three different types of data as input to the algorithm. The input datasets are based on ground rainfall measurements, radar rainfall observations and rainfall volumes derived from radar data [9]. The goal of using different input data is to identify which of them leads to better results with GP. As a general rule in R-R modelling, a better representation of the rainfall over the catchment will generate improved results but it may be possible that this aspect will not have a significant weight with symbolic regression technique. In order to present to PG algorithm some insight about the catchment the rainfall volumes were used. Rainfall volume contains aggregated information about the area of the catchment and the rain intensity. Based on this, it was supposed such type of data will deliver superior results with symbolic regression method.

2. Method

2.1 Genetic programming

Developed by Koza in the early 1990, GP is a type of evolutionary algorithm (EA) which has the ability to automatically solve a given problem without requiring in advance explicit knowledge regarding the solution’s form or structure [10]. Similar to all EAs, GPs are based on an abstraction of the natural selection principles and genetic recombination. The basic cycle of GP algorithm starts from an initial population (randomly created) that gradually evolves by selecting the fittest individuals (candidate solutions) based on their performance (objective function). Using the genetic variation operators (e.g. crossover, mutation) new and improved individuals are obtained, replacing the existing solution in the population.

The GP underlying search strategy is derived from GA, but with differences in regard to the obtained solution and chromosome encoding. GA are mainly used in optimisation problems where the populations evolves towards the optimum value for a given set of model parameters and the chromosomes are usually encoded as binary strings. On the other hand, in GP the structure of the model (e.g. algebraic expression, hierarchical program) evolves simultaneously with its parameters, while the chromosomes are usually encoded as tree-structures.

Representation

The most typical representation of GP chromosomes comes in the form of syntax tree, created as a composition of functions from a function set and a terminal set. The terminal set usually consists of independent variables, applied as inputs to the problem, and constants that are commonly identified during algorithm execution. The function set consists of domain specific functions and depending on the problem at hand may include arithmetic operators (+, -, /), mathematical functions (log, sin, cos), Boolean operator (AND, OR, NOT), logical expressions (IF-THEN-ELSE), iterative functions (DO-UNTIL), or any other user-defined function [5]. The parse trees can be defined by depth and size. The tree size is given by the maximum number of nodes, and the depth is defined as the longest path from the root mode to an endpoint [5].

In order to ensure mathematical validity of the symbolic expression represented by parse trees, the selected function and terminal set should satisfy the condition of “closure” and sufficiency” [5]. The closure property gives the ability to replace a subtree to another location in the same tree and also ensures a valid definition of functions for all possible combination of arguments. The case of division by zero is the most representative situation when closure property is not meet. This situation can be avoided by implementing various procedures (e.g. tree elimination) [5].

The sufficiency property implies that the function and terminal sets should be defined in such manner that all possible composition determined by functions and terminals include at least one solution to the problem at hand [10]. When this property is not met, the algorithm can only develop approximations but not the exact solution [5]. Depending on the specific problems, there are cases where an approximation may be as useful as the solution itself.

Fitness evaluation and selection

Corresponding to natural selection principles, better individuals are chosen for building a mating pool, due to the fact that they contain useful components that can generate improved solutions. The selection process is driven by the fitness of population members as a measure of how good they perform in solving a certain problem. Using a fitness function (objective function) the accuracy of each individual is calculated as the difference between simulated and actual (target) values but, depending on the objective of the modelling problem different error metrics such as mean square error (MSE) or root mean-square error (RMSE) can be used [5]. The selection process works in a probabilistic manner, in which individuals with better fitness have a higher chance to pass important genetic features into next generations. It can be mentioned that the fitness function can also be used as a stopping criterion of GP algorithm, if a predetermined threshold value is defined.

For the selection process, various selection strategies (e.g. fitness proportionate selection) developed within CE domain [11] are also used in GP, the most popular of them being the tournament selection method. In general lines, the tournament selection method works by randomly picking out a number of individuals (usually two) from the existing population. The fitness of each individual is compared, selecting as parent the one with best fitness. Due to the fact that only one individual is selected per tournament, the algorithm is applied multiple times until the desired number of parents for the mating pool is reached. Mathematical description of tournament selection method can be found in [12].

Genetic operators

For obtaining new population of symbolic expressions, the best individuals from the mating pool are transformed using genetic operators. The main genetic operators employed by GP algorithm are crossover and mutation. The crossover operator creates new offspring containing genetic material from two parents by swapping a randomly chosen subtree from one parent with another randomly chosen subtree from the other parent (fig. 1). This variation operator ensures the diversity of the population by promoting the best inherited features of different parents and is the primary tool for exploring a problem domain. The crossover operator is applied in a probabilistic manner using a predefined probability of crossover pc.

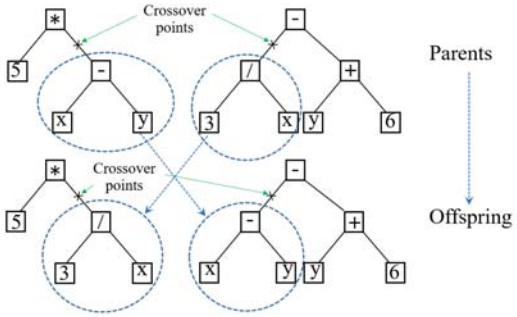


Fig. 1 - Example of crossover in GP (after [10])

The mutation operator performs by selecting one parent and modifying its structure by substituting a randomly chosen subtree with a randomly generated new subtree (fig. 2). The main advantage of this operator is that it ensures the diversity in the population due to the insertion of any functional subtree into a parent, that did not exist in the current population, whereas crossover can only insert already existing subtrees in the current generation. Similar to crossover the mutation operator is used based on a probability of mutation p_m.

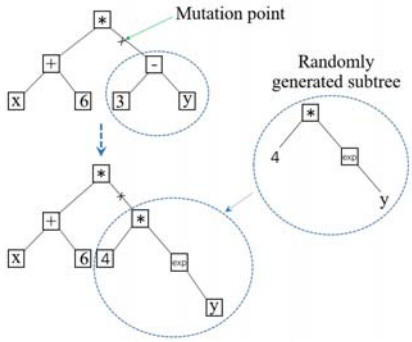


Fig. 2 - Example of mutation in GP (after [10])

The evolution process is applied over multiple generations, in a sequential manner, until a predefined termination criterion is satisfied (e.g. number of generations). Typically found in the generations developed closed to the algorithm termination, the individual that delivers the most accurate relationship for the modelled system is selected as the result generated by the algorithm.

Symbolic regression

Linear or nonlinear regression techniques are frequently employed for forecast purposes by using various relationships between independent and dependent variables. When regression techniques are applied, the structure of the model is selected in advance and can have different forms such as linear, polynomial or logarithmic expressions. The main objective in this approach is to estimate the model coefficients via an optimization technique, based on the available data. ANNs models can be grouped in the same class with nonlinear regression techniques. Similar to regression, the ANNs architecture must be selected in advances and only after a training process they can be used as models. There are two main drawbacks associated with these techniques. The

first disadvantage is given by the difficulties in finding the optimal model structure (expression or architecture) that best reflects the analysed process. The second issue is that they deliver no insight regarding the underlying mechanisms of the analysed process.

Having the ability to generate functional mathematical structures, GP is also regarded as type of regression technique, called symbolic regression [5]. GP has an advantage over conventional regression or ANN models because the specific model structure and its parameters are not selected in advance but are found during the search process. By generating functional expressions GP solutions may retain some physical meaning regarding analysed process. A drawback of this technique is given by the fact that it can generate bloated expression for which is virtually impossible to find any physical interpretation.

2.2 Study area

The study was undertaken for Bahluet catchment, situated in Iasi county, in the north-eastern part of Romania in Bahlui watershed. Located at the intersection of Moldavian Plain and Suceava Plateau, Bahluet catchment area is around 551km^2, with an average elevation around 250 m.a.s.l, a catchment slope of 7.4‰, while the length of Bahluet river is around 41 km. The main focus of the study was the upper part of Bahluet basin (fig. 3), with the closing section at Targu Frumos gauging station. The analysed area is roughly 66 km^2 and the river has a length of 21km. For Tg. Frumos gauging station the average multiannual precipitation has a value of 550 mm/year and the average multiannual flow is 0.15m^3/s.

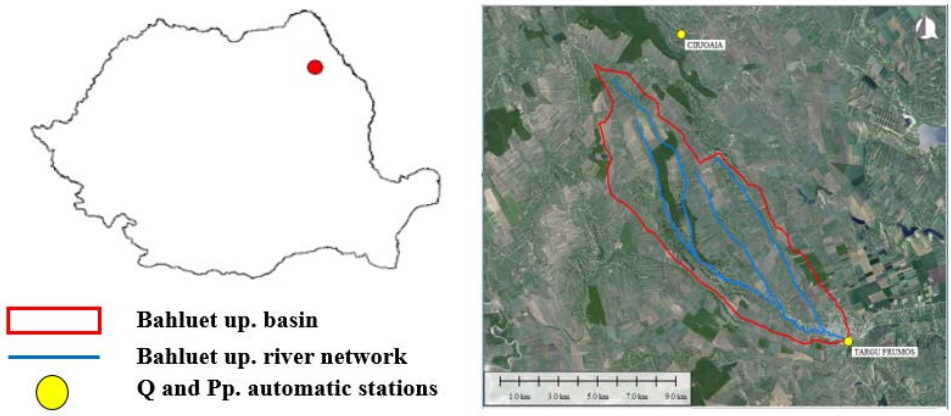


Fig. 3 - Bahluet basin map [9]

The hydrologic regime of Bahluet river is characterised by low flows all year round, with increases in discharges in early spring due to snow melting and in last period of autumn due to rainfall. Usually, in summer months the watershed is affected by scattered short strong rain events which represent the dominant process in flood generation [9]. Based on historical measurements, the daily maximum values for precipitation are registered for storm events that occur during July and August; consequently, this is the period with biggest measured flood waves. Due to the fast response and short lag times, this sub catchment can be catalogued as a flashy [9].

Due to the fast response of the catchment and the lack of any other station upstream of Targu Frumos, the possibilities for early flood warning are limited, which makes this area to be of high interest [9] from hydrological point of view.

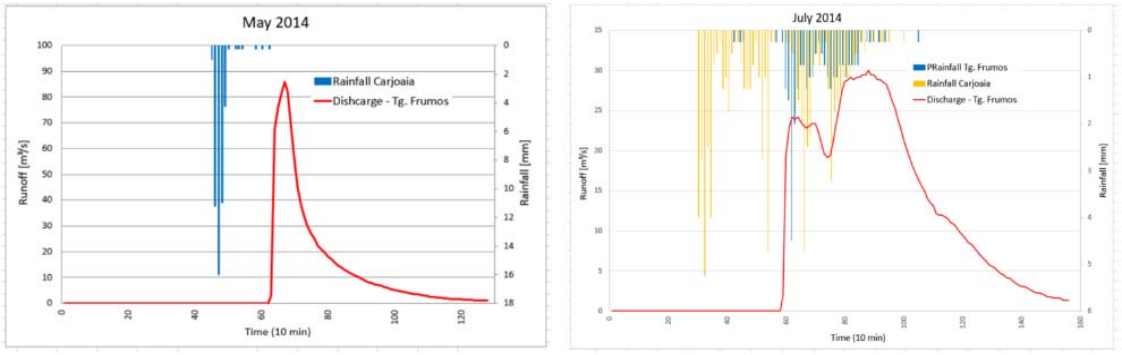
2.3. Available data

In line with study objective there were collected flow data, rainfall observation from gauging stations and radar rainfall estimates. For an event-based approach the collected datasets covered

two floods measured at Targu Frumos station. The first event occurred in 21.05.2014 and the second event was measured in 21-22.07.2014. All collected data were processed with a 10-minutes time resolution. The limitation on the number of events used is due to a relatively recent implementation of radar rainfall technology in the studied area. As a result, there is a limited number of rainfall events observed by both radar and ground automatic station.

2.3.1 Rainfall and runoff observation

In Bahlui catchment there are 14 automatic stations, equipped with flow and rainfall measuring instruments, covering an area of approximately 2000 km^2. Due to location of the studied catchment close to the water divide, only 2 rainfall stations are representative for the area. Station Carjoaia is located in the neighbouring catchment, very close to the northern extremity of the area of interest (less than 5 km). Targu Frumos station is located at the closing section of the watershed and it takes observations for both flows and rainfall. The rainfall measuring instruments use the tipping bucket principle with a 0.1mm resolution and with 10-minute time resolution, the same time resolution as for runoff measurements. The automatic monitoring network was implemented in DESWAT project and data are officially controlled and verified by ABA Prut-Barlad (regional water basin administration) [9].



a) b)

Fig. 4 - Registered flood-waves at Tg. Frumos gauging station [9]

The event of interest in this analysis is the one measured in 21.05.2014 (fig. 4-a) which presents the particularities of a flash-flood. The peak discharge had a value of 85,9 m^3/s and the peak flood stage measured at Tg. Frumos station was 336 cm, exceeding the maximum flood threshold (peril level) defined for this section (330 cm) [9]. The flood-wave was caused by a short but strong rainfall that occurred in the most upstream part of the catchment. Because the rainfall was concentrated in the upper part of the basin, there were no rainfall observations at Tg. Frumos gauging station. Due to its location, the rainfall observations for this event were available only for Carjoaia gauging station. Based on rainfall measurements the lag time for this event was around 3.5 hours [9].

The event observed in 23.07.2014 (fig. 4-b) had lower runoff values than the previous one but, according to the measurements it exceeded the flood warning level determined for Tg. Frumos gauging station (200 cm). The peak discharge had a value of 30 m^3/s and the maximum observed water level was 202cm. The event was generated by a rainfall distributed over the whole basin, with rainfall measurements at both Tg. Frumos and Carjoaia gauging stations.

2.3.2 Radar rainfall data

For the analysed area, the radar rainfall estimates are generated by a weather radar station located in Barnova, Iasi County. The station uses a Doppler high-resolution S-band weather radar, taking observations at a 6-minute interval and the spatial resolution of the volume element is 1 km^3 [9]. The weather radar station has a 230 km range and is situated about 46 - 65 km from the studied basin.

The rainfall rate is estimated using an empirical relationship that takes as input the radar reflectivity factor *Z*.

$$Z = aR^b \tag{1}$$

where *R* is the rain rate in mm/h, *Z* is the reflectivity factor in mm^6/m^3, and the coefficients *a, b* are empirical parameters. For Barnova weather radar station the value of the parameters varies, depending on the season; the summer the values of the parameters are $a = 300$, $b = 1.4$ while for winter period $a = 200$ and $b = 1.6$. [9]

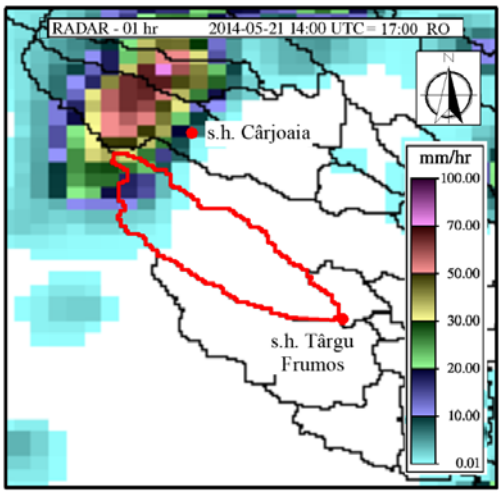
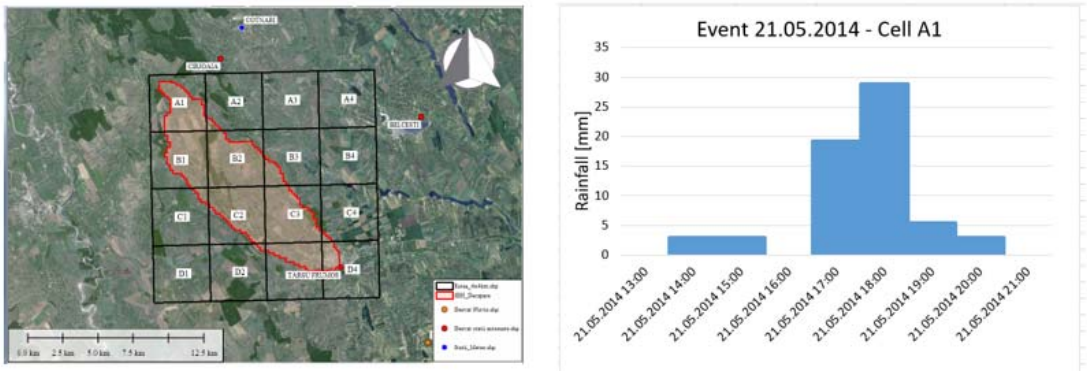


Fig. 5 - Example of radar rainfall image (May 2014 event) [9]

The radar data for each of the events previously mentioned were collected from ABA Prut-Barlad being generated by SIMIN system (National Integrated Meteorological System). The precipitation estimates are delivered as accumulated values over one hour. In fig. 5 an example of a processed radar image taken in 21.05.2014 at 17:00 hour is presented [9].

2.4 Data processing for PG algorithm input

In order to obtain the rainfall information necessary for the study, the radar files were pre-processed using GIS instruments. Radar rainfall data use a grid with 1km spatial resolution for providing the accumulated rainfall values. Applying this grid for extracting the rainfall data for the studied area leads to a very large input vector. In order to reduce the input vector, the solution of averaging the rainfall values over a coarser grid was adopted. Using a grid with 4km spatial resolution for each of the two events 12 rainfall time series with the same time resolution as the radar files (1 hour) were generated. The rainfall time series were extracted only for the cells intersecting the catchment area. In fig. 6 the extraction grid overlaid on the catchment surface together with an example of a hyetograph extracted for cell A1 (upper left cell of the grid) are presented.



a) b)

Fig. 6 - Extraction grid for radar rainfall data and hyetograph resulted for A1 grid cell [9]

As previously mentioned, the rainfall volumes determined from radar data were also used as inputs to GP algorithm. The volumes were computed using the rainfall obtained from radar files and the surface of 4km grid cells that intersected the studied catchment. Similar to previous case, were obtained 12-time series with a time resolution of 1-hour.

Due to differences of time resolution of radar and ground rainfall measurements, the data sets were homogenised. The radar data were processed in order to obtain time series with a 10-minute resolution, matching the automatic sensors observations. The time distribution for radar data was considered uniform over a one-hour interval.

The ground and radar data were presented to GP algorithm in matrix form. The input vector determined by ground rainfall measurements was structured as matrix of 2 columns, representing Carjoaia and Tg. Frumos station and by *n* rows, representing the number of time steps for each of the events. The input vector obtained from radar data was structured in the same manner, using a matrix of 12 columns (representing the cells intersecting the basin) by the same time step number as ground measurements. For the grid cell or the stations with no rainfall zero values were implemented in the input vectors [9].

3. Results

As previously mentioned, the aim of the study was to assess the model induction capacities of PG algorithm to rainfall-runoff modelling for a Bahluet catchment. Three scenarios were analysed, generated by the use of tree different input vectors. For conducting the study, the open-source software package *HeuristicLab* was used, that has implemented various evolutionary and heuristic techniques, including a genetic programming algorithm developed for time series predictions.

The symbolic regression for time-series prediction implemented in HeuristicLab allows the use of autoregressive target and the mathematical relationship can be generally expressed as:

$$Q(t+1) = f\left(R(t), R(t-1), ..., R(t-n_r), R(t), Q(t), Q(t-1), ..., Q\left(t-n_q\right)\right) + \varepsilon \quad (2)$$

where $R(t), R(t-1), \ldots, R(t-n_r)$ represent current and previous observations of the input vector (rainfall and volumes), $Q(t), Q(t-1), \ldots, Q(t-n_q)$ are the current and previous measurements for the target variable (discharges), n_r and n_q are the maximum number of past values for input variables and targets and ε is the error term. In essence the expression (2) represents a nonlinear autoregressive model with exogenous inputs.

For increasing the generalisation capacity, the datasets were split into training subset and testing subset with a 0.7 weight attributed for the training subset and the remaining 0.3 for the test subset.

The parameters used by the PG algorithm in search of the model structure that best fit the input data are shown in table below:

Table 1

PG parameters used for training in all three cases

PG parameters	Used values
Maximum symbolic tree depth	100
Maximum symbolic tree length	500
Function set	+, -, *, /, EXP, LOG, SQRT, POW, lagged variables, Autoregressive variable
Maximum generations	200
Cross over probability	0.85
Mutation probability	0.15
Population size	100
Number of past measured values	10

3.1 GP training

For the training phase of GP algorithm was used the event observed in July 2014, because of the better spatial representation over the analised catchment. The performances of the induced expression generated via symbolic regression were evaluated using graphical method (observed vs. simulated hydrographs) together with goodness-of-fit coefficients. The goodness-of-fit measures used to evaluate the model's prediction are root mean square error (RMSE) and coefficient of efficiency (CE). The former performance index gives quantitative information regarding the error of the models and the latter assess their predictive capacity.

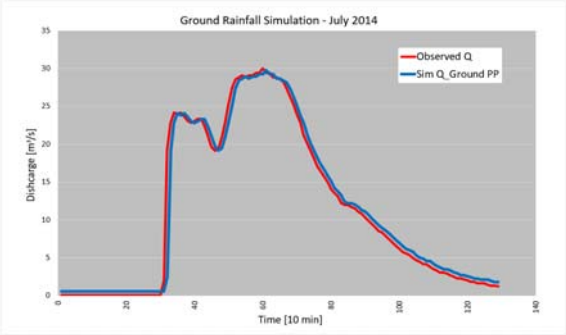


Fig. 7 - Simulated vs measured runoff hydrograph (July 2014) – ground rainfall observations (model A8L10)

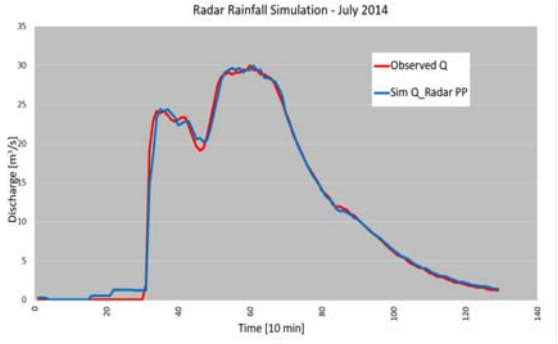


Fig. 8 - Simulated vs measured runoff hydrograph (July 2014) – – radar rainfall observations (model A41L286)

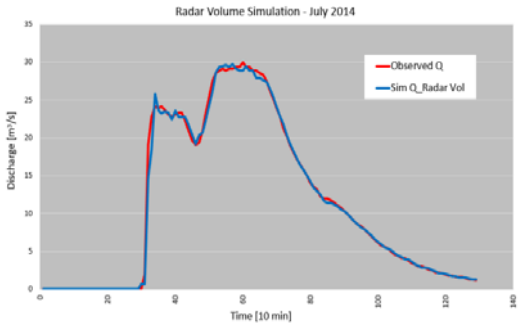


Fig. 9 - Simulated vs measured runoff hydrograph (July 2014) – rainfall volume (model A47L400)

Table 2

Goodness-of-fit coefficients for model training

	RMSE	CE
Ground PP - Model A8L10	1,682	0,975
RadarVol - Model A41L286	0,781	0,995
Radar PP - Model A47L400	0,672	0,996

All simulated flood waves (blue line), presented in the figures 7, 8 and 9, show an almost perfect agreement with the observed hydrographs (red line) with almost no perturbation affecting the results. In all training cases, the peak discharges are well captured in terms of value and time of occurrence, with the exception of the first hydrograph, that presents a delay of one-time step. The goodness-of-fit coefficients (table 2) presents more than satisfactory values, for all presented cases, suggesting good performances for all expressions identified with GP algorithm.

The expressions induced by GP algorithm are labelled based on tree depth and length (e.g. the model A8L10 has a tree depth of 8 and a length of 10). For the first training case (ground rainfall measurements) a reasonable symbolic expression was obtained, regarding the number of used elements. For the other two cases the expressions are very large and are induced mainly by the dimension of the input vector, determined from radar data. The expression obtained for the first training (fig. 7) case is presented below:

$$Q(t) = \left(log\left(exp\left(c_0 Q(t-1)\right)\right)\right)c_1 + c_2 \qquad (2)$$

where c_0=0.956, c_1=1.02, c_2=0.51. According to eq. (3), the algorithm generated the expression based only on discharge and leaving out any rainfall information. In general, the GP capacity to choose

only the most influential input variables and disregard other input data is seen as a beneficial property of the algorithm. However, by disregarding the rainfall and generating an autoregressive model based only on discharges, no real information can be extracted with respect to the influence of precipitation as the main mechanism of flood generation. Consequently, this model was not considered fit to simulate the rainfall-runoff process and was abandoned from further testing.

Regarding the other two cases, due to the large dimension of the symbolic expression (fig. 10-11), it is impossible to find any physical interpretation of the underlying process and the expression has the properties of a black-box model.

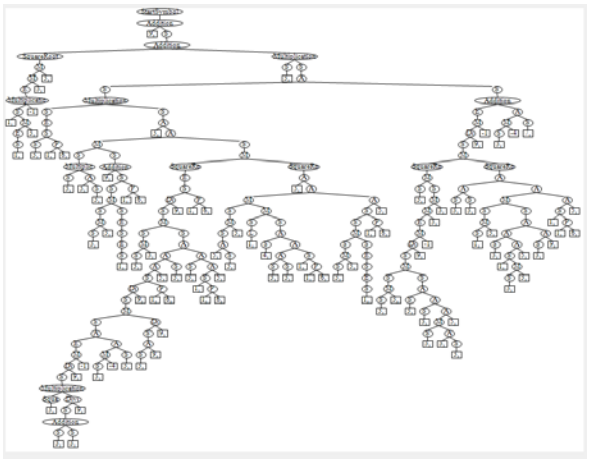


Fig. 10 - Symbolic tree generated for radar rainfall estimates (model A41L286)

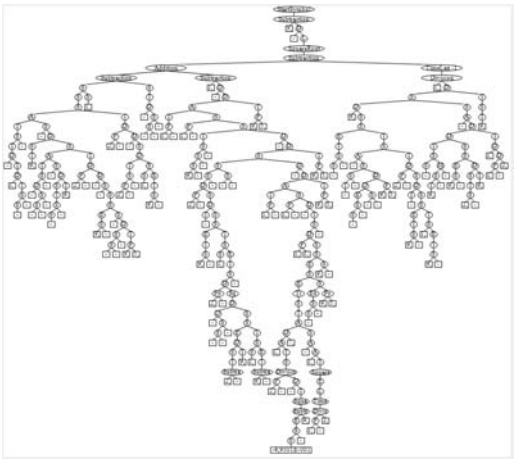


Fig. 11 - Symbolic tree generated for radar precipitated volumes (model A47L400)

3.2 GP Testing

The validation of symbolic expression was done using the observed data for May 2014 event. The computed hydrographs are presented for the remaining two cases in the pictures below and the performance coefficients are given in table 3.

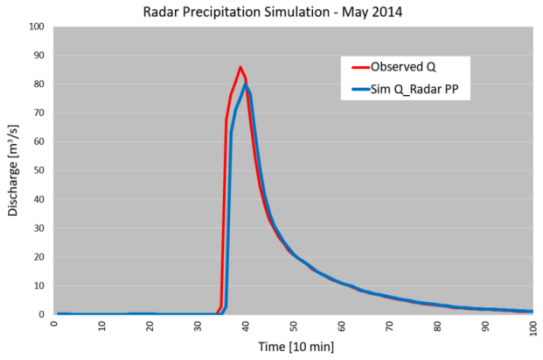


Fig. 12 - Simulated vs measured runoff hydrograph (May 2014) – radar precipitation observations (model A41L286)

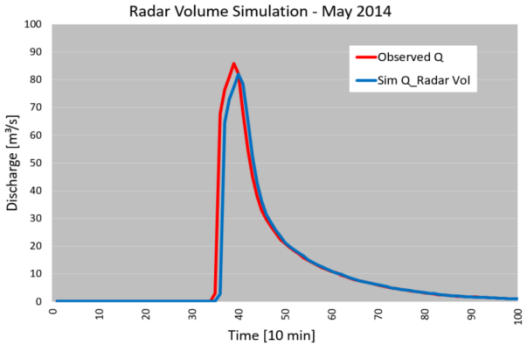


Fig. 13 - Simulated vs measured runoff hydrograph (May 2014) – rainfall volume (model A47L400)

Table 3

Goodness-of-fit coefficients for model validation

	RMSE	CE
Radar PP - Model A41L286	6,785	0,873
Radar Vol - Model A47L400	6,938	0,874

The simulation of May 2014 event was the main objective of the study, due to its high discharge values and short lag time. Both simulated hydrographs (fig. 12-13) present a good agreement between observation and simulation. The peak discharge for simulated data was not reached, but the differences are acceptable, with a -6,75% difference for A41L286 model and a -4,55% for A47L400 model. In this situation, the expression obtained with precipitated volumes as input to PG generates improved results. Beside the discharge error, for these two cases a phase error of one-time step is also present, that did not occur in the training stage. However due to time resolution used for the simulations (10 minutes the time step) the phase error in both cases is acceptable.

The obtained results considering spatially distributed inputs are of great interest for hydrological forecasting purposes. Being able to use spatially distributed inputs obtained from radar products, GP algorithm can increase the lead time for flood predictions.

4. Conclusions

The aim of this study was focused on testing the GP capabilities in simulation a flash-flood using an event-based approach. Three different input data were employed in conducting the study, namely, ground rainfall measurements, radar rainfall data and rainfall volumes. In all three cases, GP was able to develop functional mathematical expression that led to acceptable results for all input data. A significant observation can be done regarding the mode of operation of the algorithm. The ability to select only some information and disregard other data from the input vector can be seen as beneficial, leading to optimal expression. However, for the case of ground rainfall observation, the algorithm placed a much greater emphasis on the target variable that led to a total disregard of rainfall data, generating an autoregressive model. Consequently, although the model developed for ground rainfall observations presented good simulated discharge values, it cannot be used for flood prediction since it is not using rainfall data and was abandoned for testing.

For the cases where spatial distributed data were used, the expression induced with symbolic regression generated acceptable results. In contrast with the ground rainfall observations, there was less emphasis put on regressed variable and more towards the rainfall data. It can be noted that the composite variable, namely rainfall volumes, led to improved results compared to the results obtained only from radar rainfall estimates, in regard to the maximum simulated discharge. Also, it can be mentioned that for spatial distributed cases the GP algorithm generated bloated expression with a large number of components with no real possibility of extracting any physical interpretation of the underlying process. It can be considered that the mathematical structures generated in this situations function as a black-box model. Regarding the use of GP algorithm for hydrological forecasts applications, the results generated based on radar products inputs are of great significance, having the possibility of improving the lead time for flood prediction in small catchments.

Acknowledgments

This article was designed as a continuation and completion of the article [9]. The same input data sets for the same watershed were used to model the rainfall-runoff process based on genetic programming algorithm.

References

[1] Solomatine, D. P. & Ostfeld, A. (2008), *Data-driven modelling: Some past experiences and new approaches*. Journal of Hydroinformatics. 10(1), 3–22. DOI: 10.2166/hydro.2008.015;

[2] Hsu, K., Vijai Gupta, H., & Sorooshian, S. (1995), *Artificial neural network modeling of the rainfall-runoff process*. Water Resources Research. 31(10), 2517–2530. DOI: 10.1029/95WR01955;

[3] Abrahart, R. J. & See, L. M. (2007), *Neural network modelling of non-linear hydrological relationships*. Hydrology and Earth System Sciences. 11, 1563–1579. DOI: 10.5194/hess-11-1563-2007;

[4] Minns, A. W. & Hall, M. J. (1996). Artificial neural networks as rainfall–runoff models. Hydrological Sciences Journal, 41(3), 399–417;
[5] Koza, John R. (1992). *Genetic Programming: On the Programming of Computers by Natural Selection*. Cambridge, MA, USA: MIT Press;
[6] Savic, D. A., Walters, G. A. and Davidson G. W. (1999). *A genetic programming approach to rainfall-runoff modeling*. Water Resources Management 13: 219-231;
[7] Babovic, V., Keijzer, M. (2002). *Rainfall runoff modelling based on genetic programming*. Nordic hydrology, vol. 33, pp. 331-346;
[8] Whigham, P. A., Crapper, P. F. (2001). *Modelling Rainfall-Runoff Relationships using Genetic Programming*. Mathematical and Computer Modelling: An International Journal, Volume 33, pp 707-721. DOI: 10.1016/S0895-7177(00)00274-0;
[9] Dinu, C., Drobot, R., Pricop, C., Blidaru, T. V. (2017). *Flash-flood modelling with artificial neural networks using radar rainfall estimates*. Scientific Journal - Mathematical Modeling in Civil Engineering, Vol. 13-No. 3: 10-20 - 2017, Doi:10.1515/mmce-2017-0008;
[10] Poli R., Langdon W.B., McPhee N.F. (2008). *A Field Guide to Genetic Programming.* Lulu Enterprises UK Limited;
[11] Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimisation and Machine Learning*. Reading, Mass., U.S.A: Addison Wesley Addison-Wesley Publishing Company, Inc.;
[12] Goldberg, D.E., Deb, K. (1991). *A comparative analysis of selection schemes used in genetic algorithms.* Foundations of Genetic Algorithms, Volume 1, pp 69-93. DOI: 10.1016/B978-0-08-050684-5.50008-2.