



## Genetic programming for turbidity prediction: hourly and monthly scenarios

### Suda bulanıklık tahmini için genetik programlama: saatlik ve aylık senaryolar

Bahrudin HRNJICA<sup>1\*</sup>, Ali Danandeh MEHR<sup>2</sup>, Behrem SEFIK<sup>3</sup>, Necati AĞIRALIOĞLU<sup>4</sup>

<sup>1,3</sup>Technical Faculty, University of Bihaç, Bihaç, Bosnia and Herzegovina.

bahrudin.hrnjica@unbi.ba, behrem.sefik@unbi.ba

<sup>2,4</sup>Civil Engineering Department, Engineering Faculty, Antalya Bilim University, Antalya, Turkey.

ali.danandeh@antalya.edu.tr, necati.agiralioğlu@antalya.edu.tr

Geliş Tarihi/Received: 07.03.2019, Kabul Tarihi/Accepted: 09.07.2019

\* Yazışılan yazar/Corresponding author

doi: 10.5505/pajes.2019.59458

Özel Sayı Makalesi/Special Issue Article

#### Abstract

This paper presents the calibration and evaluation of two genetic programming (GP) methods, namely classic GP and gene expression programming (GEP) for turbidity prediction at drinking water distribution networks. Classic GP first method was used to model turbidity at the main water source of Bihaç town (Bosnia and Herzegovina) and GEP second method was used to model turbidity at one of the water monitoring stations of city of Antalya, Turkey. The former various predictive models were built based on the mean monthly turbidity measurements recorded during 2006-2018. Moreover, hourly measurements at Gürkavak Station during low turbidity period were used. The results showed that the modelling of turbidity is a challenging task which required careful data analysis especially in the context of determining the optimum lag times/input parameters. No meaningful relation between discharge and turbidity was found at Antalya water supply pipeline. The results also indicated that the predictive models based on the presented algorithms may provide more accurate estimations in comparison to the traditional regression approach. The findings are useful for sustainable urban water management whereby a high quality water supply is aimed.

**Keywords:** Genetic programming, Gene expression programming, Turbidity, Drinking water

#### Öz

Bu çalışmada, içme suyu dağıtım şebekelerinde bulanıklık tahmini için klasik genetik programlama (GP) ve gen ekspresyon programlama (GEP) olmak üzere iki GP yönteminin kalibrasyonu ve değerlendirilmesi sunulmaktadır. İlk yöntem olan Klasik GP, Bihaç kentinin (Bosna Hersek) ana su kaynağındaki bulanıklığı modellemek için kullanılmıştır. İkinci yöntem olan GEP, Türkiye'de bulunan Antalya ili su izleme istasyonlarından birinde bulanıklık modellemesi için kullanılmıştır. Birincisinde, 2006-2018 döneminde kaydedilen ortalama aylık bulanıklık ölçümlerine dayanarak çeşitli tahmin modelleri oluşturuldu. İkincisinde ise, düşük bulanıklık dönemindeki Antalya-Gürkavak İstasyonu'ndaki saatlik ölçümler kullanılmıştır. Sonuçlar, bulanıklık modellemesinin, özellikle optimum gecikme süreleri ve girdi parametrelerinin belirlenmesi bağlamında, dikkatli veri analizi gerektiren zorlu bir görev olduğunu göstermiştir. Antalya su temin hattında debi ve bulanıklık arasında anlamlı bir ilişki bulunamamıştır. Bulunan sonuçlar ayrıca sunulan algoritmalara dayanan tahmin modellerinin geleneksel regresyon yaklaşımına kıyasla daha doğru tahminler sağlayabileceğini göstermiştir. Bulgular, yüksek kalitede su temininin hedeflendiği sürdürülebilir kentsel su yönetimi için kullanışlıdır.

**Anahtar kelimeler:** Genetik programlama, Gen ekspresyon programlama, Bulanıklık, İçme suyu

## 1 Introduction

Turbidity is one of the main quality parameter of drinking water. It can be defined as an optical determination of water clarity or cloudiness of a fluid which is caused by large number of invisible (to the bare eye) particles. Turbidity in water can be considered like smoke in air. It is defined as the measure of relative clarity of water and its optical characteristics. It is an expression of the amount of light scattering by materials in the water. The higher intensity of scattered light means the higher turbidity and consequently the lower quality. High level of turbidity in drinking water can cause problems for immunocompromised people due to suspended solids. Suspend solids can contain various types of viruses or bacteria and develop diseases while consuming such kind of water. In order to protect the people governments have set standards for allowable turbidity in drinking water. In Europe, each country developed its own standard which is based on the European standards for turbidity ISO 7027-1:2016. Usually, the turbidity is measured in nephelometric turbidity units (NTU). Different

standards define different allowable levels of NTU. In USA, allowable value of turbidity must be less than 5 NTU while as per European standards the maximum values must not be more than 4 NTU.

The main cause of sudden appearance of turbidity in water flows is typically appearance of short-term rainfall causing soil erosion. However, researches have shown that it was a bit more complex process (Nebbache et al., 2001) and mainly depends on the soil structure, level of underground water and variety of natural and anthropogenic factors and processes. Due to the fact the turbidity is very important parameter and directly influences the quality of drinking water, engineers use different approaches to predict it (Najah et al., 2009), (Makic et al. 2015a) and (Makić et al., 2015b). In this paper, turbidity models were developed based on different genetic programming methods and comparison analysis was performed. The GP method has been proved by many engineers as an effective and suitable method in modelling and developing predictive models of hydrological parameters (Danandeh et al., 2017) and (Olyaie et al., 2016). Due to its importance, the turbidity is of constant

interest for engineers using different approaches in order to develop a good and reliable model. For example, (Mulia et al., 2013) used artificial neural networks (ANN) in combination with genetic algorithm (GA) developing data-driven models in order to predict turbidity and chlorophyll-a concentrations. (Ogston et al., 2010) predicted the turbidity based on enhanced sediment resuspension which is the result of sea level rise. Recently, (Nakao et al., 2017) proposed a new method of inverse analysis of turbidity currents based on turbidity in order to reconstruct the hydraulic conditions of ancient turbidity. Using different machine learning methods, in modelling water parameters, it is proved to be very effective way of monitoring and handling environmental resources. (Wang et al., 2013) analysed large complex water data set in order to monitor 12 parameters at 13 different sites, by using different methods and techniques. In this large research study several methods like cluster analysis, principal component analysis, factor and discriminant ones were used in the water parameters evaluation. (Iglesias et al., 2014) studied the importance of the turbidity where they developed turbidity models with influence of several water quality parameters by using ANN. More recent studies about turbidity and prediction models by using different machine learning techniques can be seen in the literature (Huang et al., 2018), (Kazemi et al., 2018), (Koven et al., 2019), (Savary et al., 2018) and (Wang JD et al., 2010). Undoubtedly, turbidity is one of the most important quality parameter in managing drinking water and it is the subject of interest for many engineers. This paper developed new models in order to decrease costs and increase quality of managing this important resource for two case study locations.

## 2 Methods

The use of artificial intelligence (AI) techniques to solve a wide range of engineering problems was increased by the quick growth of information technology at the outset of the 1990s. As hardware components start to be more powerful every year AI methods and machine learning solutions have got more attention in recent years. Genetic programming (GP) widely used in water engineering (Danandeh et al., 2017) is of the recent AI methods that's frequently used to create symbolic regression models between a set of predictors and a given predictand. Comparing to earlier AI methods such as artificial neural networks (ANN), GP provides explicit model easy to be applied in practice. Thus this method was considered in the present study.

### 2.1 Genetic programming

GP evolved as a generalisation of 40 years old genetic algorithm (GA). Unlike GA where chromosomes were represented as binary numbers which was found as constraining factor, in GP chromosomes were created as tree structures representing the computer programs that conduct mutation and crossover operations and thus produces a computer program which solves a particular problem. Chromosomes in GP are represented in population shaped like a hierarchical structure which contains primitive functions and terminals for particular problem area. Set of primitive functions are in fact arithmetic operations, mathematical functions, Boolean operators and special functions for particular problems. Terminal set is also a part of chromosome structure and it is usually formed from input parameters and numerical constants. Population growth in GP is consistence with the biological principle of "survival of the fittest", through evolutionary crossover and mutation operations adopted to mating of computer programs.

GP begins with formation of initial population of randomly generated computer programs. Each computer program (chromosome) in the population possesses the fitness function by which its ability for solving the problem can be evaluated and measured. Once the initial population is evaluated the population is checked if satisfied with given termination criteria. If the termination criteria is satisfied the program terminates by returning the best solution found. Otherwise the GP goes into the iteration loop where it repeats several steps. Within the loop, each population is breeding by performing crossover, mutation and reproductions. Once the breeding phase is over the population is reevaluated and checked if satisfied with the termination criteria. When the termination criterion is met the program ends the evolutionary progresses and the best program is returned as the output. Typical flowchart of GP was shown on Figure 1 (Hrnjica et al., 2019).

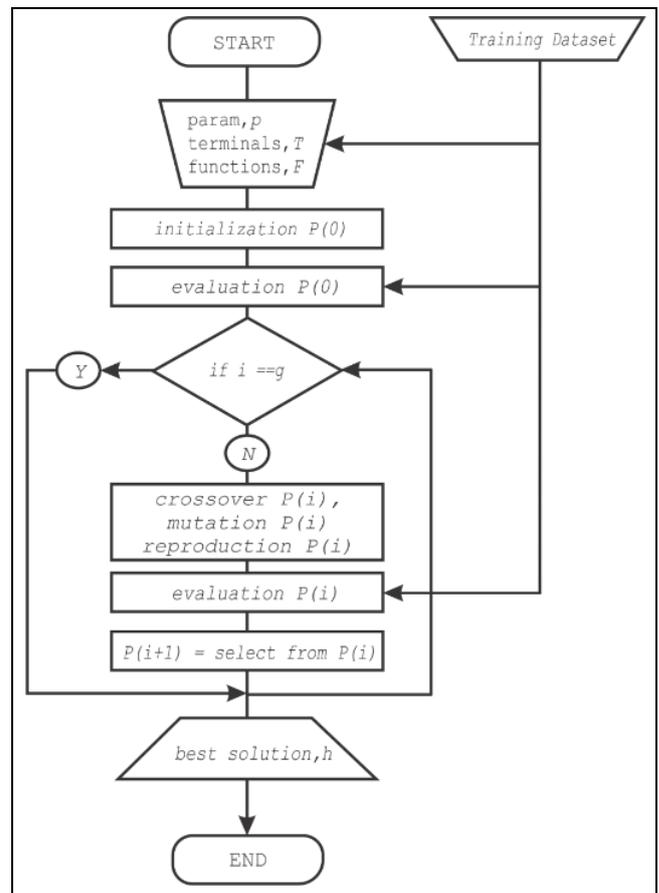


Figure: 1 GP flowchart.

The whole GP process is controlled by several GP parameters. According to GP algorithm the evolutionary process is commonly controlled through 19 parameters. Two of them are main parameters. There are also 11 secondary and 6 qualitative parameters that must be carefully chosen with alternative methods of executing the algorithm. The main GP parameters are M- population size and G- maximum number of generations. Secondary GP parameters represent a probability value that a certain event will occur or specific value level of the GP tree structure. Here is the list of the most important secondary parameters. The probability of the crossover (the recommended value should be greater than 90%), the probability of the reproduction (the recommended value should be about 20% of the total population), the probability of

the mutation (the recommended value should be about 5% of the total population), maximum initial level of the chromosome tree structure, maximum operation level of the chromosome tree structure etc. The secondary parameters are included in the algorithm depending on the method of the GP implementation and are considered as optional.

For more than 20 years, GP has been one of the most successful AI technique applied to solve a wide diversity of water engineering problems. The main reason for its great success over time is its capability to generate explicit expressions that are easily analysed and validated. One of the main issue that most of the AI techniques have to tackle is setting constraints on the complexity of the system to be developed. For instance, a polynomial regression model needs to set the polynomial order in advance or ANN needs to set the number of hidden neurons/layers within the network. Hence, those constraints limit the complexity of the problems that the polynomial or network would be able to solve. In contrast, GP has no beforehand constraints on problem complexity. Since the most common expression in GP has the shape of a tree and given the possibility that this tree may grow the corresponding mathematical expression can be of arbitrary complexity, although some measures may be taken to avoid excessively large trees like limiting the maximum height of the tree.

### 2.2 Gene Expression Programming

Gene expression programming (GEP) is an advanced AI technique from GP family that selects some of the evolved programs to link and improve them with respect to a given objective function and obtains the best combination of individuals using a linking operator. As firstly described by (Ferreira et al., 2001), there is primary difference between GEP and GP algorithms principally residing in the nature of their evolved programs. In classic GP the evolutionary algorithm output is the individual with the best fitness in the training set. However, in GEP each individual includes one or more genes called sub-expressions (sub-ET) and the best result is built by linking the sub-expressions using mathematic or Boolean functions such as AND, OR, NOT etc. For example, the GEP chromosome shown in Figure 2 d) represents a model (i.e., ET) as the linear combination of tree sub-expression functions (Hrnjica et al., 2019). Mathematically, the model was presented as follows:

$$(x_1 \times \sin x_2 + x_2 \times \cos x_1) + (x_1 \times x_2 + \cos x_2) + (c_1 \times x_2 + x_1 + x_2) \quad (1)$$

where  $x_1$  and  $x_2$  are the vectors of input variables. Each expression within the brackets represents a sub-ET and the addition function between the sub-ETs is the linking function. Therefore, equation (1) indicates a GEP program by which three nonlinear individuals were linearly combined. It can be implied that each sub-ET is a part of the best solution with less complexity than main one. Therefore, it can be concluded that GEP employs the power of small GP blocks (i.e., sub-ETs) to capture nonlinear behaviour of a complex system through a multigene structure. Bearing in mind that the best solution must be checked for the likely overfitting problem.

Despite several advantages of GEP over classic GP (see (Ferreira et al., 2001) for more details) one of its main drawbacks is the expressional complexity of the final solution that may also become overfitted to the training set. In such cases noise within the training set is treated as a feature and the generated expressions and also learned for the noise

information. Usually this effect is evidenced by very low errors on the training set, whereas the evaluation on the test set achieved high errors.

A common approach to address this issue is to use lower number of sub-ETs as well as limiting the maximum height of the associated trees. To overcome the overfitting problem, the validation dataset that allows to calculate the validation error on the expressions generated by both GP and GEP can be used. Nonetheless, this is not an easy decision because the error in the validation set may decline in future generations.

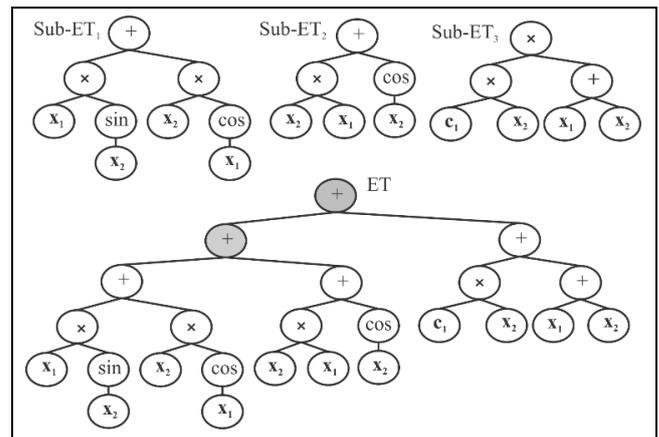


Figure 2: An example of GEP chromosome comprising 3 Sub-ETs summed by addition linking function.

### 3 Case studies

In the present study, the canonical GP and GEP were trained and tested by using observations at two different locations of Bihac town (Bosnia and Herzegovina) and Antalya (Turkey), respectively. In the first case study the GP was applied for turbidity forecasting of drinking water at the location of Klokot, the main water supply source of Bihac town in Bosnia and Herzegovina (Figure 3).

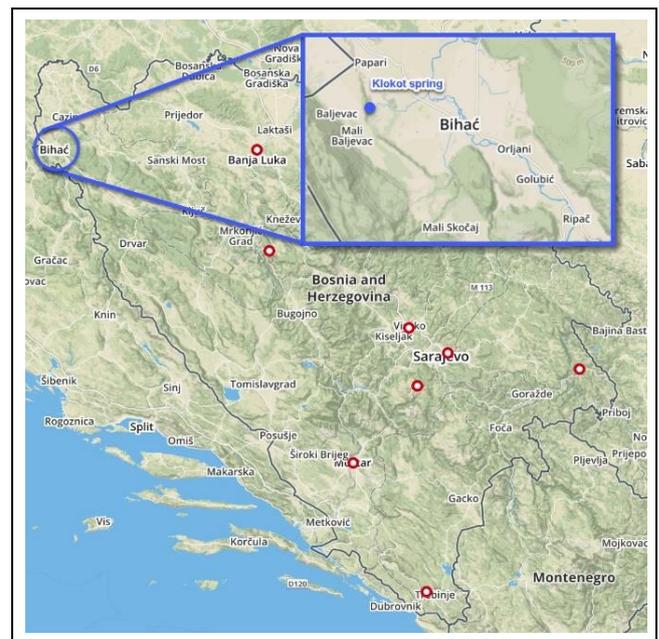


Figure 3: Klokot spring location at Bihac town.

The turbidity of drinking water was measured at the water source. Measurements were monthly data sets which were collected for the period from 2006-2018. All the collected data were split into training and testing data sets. The training set was used to build a model and the testing one was employed to validate the model considering both overfitting problem and accuracy of turbidity predictions for the next 6 months. GPdotNET computer program was used for modelling and determination of the best GP model. The software provides a user friendly framework capable of providing all three common steps of modelling: a data pre-processing, function generation (i.e., predictive modelling) and data post-processing. Technical features of the software (user guide manual) are available in (Hrnjica et al., 2019). Satisfactory application of GPdotNET in hydrological predictions was reported recently (e.g., Danandeh et al., 2018).

Figure 4 shows turbidity values measured from 2006 to 2018 with calculated three months moving average (hereafter MA), used as input data set for classic GP modelling.

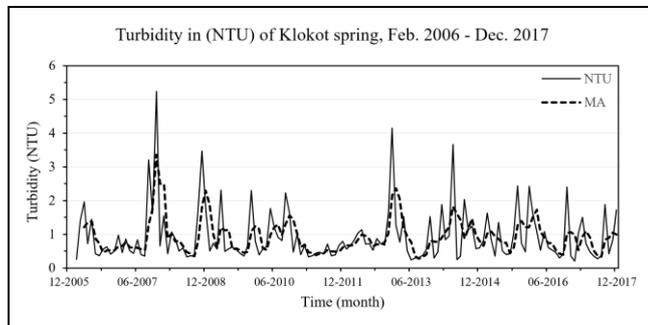


Figure 4: Turbidity of drinking water measured at Klokot spring, and MA (3): 3 months moving average.

In the second case study the GEP was applied to model the turbidity of drinking water at the 13<sup>th</sup> measurement point of Gürkavak water supply pipeline located in Antalya, Turkey (Figure 5). The station was monitored via central Supervisory Control and Data Acquisition (SCADA) control system by Antalya Water and Sewage Administration (ASAT). Beside turbidity some other water flow characteristics such as discharge and point pressure were measured and recorded continuously. The ultimate goal in this case study was to investigate if there was a noteworthy relation between low turbidity values ( $0.6 < \text{NTU} < 1.0$ ) and flow discharge or not. To this end hourly mean measurements at the station during 24.09.2017 to 12.11.2017 were used in this study (Figure 6). Figure 6 shows 1054 turbidity samples and their corresponding discharges as well (See Figure 6a).

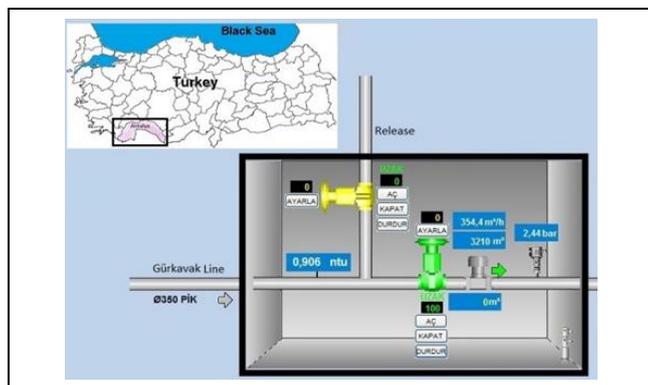


Figure 5: Location of the second case study, Gürkavak, Antalya.

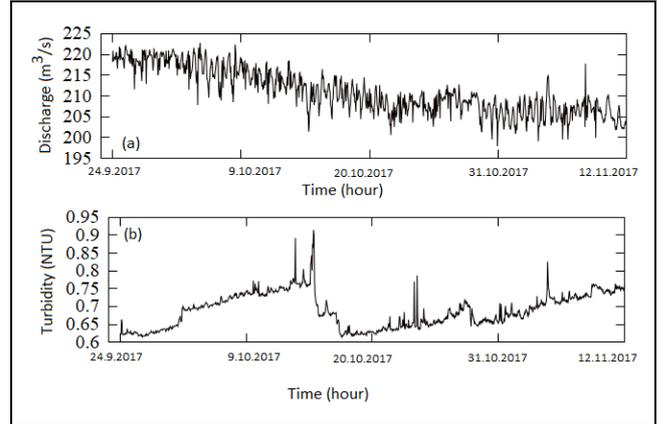


Figure 6: Hourly (a) discharge and (b) turbidity measurements at Gürkavak Station, Antalya, Turkey.

## 4 Results

In the first case study the GPdotNET computer program was used in order to determine GP model of three months moving average turbidity. The measured data were prepared in order to optimize searching algorithm by defining set of data transformation (Makic et al., 2015a). During modelling the following GP parameters were used:

- Function set  $F = \{+, -, *, /, 1/x, \tanh\}$ , terminal set  $T = \{t_{n-1}, t_{n-2}, t_{n-3}, t_{n-4}, \mathfrak{R}\}$ , where  $t_n$  - turbidity of the  $n$  month parameters,  $\mathfrak{R}$  - set of random constants.
- Population size  $M = 1000$ , number of generation  $G = 500$  and probability of the genetic operations  $p_c = 95\%$ ,  $p_m = 20\%$ ,  $p_r = 10\%$ ,  $d_{\text{initial}} = 5$ ,  $d_{\text{formed}} = 9$ .

Since the time series dataset was used in the first case study, it was necessary to define a time lag in order to create the modelling input datasets. For this case study different lag times were tested and the best result was achieved using time lag of 4 months. Once the time lag was determined the searching and improving process of GP started. Firstly, the training and testing datasets were defined by splitting experimental measurements into two datasets with the ratio of 75% - 25%. In other words, 75% of all data formed the training dataset and the remaining 25% (27 months) formed validation dataset. Once the datasets were setup the termination criterion was setup on 500 generations. At the beginning of the searching process several different configurations were applied in order to get best possible combination of all GP parameters. The time execution of the searching process in GP depends on the complexity of the process and the accuracy to be achieved. When an acceptable GP model is developed post-processing of the results may require to distill knowledge from the GP model. GP model of turbidity is given in analytical form in the following expression.

$$\begin{aligned}
 &= 0.8(t_{n-1} - 0.28) \left( \frac{0.65(t_{n-2} + 2.44)(t_{n-1} - 0.28)^3(1.27 t_{n-4} + t_{n-1} - 0.64)}{(t_{n-2} + 2.5)} \right. \\
 &\quad \left. + \frac{1/(7.77t_{n-4}t_{n-3} - 2.17t_{n-4} + 6.11t_{n-3}t_{n-1} - 3.89t_{n-3} + t_{n-1}^4 - 1.12t_{n-1}^3 + 0.47t_{n-1}^2 - 1.8 t_{n-1} + 1.1)}{t_{n-1} + t_{n-4} - 0.56} - 0.33(t_{n-3} - 0.28) \right. \\
 &\quad \left. - 0.33(t_{n-1} - 0.28) + 1.73) + 0.28 \right) \quad (2)
 \end{aligned}$$

where  $t_n$  is turbidity of the current month  $n$ , and  $t_{n-1}$ ,  $t_{n-2}$ ,  $t_{n-3}$ ,  $t_{n-4}$  turbidity of the previous 1, 2, 3 and 4 months.

Once the GP model is calculated the prediction for the next two years is evaluated and the result is shown on Figure 7. It can be seen that the predicted values follow the actual turbidity values in most of the points which means the GP model is trained within satisfactory prediction ranges. However, it must be declared that the solution (equation 2) is complicated and implied a chaotic relation among turbidity and its historical values.

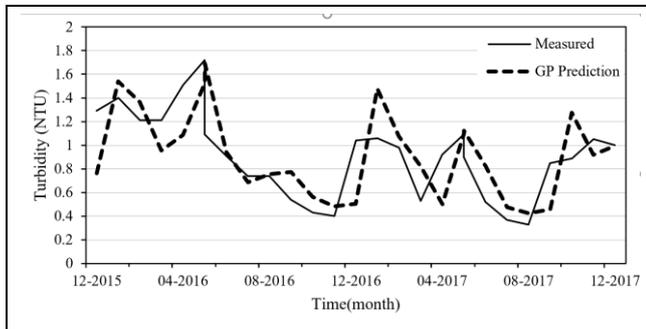


Figure 7: Observed and predicted turbidity at testing period.

Table 1 shows performance parameters for the training and testing sets of the monthly scenario. Four performance parameters were calculated: correlation coefficient (R), root mean squared error (RMSE), mean absolute percentage error (MAPE), and mean squared error (MSE). The parameters indicated that the model had higher performance for training set. This also may indicate overfitting of the model in some amount but also randomness of the data.

Table 1: Performance result of the GP model.

	Training	Testing
R	0.9022	0.6730
RMSE	0.2332	0.2898
MAPE	0.1983	0.2720
MSE	0.0544	0.0840

In the second case study GEP was trained using 70% of the standardized observed data and tested by the remaining 30 percent to formulate the relationship between low turbidity values and flow discharge. 20 different functions were randomly selected to setup GeneXproTool and the tool was allowed to use five genes and run up to 3000 generations. To avoid overfitting problem the toll was forced to stop the generation of new populations once it starts to decrease the model accuracy at validation set. The prediction results at testing period were depicted in Figure 8.

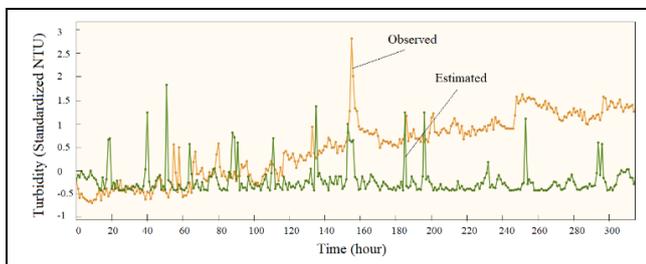


Figure 8: Observed and predicted turbidity at Gürkavak Station (Antalya).

It is clear that even the best solution was unable to provide reliable prediction. The best GEP model produced  $R^2$  (i.e., coefficient of efficiency) values less than 0.2 in both training and testing periods. Regarding to the prediction accuracy root mean squared error and root relative squared error measures

were calculated at testing period that were 1.04 and 1.54, respectively. These were standardised errors that were rather high with respect to standardised NTU (-3 to 3) range at the station. Thus, it can be concluded that there was no trustworthy interrelation between flow discharge and turbidity in the case study station. Such conclusion was also found by the application of classic GP tool (not given here).

## 5 Conclusion

GP is an evolutionary modelling technique which is able to solve different problems through the development of mathematical expressions. While the classic GP evolves a single expression the GEP provides several expressions that are combined using a user-specified mathematic or Boolean linking function. In the present study classic GP and GEP techniques were used for turbidity forecasting of drinking water at Bihac town (Bosnia and Herzegovina) and Antalya (Turkey) respectively. To this end, GPdotNET a free open source GP software developed by the authors was introduced and applied together with the well-known commercial GeneXproTools®. In order to avoid overfitting problem in the use of these evolutionary programming tools 25% to 30% of the observed data were used only for testing (validation) the evolved models.

In both use cases the results of the study pointed out that modelling of turbidity was a challenging task which required careful data analysis especially in the context of determining the optimum lag times/input parameters. Considering the results at Gürkavak Station no meaningful relation between the flow discharge and low turbidity amounts was found. Although the predictive models based on the GEP and GP algorithms may provide more accurate estimations in comparison to the traditional regression approaches it would be informative to conduct a similar study where other AI techniques are tried to model the turbidity process. Evaluation of hybrid AI models and ensemble modeling methods might be of great interest for future studies.

## 6 References

- [1] Nebbache S, Feeny V, Poudevigne I, Alarda D. "Turbidity and nitrate transfer in karstic aquifers in rural areas: the brionne basin case-study". *Journal of Environmental Management*, 62(4), 389-98, 2001.
- [2] Najah A, Elshafie A, Karim, OA, Jaffar O. "Prediction of Johor River water quality parameters using artificial neural networks". *European Journal of Scientific Research*, 28(3), 422-435, 2009.
- [3] Makić H, Koričić A, Hrnjica B. "Development of a model of predicting the sweetness level by using genetic programming". *International Conference SVAROG*, Banja Luka, Bosnia and Herzegovina, 22-23 May 2015.
- [4] Makić H, Ibrahimpašić J, Hrnjica B. "Modeling and forecasting of some una water quality parameters by genetic programming". *11th International Scientific Conference RIM*, Dubrovnik, Croatia, 2-5 October 2015.
- [5] Danandeh Mehr A, Nourani V. "A pareto-optimal moving average-multigene genetic programming model for rainfall-runoff modelling". *Environmental Modelling and Software*, 92, 239-51, 2017.
- [6] Olyae E, Abyaneh HZ, Danandeh Mehr A. "A comparative analysis among computational intelligence techniques for dissolved oxygen prediction in delaware river". *Geoscience Frontiers*, 8(3), 517-27 2016.

- [7] Mulia IE, Harold T, Roopsekhar K, Pavel T. "Hybrid ANN-GA model for predicting turbidity and chlorophyll-a concentrations". *Journal of Hydro-Environment Research*, 7(4), 279-99, 2013.
- [8] Ogston SA, Field EM. "Predictions of turbidity due to enhanced sediment resuspension resulting from sea-level rise on a fringing coral reef: evidence from Molokai, Hawaii". *Journal of Coastal Research*, 26, 1027-37, 2010.
- [9] Nakao K, Hajime N, Shuichi T. "Inverse analysis to reconstruct hydraulic conditions of non-steady turbidity currents based on multiple grain-size classes: application to an ancient turbidite of the kiyosumi formation of the Awa Group, Boso Peninsula, and Central Japan". *JPGU-AGU Joint Meeting*, Mkuhary Messe, Japan, 20-25 May 2017.
- [10] Wang Y, Wang P, Bai Y, Tian Z, Li J, Shao X, Mustavich LF, Li BL. "Assessment of surface water quality via multivariate statistical techniques: a case study of the songhua river harbin region, China". *Journal of Hydro-Environment Research*, 7(1), 30-40, 2013.
- [11] Iglesias C, Martínez Torres J, García Nieto PJ, Alonso Fernández JR, Díaz Muñoz C, Piñeiro JI, Taboada J. "Turbidity prediction in a river basin by using artificial neural networks: a case study in northern Spain". *Water Resources Management*, 28(2), 319-31, 2014.
- [12] Huang M, Tian D, Liu H, Zhang C, Yi X, Cai J, Ying, G. "A hybrid fuzzy wavelet neural network model with self-adapted fuzzy c-means clustering and genetic algorithm for water quality prediction in rivers". *Complexity*, 8241342, 1-11, 2018.
- [13] Kazemi E, Mounce S, Husband S, Boxall J. "Predicting turbidity in water distribution trunk mains using nonlinear autoregressive exogenous artificial neural networks". *13<sup>th</sup> International Conference on Hydroinformatics*, Palermo, Italy, 16 July 2018.
- [14] Koven W, Gisbert E, Nixon O, Solovyev MM, Gaon, A, Allon G, Rosenfeld H. "The effect of algal turbidity on larval performance and the ontogeny of digestive enzymes in the grey mullet (*Mugil cephalus*)". *Comparative Biochemistry and Physiology -Part A: Molecular and Integrative Physiology*, 228, 71-80, 2019.
- [15] Savary M, Johannet A, Massei N, Dupont JP, Hauchard E. Limits in Using Multiresolution Analysis to Forecast Turbidity by Neural Networks. Case Study on the Yport Basin, Normandie-France. In: Bertrand C, Denimal S, Steinmann M, Renard P. (Eds) Eurokarst 2018, Besançon. *Advances in Karst Science*. Springer, Cham, 2019.
- [16] Wang JD, Li PY, Zhang YM, Qi, WG. "River water turbidity forecasting based on phase space reconstruction and support vector regression". *2010 International Conference on Intelligent Computation Technology and Automation, ICICTA*, Changsha, China, 11-12 May 2010.
- [17] Danandeh Mehr A, Nourani V, Kahya E, Hrnjica B, Sattar A M, Yaseen ZM. "Genetic programming in water resources engineering: a state-of-the-art review". *Journal of Hydrology*, 566, 643-667, 2018.
- [18] Hrnjica B, Danandeh Mehr A. *Optimized Genetic Programming Applications: Emerging Research and Opportunities*. Hershey, USA, IGI Global, 2019.
- [19] Ferreira C. "Gene expression programming: a new adaptive algorithm for solving problems". *Complex Systems*, 13(2), 87-129, 2001.
- [20] Danandeh Mehr A, Nourani V, Hrnjica B, Molajou A. "A binary genetic programming model for teleconnection identification between global sea surface temperature and local maximum monthly rainfall events". *Journal of Hydrology*, 555, 397-406, 2017.