Uncertainty Quantification of Granular Computing-neural Network Model for Prediction of Pollutant Longitudinal Dispersion Coefficient in Aquatic Streams

Behzad Ghiasi
University of Tehran
Sun Yuanbin
Hohai University
Roohollah Noori (roohollahnoori@gmail.com)
Iran University of Science and Technology
Hossein Sheikhian
University of Tehran
Amin Zeynolabedin
University of Tehran
Changhyun Jun
Chung-Ang University
Mohamed Hamouda
National Water Center, United Arab Emirates University
Sayed M. Bateni
University of Hawaii at Manoa
Soroush Abolfathi
University of Warwick

Research Article

Keywords: Granular Computing, Prediction of Pollutant, Aquatic Streams

Posted Date: November 10th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1036627/v1

Version of Record: A version of this preprint was published at Scientific Reports on March 17th, 2022. See the published version at https://doi.org/10.1038/s41598-022-08417-4.

Uncertainty quantification of granular computing-neural network model for prediction of pollutant longitudinal dispersion coefficient in aquatic streams

Behzad Ghiasi[1,ǂ], Sun Yuanbin[2], Roohollah Noori[3,*,ǂ], Hossein Sheikhian[4,ǂ], Amin Zeynolabedin[5], Changhyun Jun[6], Mohamed Hamouda[7], Sayed M. Bateni[8], and Soroush Abolfathi[9]

[1]School of Environment, College of Engineering, University of Tehran, Tehran, 1417853111, Iran.

[2]College of Hydrology and Water Resources, Hohai University, Nanjing, 210098, China.

[3]School of Civil Engineering, Iran University of Science and Technology, Narmak, Tehran, 1684613114, Iran.

[4]Department of Geospatial Information Systems, College of Engineering, University of Tehran, Tehran, 1439957131, Iran.

[5]School of Civil Engineering, College of Engineering, University of Tehran, Tehran, 1417613131, Iran.

[6]Department of Civil and Environmental Engineering, College of Engineering, Chung-Ang University, Seoul, 06974, Korea.

[7]Civil and Environmental Engineering and the National Water Center, United Arab Emirates University, Al Ain 15551, Abu Dhabi, United Arab Emirates.

[8]Department of Civil and Environmental Engineering and Water Resources Research Center, University of Hawaii at Manoa, Honolulu, HI 96822, USA.

[9]School of Engineering, University of Warwick, Coventry, CV4 7AL, UK.

[*]roohollahnoori@gmail.com

[ǂ]These authors contributed equally to this work

Abstract

Discharge of pollution loads into natural water systems remains a global challenge that threatens water/food supply as well as endangers ecosystem services. Natural rehabilitation of the polluted streams is mainly influenced by the rate of longitudinal dispersion (D_x), a key parameter with large temporal and spatial fluctuates that characterizes pollution transport. The large uncertainty in estimation of D_x in streams limits evaluation of water quality in natural streams and design of water quality enhancement strategies. This study develops a sophisticated model coupled with granular computing and neural network models (GrC-ANN) to provide robust prediction of D_x and its uncertainty for different flow-geometric conditions with high spatiotemporal variability. Uncertainty analysis of D_x GrC-ANN model was based on the alteration of training data fed to tune the model. Modified bootstrap method was employed to generate different training patterns through resampling from a 503 global database of tracer experiments in streams. Comparison between the D_x values estimated by GrC-ANN to those determined from tracer measurements show the appropriateness and robustness of the proposed method in determining the rate of longitudinal dispersion. GrC-ANN model with the narrowest bandwidth of

estimated uncertainty (bandwidth-*factor* =0.56) that brackets the most percentage of true D_x data (i.e., 100%) is the best model to compute D_x in streams. Given considerable inherent uncertainty reported in other D_x models, the D_x GrC-ANN model is suggested as a proper tool for further studies of pollutant mixing in turbulent flow systems such as streams.

Introduction

Discharge of pollution loads into streams (i.e. rivers, manmade channels, and laboratory flumes) threatens water/food supply for mankind and aquatic biodiversity at a global scale[1,2]. The natural rehabilitation of polluted streams is mainly characterized by the rate of longitudinal dispersion (D_x or K_x), a key parameter in river water quality models with large temporal and spatial fluctuates. A challenging task in the study of the pollutant fate and transport in turbulent flow systems (e.g., streams) is determining D_x for numerical and analytical water quality models[3,4]. This is because D_x is the most predominant factor influencing the pollutant concentration at downstream of the point of accidental pollution[5,8]. Starting from the late 1960s, the mechanism of D_x determination in streams was introduced by Fischer[9]. From this, Fischer[10] proposed an analytical formula to estimate D_x that required the full and detailed knowledge of the flow-geometric conditions of the system under study.

Given that the flow-geometric data for streams are highly variable in temporal and spatial scales, such data are not readily measured and available. Also, the complex numerical procedures required to solve Fischer[10] equation, have led to several simplifications to determine D_x. Hence, the predictions of D_x from the simplified models often deviate from the field-estimated measurements up to several orders of magnitude[11,12]. These simplifications are mainly exclusion of some difficult accessible variables such as flow-geometric irregularities that influence dispersion mechanism in streams. Although in many cases the impact of the excluded variables is somewhat embedded in other variables used in D_x estimation models, they do not fully represent the complex interactions between the absent variables and D_x. For example, friction term (i.e., rate of flow velocity to shear velocity – U/U^*), as a simply accessible input for D_x estimation models, to some extend can represent the impact of lateral and vertical irregularities in streams that affect the rate of dispersion[13]. However, these irregularities produce shear flows and secondary currents that can alternate the D_x. Simultaneously, the former causes an increase in D_x whilst the latter decreases D_x[8,14-18]. The complex interactions between the flow-geometric data and dispersion mechanism prohibit reaching an accurate estimation of D_x in streams whilst some effective variables on dispersion mechanism are excluded (e.g., stream bed shape factor and sinuosity).

In recent decades, and with the advancement in artificial intelligence (AI) models, they became powerful tools to solve complex engineering problems[19-29]. A number of AI-based studies have conducted to enhance the accuracy of D_x estimation in turbulent flow systems such as streams[30,33]. Given that AI techniques are able to map the complex non-linear input-output relationships even when some important information are missing[34], their applications in estimating the D_x have been investigated by several studies[30,33,35-44]. However, complex nature of dispersion mechanism in turbulent flow systems with variations in both spatial and temporal scales, as well as the inevitable simplification assumptions that are needed for modelling will result in uncertainty of D_x using AI models. In a study conducted by Noori et al.[45], they reported that although the AI techniques outperform traditional empirical models for estimation of D_x, they are still subjected to uncertainty induced by changes in their training patterns. The inaccuracy in estimation of the D_x using AI models can limit water quality assessment and design of appropriate measures to improve the water quality of aquatic flows. Hence, developing robust methodological framework to quantify the prediction uncertainty of the D_x from AI models is essential for developing appropriate AI-based water quality models.

Granular computing (GrC) model is a highly efficient AI-based model which has recently shown an excellent potential to solve complex problems in engineering community[42,46,47]. GrC model is a novel tool

capable of applying the granules in the process of nonlinear problem solving[48]. In the GrC model, the natural rules between the data are extracted by means of the rule mining algorithm, operating on a set of information arranged as information table. The granule measures involved in the process of information mining, has made GrC as a powerful tool to map a set of inputs to a set of outputs in different fields of science and engineering[49,50].

However, similar to other AI models, the GrC performance can be adversely influenced by the selection of training patterns. Therefore, the effects of changes in training patterns on the performance of GrC model should be investigated, to understand and quantify the degree of uncertainty in the model's prediction of D_x in water quality assessments. In this study, we first replaced an artificial neural network (ANN) with rules' information in the GrC to improve its performance (GrC-ANN). Then, we developed a D_x estimation model using GrC-ANN model. Finally, a comprehensive uncertainty analysis method was proposed for the model-estimated D_x in streams, to compare the accuracy of D_x predicted by GrC-ANN with other AI-based D_x models in the literature.

By considering the above, this study, for the first time, highlights the application, strengthens and weaknesses of GrC-ANN model in the field of hydrological and water quality engineering. More importantly, this study introduces a framework for uncertainty determination of the GrC-ANN model, a subject with no documentation in literature, that is beneficial for reliable application of this model in water-environmental community and other fields of science/engineering.

Methods

Longitudinal dispersion. Non-reactive pollutant mixing process in aquatic systems is a complex three-dimensional (3-D) flow process, consists of molecular and turbulent diffusion, and shear dispersion (simply referred to "dispersion") mechanism where the latter is the net trace of velocity shear over the flow width and depth, and the turbulent mixing[11]. In the natural streams that are specifically much longer than width or depth of the flow, the pollutants become well-mixed in the vertical and transverse directions rather than the longitudinal mixing (Fig. 1). Therefore, pollutant fate and transport in streams is usually studied by the application of 1-D dispersion model quantified by the advection-dispersion equation, i.e., Eq. (1)[51,52].

$$\frac{\partial C}{\partial t} + U\frac{\partial C}{\partial x} = D_x\frac{\partial^2 C}{\partial x^2} \quad (1)$$

In Eq. (1), C and U, are the average cross-sectional concentration and average longitudinal velocity, respectively; t and x are time and longitudinal coordinate in the stream-wise direction, respectively.

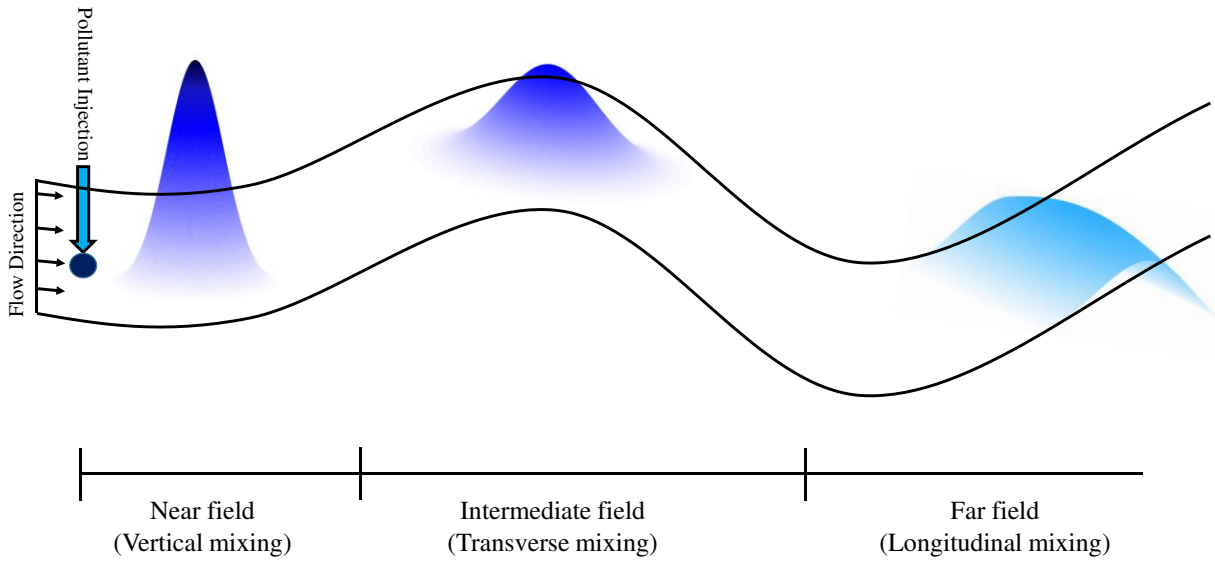


Figure 1. Schematic view of pollutant mixing in streams (adapted from Kilpatrick and Wilson[53]).

Ideally, vertical and (transverse) dispersion in streams takes place close to (in intermediate fields from) the pollutant discharge location, whilst the longitudinal dispersion occurs far from the point of pollutant discharge where solute become readily well-mixed in both vertical and transverse directions (Fig. 1). In streams, the longitudinal dispersion usually varies form 10^{-1} to 10^{7} m^2/s[10,13,54,55] and the diffusion coefficient ranges from 10^{-9} (molecular) to 10^{-2} m^2/s (turbulent)[5]. Therefore, dispersion is the dominant mechanism of mixing process, by several orders of magnitude[56], highlighting the necessity of developing robust methodological approach to quantify the dispersion and mixing coefficient in the streams.

***D_x* parametrization.** Pioneering work on quantification of dispersion mechanism in pipes date back to Taylor's studies[57,58]. Thereafter, Taylor's approach was used for quantifying dispersion in streams with the assumption of no limits for the width of the channel[59]. However, the Elder's formula underestimates the dispersion in natural streams since it does not consider the lateral velocity shear[10,60]. In streams, the lateral velocity shear mechanism plays a more dominant role in determining the mixing compared to the vertical shear. On this basis, Fischer[9] derived an analytical formula for determining D_x:

$$D_x = \left\{\int_0^W h(y)\acute{u}(y) \int_0^y [1/\varepsilon_t(y)h(y)] \int_0^y h(y)\acute{u}(y)\, dy\, dy\, dy\right\}/A \quad (2)$$

where, W denotes the local flow width, x is the longitudinal coordinate, y is the lateral coordinate, $\acute{u}(y)$ is the local velocity deviation, $h(y)$ represents the local flow depth, $\varepsilon_t(y)$ is local lateral mixing coefficient, and A represents the local flow cross-sectional area.

In Eq. (2), the flow is supposed to be 1-D, i.e., the pollutant is mixed in both vertical and lateral directions well, a condition that is rarely satisfied in turbulent flow systems such as natural streams and even in laboratory flumes due to existence of secondary currents[61]. Also, Fischer[9] equation has been derived based on the assumption that the dispersion is controlled by lateral shear rather than vertical shear, a condition that may not satisfy well for the narrow and deep rivers where aspect ratio (i.e., river flow width to depth – W/H) is small[5]. Such drawbacks of Eq. (2) lead to inaccurate estimation of D_x compared to those values determined from tracer experiments. The deviation between D_x values estimated by Eq. (2) and those true values is maximum for the case of non-uniform flow in real streams, although Fischer[9] model can well approximate the dispersion for the case of uniform flow[62]. In addition to the inherent drawbacks in practical application of the Eq. (2), it also requires detailed information on the

geometrical properties (i.e. cross-section, bathymetry) of stream, as well as the lateral flow velocity profiles. Collecting such information is rather costly and time consuming, and often requires very detailed flow measurements which are not readily available. Therefore, practical application of Fischer[9] model is limited.

To address the difficulties in using Eq. (2), Fischer[63] suggested a simplified empirical equation that correlated D_x with pertinent dimensionless variables of W/H and U/U^* (Eq. (3)).

$$\frac{D_x}{HU^*} = a\left(\frac{W}{H}\right)^b \left(\frac{U}{U^*}\right)^c \quad (3)$$

Fischer[63] modified formula for determining the dispersion coefficient (Eq. (3)), has been widely used and validated by other researchers and rely on parameters which can be practically determined.

Data collection. This study aims to estimate D_x in streams using GrC-ANN model. In this regard, a global tracer database consisting of 503 observations from natural streams and laboratory flumes was used to develop the model and validate the performance of the proposed GrC-ANN model. This database was compiled by Riahi-Madvar et al.[64], consisting of a comprehensive range of flow conditions, stream geometrical properties, and pollution loading. It should be noted that although the database used in this study is more comprehensive compared to other studies on D_x estimation, it does not fully include high extreme values of D_x[12].

GrC-ANN development. In-depth description and model development for the GrC and GrC-ANN approach for environmental pollution modelling were given in Noori et al.[42] and Ghiasi et al.[46], respectively. and detailed information about these models documented by Sheikhian et al.[47,48] conducted comprehensive review of the models and their performance. Hence, we shortened the descriptions of GrC-ANN model developed in this study.

The GrC applies different measures to construct the high-quality classification rules in the form of IF-THEN statements[65]. In this study, GrC extracts the rules from the global tracer database consisting of 503 observations from natural streams and laboratory flumes based on conditional entropy (*CE*) and absolute support (*AS*) so that rules with the minimum *CE* value and the maximum *AS* are extracted from the database. To form a granular decision tree, the priority of rules in the tree is determined based on higher generality (*G)* and coverage (*CV)*.

A basic GrC model has two major deficiencies. First, it prioritizes rules based on their obtained parameters and uses the first rule satisfied by input data to define its output[66,67]. Second, after classification, some of the input patterns may remain unclassified if they do not meet the requirements of the classification rule set. Hence, the GrC-ANN model proposed in this study uses an integration of GrC rule generation and ANN model (Fig. 2). The proposed modelling approach benefiting the information gathered from existing relationship among extracted rules from GrC algorithm, by applying a group of rules providing information about the input pattern of data (i.e., W/H, U/U^* and D_x/HU^*), ranked by the rule relevance measurements undertaken by the GrC. This approach allows the model to use the mentioned rule quality parameters to construct the approximator structure, instead of common time-consuming iterative learning procedure[46].

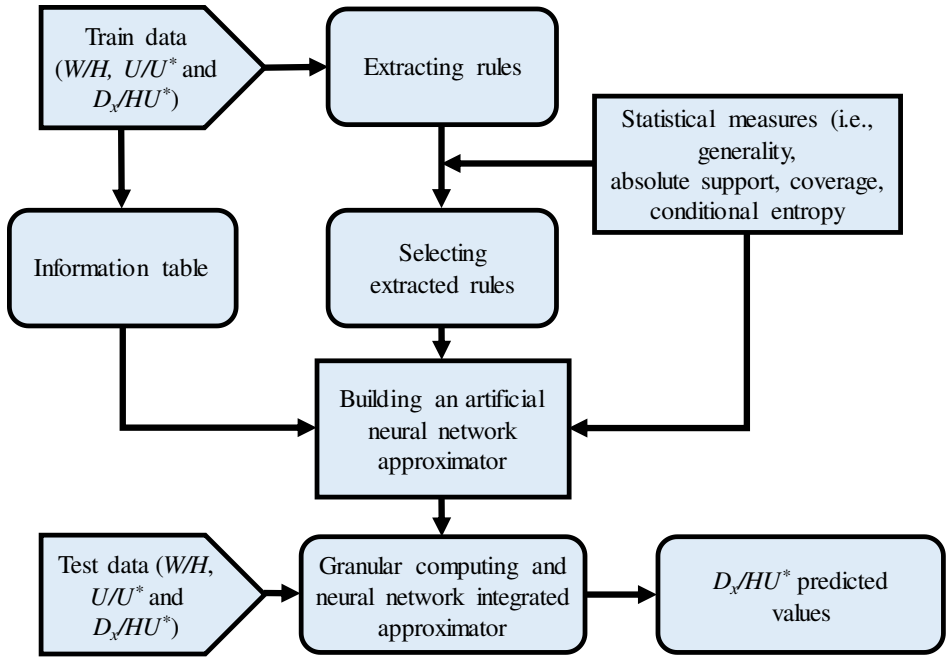


Figure 2. Procedure of integrating GrC and ANN for determining longitudinal dispersion coefficient.

The GrC-ANN structure proposed in this paper comprises layers as a conventional neural network including the input layer, two computing layers, and the output (aggregation) layer (Fig. 3). The layers are customized to support the proposed idea. The input layer contains nodes equal to attributes of the data records (i.e., W/H, U/U^*, and D_x/HU^*). Computing layers comprise from two inner-connected layers including pattern layer and rule firing layer. These layers within the computing layer, receive values that are valid according to the criteria presented in the input layer. The third layer contains the set of qualified extracted rules by GrC-ANN and embeds the classification rules. The aggregation layer assigns an output value to the input pattern of the data. The connection weights of the rule-firing layer and the aggregation layer are given by the statistical measure of absolute support provided by the corresponding rule to its output value, to consider the accuracy of the rules in determining that output value.

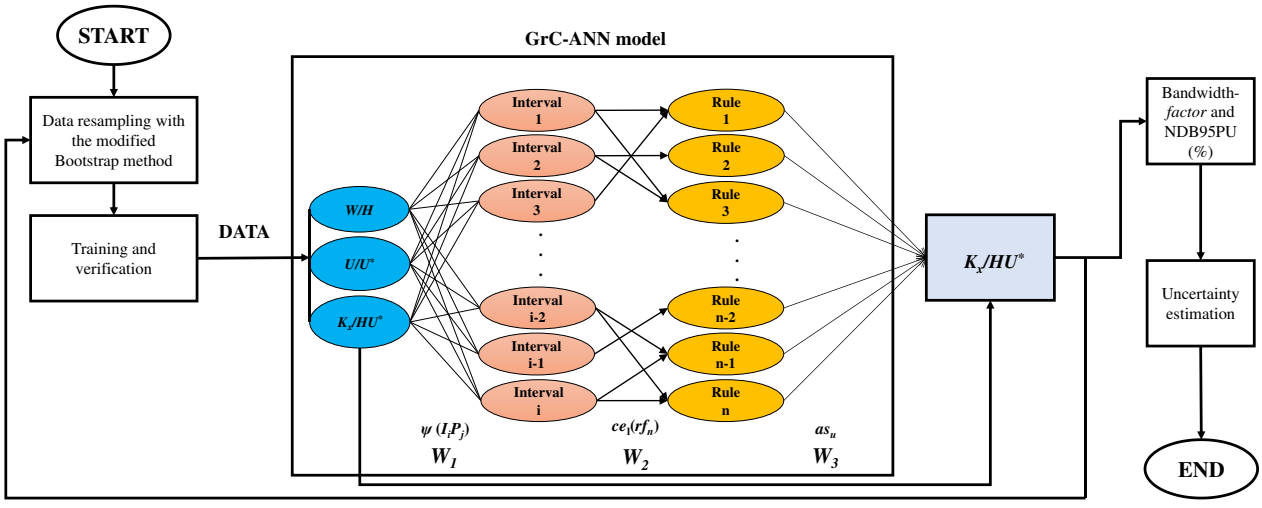


Figure 3. Schematic of the GrC-ANN modelling details of this study.

Uncertainty determination. Similar to other data-driven models, the GrC-ANN model minimizes the error function based on the fed data with the aid of a supervised algorithm throughout the training process[49]. Hence, training plays a vital role in quantification of the model's uncertainty caused by different tuning sets. In this study, the GrC-ANN model was tuned to map inputs W/H and U/U^* to target D_x/HU^* based on finite training patterns resampled from 503 observations of the global tracer database. Probabilistically, each training pattern used for tuning the GrC-ANN model is different from others resampled from the global database of tracer experiments. Thus, each training pattern could produce different set of GrC-ANN parameters as well as different outputs for estimation of D_x/HU^*.

The modified bootstrap method suggested by Noori et al.[12] was used to resample distinct training patterns for tuning D_x/HU^* GrC-ANN model. This method ensures that the chosen training patterns are fully representative of the statistical characteristics of the 503 tracer experiments of the global database used in this study. This is important since the global database used in this study rarely has large D_x instances[12], denoting that these large dispersion values are likely under-represented in the training patterns chosen by the traditional bootstrap technique. This issue can result in poor training of D_x/HU^* GrC-ANN model and consequently increase the model's uncertainty. Detailed description on the bootstrap method is given by Efron and Tibshirani[68], while Noori et al.[12] modified the bootstrap method.

However, we used different outputs of the D_x/HU^* GrC-ANN model in verification stage, resulted from the change in the training patterns, as a measure of the model's uncertainty[69]. We computed an interval band of the GrC-ANN estimations of D_x/HU^* with a level of significant of 95%. Then, two measures were used to assess variations in the different responses of the D_x/HU^* GrC-ANN model in verification stage including bandwidth-*factor* and the number of bracketed D_x/HU^* data using 95% of predicted uncertainties (NBD95PU) as shown in Eqs. (4) and (5), respectively[70]. Given these two measures, the uncertainty in estimation of the D_x/HU^* GrC-ANN model in verification stage was quantified.

$$\text{bandwidth-}factor = \left\{\left(\frac{1}{n}\right)\sum_{i=1}^{n}(X_U - X_L)\right\}/\sigma_x \quad (4)$$

$$\text{NBD95PU}\ (\%) = (1/n)\text{count}\{Q|(X_L \le Q \le X_U)\} \quad (5)$$

where σ_x is the standard deviation of the target D_x/HU^*, and X_U and X_L are the maximum and minimum of the estimated D_x/HU^* for each training pattern, respectively.

Figure 3 illustrates a detailed description of the model development and uncertainty quantification process proposed for this study.

3. Results and discussion

Tuned GrC-ANN models. The correlation among W, H, U, U^* and D_x is shown in Fig. 4A. The correlation coefficients for the model variables in dimensionless format, i.e. W/H, U/U^* and D_x/HU^* and the corresponding statistical significance level are illustrated in Fig. 4B. In dimensional form, D_x/HU^* is more correlated with the geometrical configuration W/H of the stream (correlation coefficient =0.21, p-value <0.1) than the flow characteristic U/U^* (correlation coefficient =0.002, p-value >0.1), which confirm the results reported by Noori et al.[12].

W	0.72 (p-value <0.01)	0.15 (p-value <0.01)	0.00 (p-value >0.1)	0.18 (p-value <0.01)
0.72 (p-value <0.01)	H	0.08 (p-value <0.01)	0.00 (p-value >0.1)	0.11 (p-value <0.01)
0.15 (p-value <0.01)	0.08 (p-value <0.01)	U	0.07 (p-value < 0.01)	0.22 (p-value <0.01)
0.00 (p-value >0.1)	0.00 (p-value >0.1	0.07 (p-value < 0.01)	U^*	0.03 (p-value <0.01)
0.18 (p-value <0.01)	0.11 (p-value <0.01)	0.22 (p-value <0.01)	0.03 (p-value <0.01)	D_x

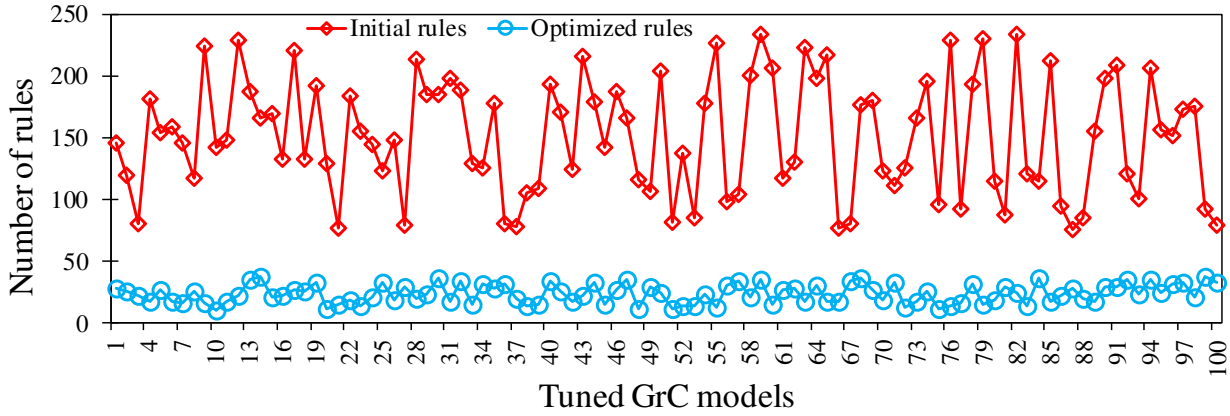
(A)

W/H	0.002 (p-value >0.1)	0.21 (p-value <0.01)
0.002 (p-value >0.1)	U/U^*	0.01 (p-value >0.1)
0.21 (p-value <0.01)	0.01 (p-value >0.1)	D_x/HU^*

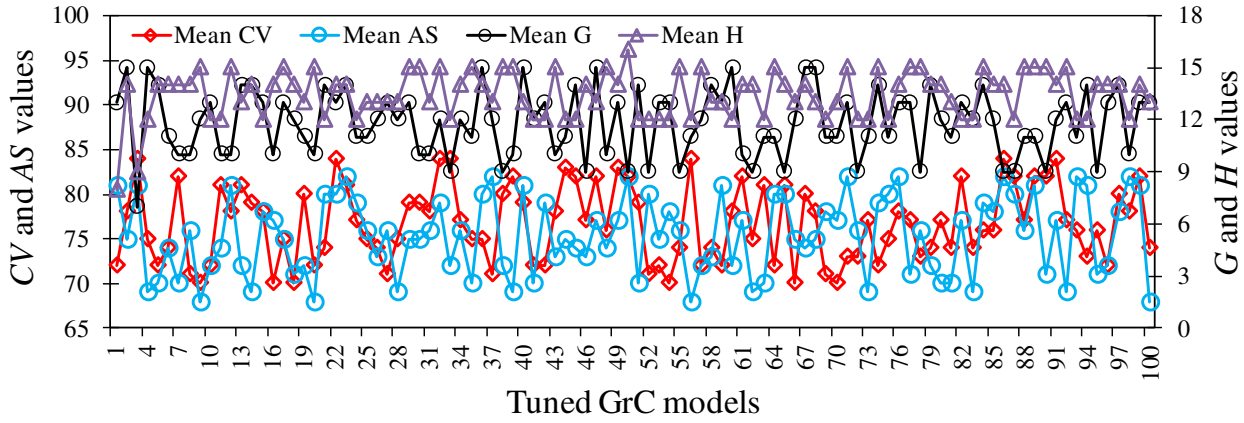
(B)

Figure 4. The correlation coefficient plots of (A) W, H, U, U^*, and D_x, (B) W/H, U/U^* and D_x/HU^*.

To apply the GrC-ANN model, the 503 observations from real streams and laboratory flumes were scaled between 0 and 1. 40 data were excluded away from the 503 observations of global tracer experiments for the model verification. Then, 100 distinct training patterns were randomly resampled from the remaining database, i.e. 463 observations, with replacement to tune 100 different D_x/HU^* GrC-ANN models. Each training pattern consists of 80 data, and the 40 pre-assigned verification data. The model inputs include, aspect ratio and friction term, and dimensionless target D_x/HU^* were clustered based on their indiscernibility in the given attributes. To form final rule network, the GrC-based rule extraction algorithm was used to select the best granules of information by considering the measurements CE, AS, G, and CV computed for each rule. In this regard, AS and CE indices were employed to extract the set of possible valid rules by considering minimum and maximum threshold values of 0.75 and 0.5 for these parameters, respectively, in accordance to the similar studies in the literature[42,71,72]. At this stage, if a rule caused redundancy in the rule set, it was considered as an active granule and was replaced with a granule that had more consistency in the set of rules. Using the proposed methodology led to extraction of a range of rules, varied from 76 to 234, for tuning the GrC models based on the training patterns (Fig. 5A).



(A)



(B)

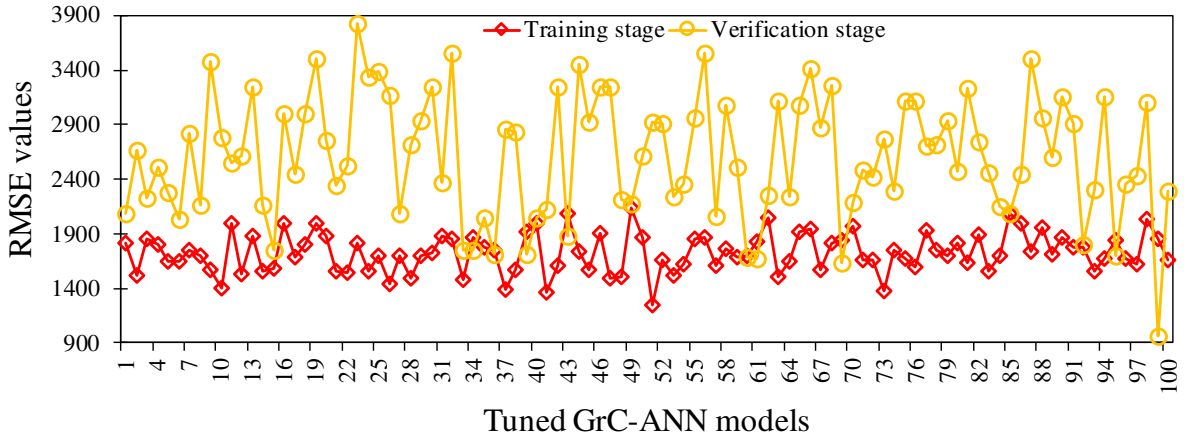
Figure 5. (A) Number of initial and optimized rules, and (B) the mean values of quality indices for the final rules for each tuned GrC models.

In the next step, the *CV* and *G* indices were applied to prioritize the rules resulted in the final rule sets. The optimized rules varied from 10 to 38 for the models tuned based on the training patterns (Fig. 5A). The mean values of quality indices for the final rules selected for each tuned model are illustrated in Fig. 5B. According to Fig. 5B, the *G* values ranged between 0 and 0.4 indicating the rules' generality does not pertain to big values of *G*, which confirms the results of previous GrC modelling studies[42,66]. The *CV* varied between 0 and 1, pertaining to the numbers of extracted rules by each class and dataset covered by each rule, following Yao and Yao[66] suggestion.

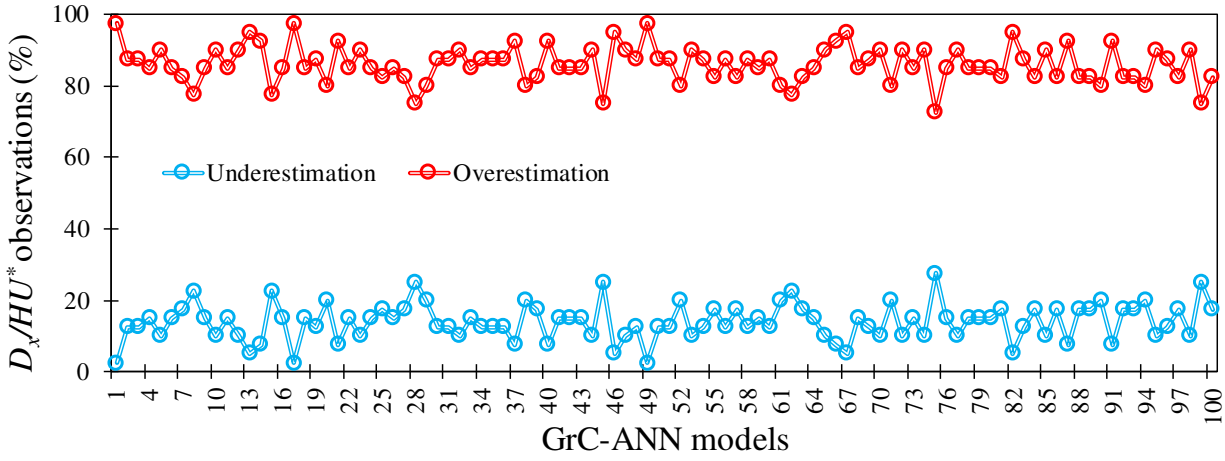
100 optimized rule sets computed correspond to one hundred distinct training patterns are then fed to the GrC-ANN modelling structure. In this regard, the rule quality indices were embedded into an ANN structure instead of initial weights, forming a GrC-ANN model corresponding to each optimized rule set. The best network structures describing the relations between the inputs (W/H and U/U^*) and the output (D_x/HU^*) data were determined based on the quality index of root mean square error (RMSE) for each GrC-ANN model tuned by the distinct training patterns (Fig. 6A). Analysis of the results show the RMSE values for the tuned D_x/HU^* GrC-ANN models, in training and verification stages varied from 1251 to 2142 and 966 to 3826, respectively (Fig. 6A).

Figure 7 shows the difference between the true (field-estimated) D_x/HU^* values and those predicted by each tuned D_x/HU^* GrC-ANN model. The minimum (i.e., -10934) and the maximum (i.e., 7471) errors were produced in D_x/HU^* GrC-ANN models #42 and #100, respectively. In general, the

GrC-ANN models overestimate the D_x/HU^* values for around 86% of observations (Fig. 6B). Such overestimation of D_x was reported by Etemad-Shahidi and Taghipour[73] for the D_x models proposed by Liu[55], Seo and Cheong[13], Deng et al.[51], and Sahay and Dutta[74]. However, using the overestimated D_x/HU^* values in corporation with 1-D ADE models give lower maximum concentration rate at locations which are far from the pollutant injection point[12]. Therefore, the tuned D_x/HU^* GrC-ANN model must be used with caution in hydro-environmental studies such as outfall design and/or risk assessment researches from accidental hazardous pollution.



(A)



(B)

Figure 6. (A) Root mean square error (RMSE) values calculated for the tuned D_x/HU^* GrC-ANN models in training and verification stages, and (B) D_x/HU^* observations (%) with underestimation and overestimation in GrC-ANN models tuned by the distinct training patterns.

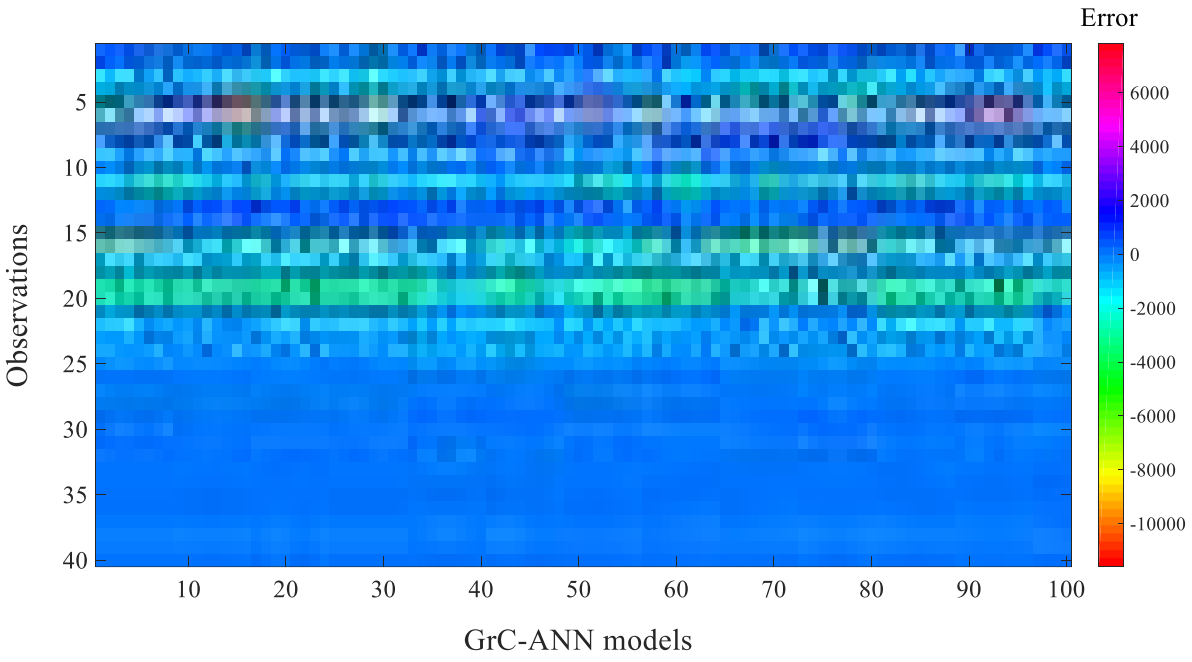


Figure 7. Difference between the true D_x/HU^* values and those predicted using GrC-ANN models tuned by the distinct training patterns.

The GrC prioritizes rules based upon their obtained parameters and uses the first rule satisfied by aspect ratio and friction term data to define dimensionless term D_x/HU^* as the model output. In addition, some of the input patterns may remain unclassified if they do not meet the requirements of the classification rule set in the GrC model[67]. Meanwhile, GrC-ANN uses the final rule set selected by GrC instead of the initial weights[46]. In other words, a conventional ANN model provides results which is influenced by initial weights generated in a random manner, yielding to different results from the same set of training information. In addition, the proposed D_x/HU^* GrC-ANN model replaces the learning part of the ANN with the information from rule quality measures and ensures that no connection or node are remained without a transparent description. The GrC-ANN modelling approach described in the present study is a significant improvement to the conventional ANN modelling approaches which contain hidden neurons obtaining their connection weights by learning through a black-box learning algorithm. Hence, the tuned D_x/HU^* GrC-ANN theoretically provide a more robust framework than both GrC and ANN models. Previous studies also confirmed the performance superiority of GrC compared to ANN and adaptive neuro fuzzy inference system (ANFIS) developed for D_x/HU^* predictions (e.g., Ghiasi et al.[46]). Comparative analysis of the tuned GrC-ANN models developed in this study, and other AI models including model tree (MTree), gene-expression programming (GEP), evolutionary polynomial regression (EPR), support vector machine (SVM), and multivariate adaptive regression splines (MARS), developed by Najafzadeh et al.[75], highlights that the proposed GrC-ANN models are capable of better and more robust approximation of longitudinal dispersion (D_x/HU^*) in stream (Fig. 8). As shown in Fig. 8, the determination coefficient (R^2) values determined for the GrC-ANN models in verification stage, are much larger than those reported for ERP and MARS models.

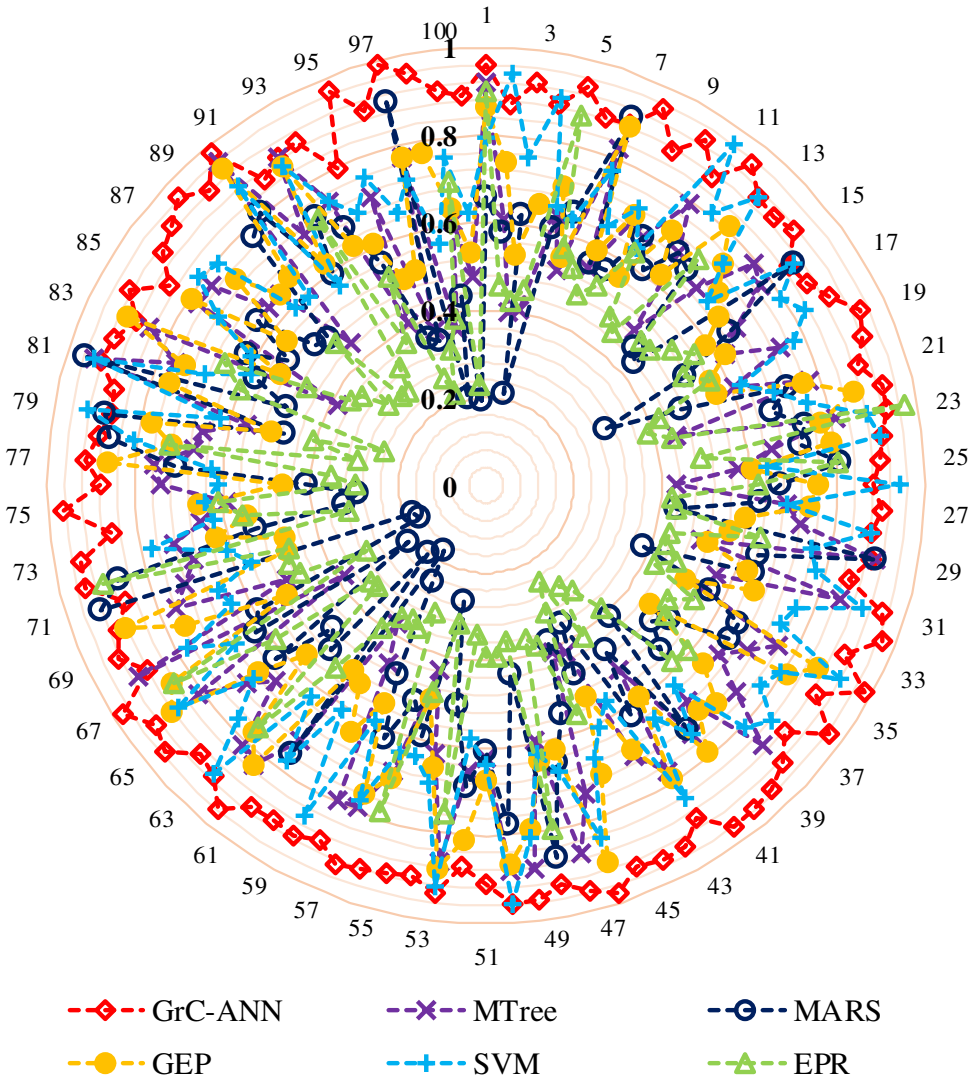


Figure 8. Determination of coefficient (R^2) values calculated for D_x/HU^* prediction during the verification step of the tuned GrC-ANN models developed in this study, and those reported for model tree (MTree), gene-expression programming (GEP), evolutionary polynomial regression (EPR), support vector machine (SVM) and multivariate adaptive regression splines (MARS)) by Najafzadeh et al.[75].

3.2. GrC-ANN uncertainty. The D_x/HU^* values estimated during the verification stage by the 100 GrC-ANN models tuned under distinct training patterns were used to measure the model uncertainty. In this regard, prediction intervals corresponding to each D_x/HU^* observation was computed by considering the level of significant of 95% (Fig. 9). These prediction intervals show the deviation from the true D_x/HU^* values, denoting the uncertainty associated with the GrC-ANN predictions of longitudinal dispersion in streams.

Figure 9 shows that the true D_x/HU^* values are fully located between the lower and upper bands of the uncertainty, concluding the appropriate performance of the GrC-ANN model based on the NDB95PU (%) index. Also the small value of the bandwidth-*factor* (= 0.56) indicates the small deviation of the predicted D_x/HU^* values by the GrC-ANN models from the measured values, leading to low uncertainty of the model. Fig. 9 shows that the proposed GrC-ANN model has good performance in predicting both large and small D_x/HU^* values with a narrow bandwidth of uncertainty, highlighting the model

superiority in predicting the D_x/HU^* compared to other AI models which are suffering from large uncertainty in estimation of D_x[12,42,45,75].

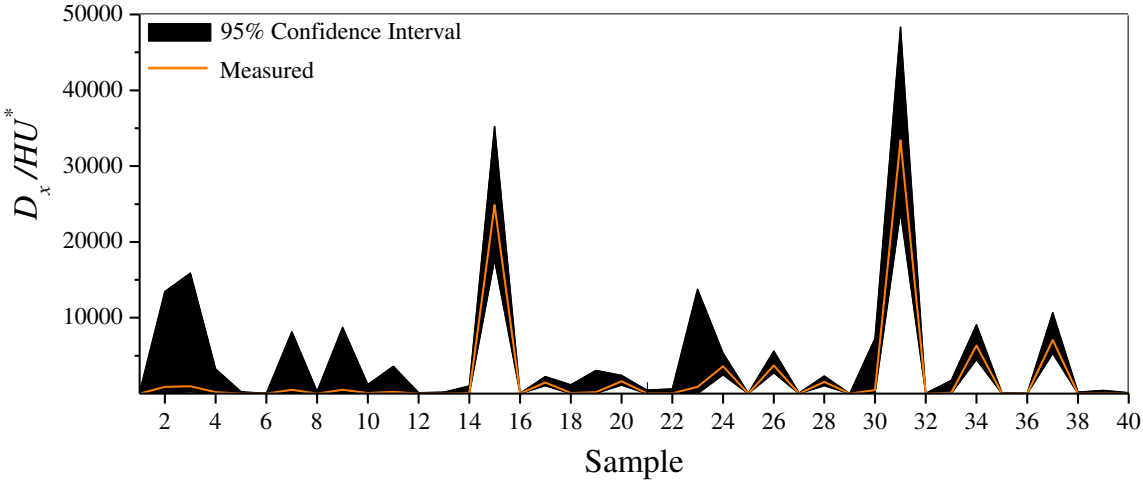


Figure 9. GRC-ANN model uncertainty for estimation of D_x/HU^* in streams.

However, neither the GrC-ANN model nor other mathematical and statistical models can fully understand and predict the dispersion processes in real streams. Therefore, the results illustrated in Fig. 9 still contain some degree of uncertainty in the prediction of D_x/HU^* from GrC-ANN model. To compare the uncertainty of the predicted D_x/HU^* from GrC-ANN with other AI models, the bandwidth-*factor* and NDB95PU (%) values computed for these models are illustrated in Fig. 10. This figure shows that D_x/HU^* GrC-ANN model has the smallest bandwidth-*factor* value amongst the 9 AI-based models examined in this study. Also, D_x/HU^* GrC-ANN model has the largest NDB95PU (%) value compared to other AI models (i.e., EPR, MTree, GEP, SVM, MARS, ANN, and ANFIS). These measures suggest that the uncertainty in the prediction of D_x/HU^* from GrC-ANN model is far less than those reported for other well-established AI models for the case of pollutant transport in streams.

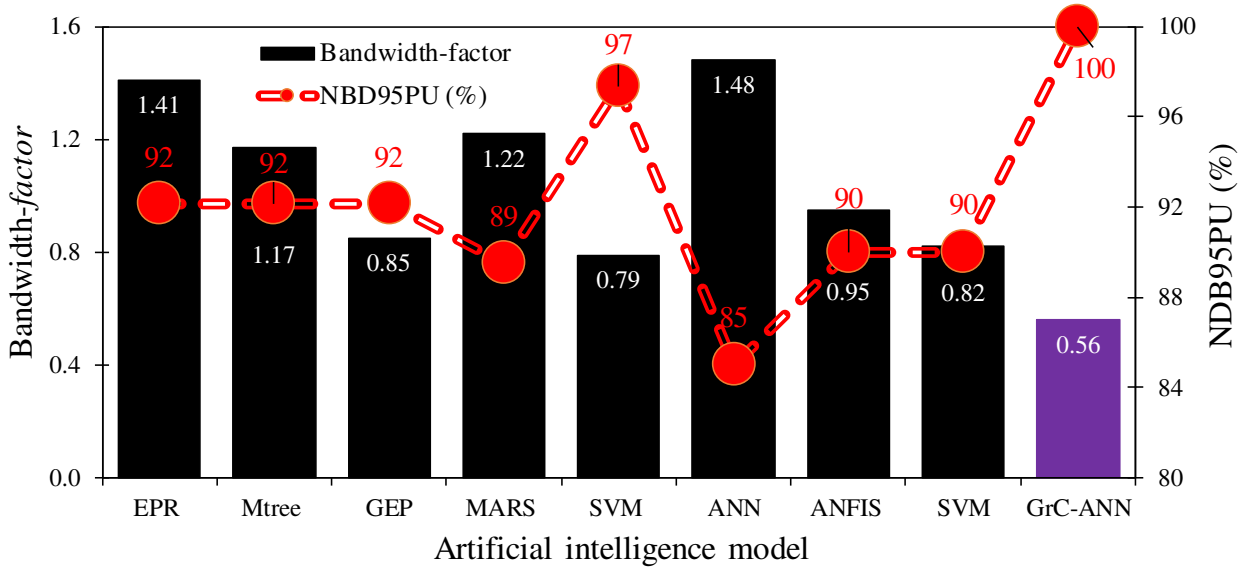


Figure 10. Comparison of the bandwidth-*factor* and the NDB95PU (%) values of the GrC-ANN model developed in this study (i.e., D_x/HU^* GrC-ANN), with ANN and ANFIS models[45], SVM, GEP, MTree, MARS, and EPR models[75].

A review of the past studies indicates that the accuracy of the models suggested for estimation of D_x largely varies depending on the choice of AI modelling approach and the dataset used[76]. For example, the accuracy of ANN model developed for estimation of D_x, measured as R^2, varied from 0.34[76] to 0.98[77]. The R^2 value for the SVM model is ranging from 0.4[76] and 0.76[78]. Also, ANFIS model developed for the estimation of D_x in streams show an R^2 value varying from 0.50[79] to 0.96[46]. The same results have also been reported for traditional empirical studies where R^2 value varies from 0.36[80] to 0.96[42]. Such large variations in the accuracy of D_x estimation clearly specify the importance of training patterns used to tune the models for predicting the rate of longitudinal dispersion in streams, an important issue that generally overlooked in the previous studies. Therefore, introducing a D_x model tuned based on a single training pattern does not provide a comprehensive and robust understanding of the possible change of D_x in real case studies, where D_x largely fluctuates and influenced by the irregularities in flow hydrodynamics and the geometric characteristics of streams.

However, study of the Fig. 9 reveals that despite modified and enhanced training patterns adopted in this study, there remains some uncertainty in the prediction of the $D_x/HU^*$from GrC-ANN model, which can be considerable at times and leading to a wide confidence interval band for some samples. In fact, in the D_x/HU^* GrC-ANN modelling process, some rules are eliminated due to low criteria values (i.e., *G*, *AS*, *CV*, and *CE*). Therefore, the selected rules, which govern the final prediction of the model, do not fully represent the complex mechanisms of the longitudinal dispersion in streams, leading to inevitable uncertainty in the predictions by GrC-ANN model. In addition, diversity of streams and the irregularities in geometric characteristics and nonlinearity of the flow hydrodynamics add to the complexity of the mixing mechanisms in the streams. Therefore, full identification, quantification and inclusion of these intricate natural processes in a mathematical or statistical model is not possible. This is correct even for the non-simplified models for prediction of D_x, i.e. Eq. (2), where estimated D_x values are still not in full agreement with those values measured in the field. For example, the minimum error between the estimated and field-measurement of D_x values occurs for the case of a uniform flow, that is usually less than 30%[62]. In the case of non-uniform flow in large meandrous streams with severe irregularities in bathymetry, and spatiotemporal variations in flow hydrodynamics, the estimated D_x using Eq. (2) usually deviates from the field measurements by several orders of magnitude[11]. The problem of inaccuracy in modelling predictions raises up when using Eq. (3), derived based on simplified assumptions for Eq. (2), and by exclusion of important parameters influencing D_x such as S_f and S_n[5,11,16,80,811]. These excluded parameters are seldom monitored in streams due to the difficulties associated with their measurement. Another factor that contribute to the uncertainty in prediction of longitudinal dispersion from GrC-ANN model is the rare presence of very large D_x values in the dataset used in this study. Analysis of the dataset used in this study shows that only around 1% of the 503 global dataset of tracer experiments consists of D_x >1000 m^2/s, whilst the maximum value of D_x in the dataset is around 1800 m^2/s[12]. This absence of very large D_x in the dataset, is leading to uncertainty in the D_x/HU^* predicted by the GrC-ANN model.

Conclusions

Rate of longitudinal dispersion (D_x) dominantly influences the pollutant transport and fate in streams. Given the high spatiotemporal variability of D_x, previous single-based trained models cannot robustly understand the complex nonlinear interactions between the parameters that govern the longitudinal dispersion in the streams and cannot capture the uncertainty associated with the predictive models for D_x in streams. This provides rigorous methodological approach to examine and quantify the uncertainty in the prediction of D_x/HU^* from the proposed GrC-ANN model, by studying the distinct training patterns used to optimize the model's parameters. The detailed analysis of the results highlights that although D_x/HU^* predicted by GrC-ANN model outperform other AI-based dispersion models, there remains

some uncertainty in the predicted D_x from the model which need careful consideration and evaluation. This finding suggests that river water quality assessments and management studies should consider the impacts of uncertainty associated with the D_x estimation on the pollutant concentrations, that could result in detrimental impacts on aquatic biodiversity, and ecosystem function in streams as well as human health. Enhanced data on the flow hydrodynamics and the geometric features in streams (e.g., stream sinuosity and bed shape factor) for the D_x models can further reduce the uncertainty in estimation of longitudinal dispersion parameter.

Data availability

The data used in this study can be obtained from https://doi.org/10.1007/s11269-018-2139-6

References

1. Atique, U., Kwon, S. and An, K.G., 2020. Linking weir imprints with riverine water chemistry, microhabitat alterations, fish assemblages, chlorophyll-nutrient dynamics, and ecological health assessments. *Ecological Indicators*, 117, 106652. https://doi.org/10.1016/j.ecolind.2020.106652
2. Kim, J.Y., Atique, U. and An, K.G., 2021. Relative Abundance and Invasion Dynamics of Alien Fish Species Linked to Chemical Conditions, Ecosystem Health, Native Fish Assemblage, and Stream Order. Water, 13(2), 158. https://doi.org/10.3390/w13020158
3. Ramezani, M., Noori, R., Hooshyaripor, F., Deng, Z. and Sarang, A., 2019. Numerical modelling-based comparison of longitudinal dispersion coefficient formulas for solute transport in rivers. *Hydrological Sciences Journal*, *64*(7), 808-819. https://doi.org/10.1080/02626667.2019.1605240
4. Abolfathi, S., Cook, S., Yeganeh-Bakhtiary, A., Borzooei, S. and Pearson, J. M., 2020. Microplastics transport and mixing mechanisms in the nearshore region. *Coastal Engineering Proceedings 36v*. https://doi.org/10.9753/icce.v36v.papers.63
5. Rutherford, J C. (1994). River Mixing. John Wiley & Sons, Chichester, U K, 347 pp.
6. Abolfathi, S., and Pearson, J. M., 2017. Application of smoothed particle hydrodynamics (SPH) in nearshore mixing: a comparison to laboratory data. *Coastal Engineering Proceedings (35)*. https://doi.org/10.9753/icce.v35.currents.16
7. Cook, S., Chan, H. L., Abolfathi, S., Bending, G. D., Schäfer, H. and Pearson, J. M., 2020. Longitudinal dispersion of microplastics in aquatic flows using fluorometric techniques. *Water Research*, 170, 115337. https://doi.org/10.1016/j.watres.2019.115337
8. Cheme, E.K. and Mazaheri, M., 2021. The effect of neglecting spatial variations of the parameters in pollutant transport modeling in rivers. *Environmental Fluid Mechanics*, 21(3), 587-603. https://doi.org/10.1007/s10652-021-09787-5
9. Fischer, H.B., 1967. The mechanics of dispersion in natural streams. *Journal of the Hydraulics division*, *93*(6), 187-216. https://doi.org/10.1061/JYCEAJ.0001706
10. Fischer, H.B., 1968. Methods for Predicting Dispersion Coefficients in Natural Streams: With Applications to Lower Reaches of the Green and Duwamish Rivers, *Washington* (Vol. 582). US Government Printing Office.
11. Deng, Z.Q., Bengtsson, L., Singh, V.P. and Adrian, D.D., 2002. Longitudinal dispersion coefficient in single-channel streams. *Journal of Hydraulic Engineering*, *128*(10), 901-916. https://doi.org/10.1061/(ASCE)0733-9429(2002)128:10(901)
12. Noori, R., Mirchi, A., Hooshyaripor, F., Bhattarai, R., Haghighi, A.T. and Kløve, B., 2021. Reliability of functional forms for calculation of longitudinal dispersion coefficient in rivers. *Science of The Total Environment*, 148394. https://doi.org/10.1016/j.scitotenv.2021.148394
13. Seo, I.W. and Cheong, T.S., 1998. Predicting longitudinal dispersion coefficient in natural streams. *Journal of hydraulic engineering*, 124(1), 25-32. https://doi.org/10.1061/(ASCE)0733-9429(1998)124:1(25)

14. Nezu, I., Tominaga, A. and Nakagawa, H., 1993. Field measurements of secondary currents in straight rivers. *Journal of Hydraulic Engineering*, *119*(5), 598-614. https://doi.org/10.1061/(ASCE)0733-9429(1993)119:5(598)
15. Deng, Z.Q. and Singh, V.P., 1999. Mechanism and conditions for change in channel pattern. *Journal of Hydraulic Research*, *37*(4), 465-478. https://doi.org/10.1080/00221686.1999.9628263
16. Marion, A. and Zaramella, M., 2006. Effects of velocity gradients and secondary flow on the dispersion of solutes in a meandering channel. *Journal of Hydraulic Engineering*, *132*(12), 1295-1302. https://doi.org/10.1061/(ASCE)0733-9429(2006)132:12(1295)
17. Bashitialshaaer, R., Bengtsson, L., Larson, M., Persson, K.M., Aljaradin, M. and Hossam, A.I., 2011. Sinuosity effects on longitudinal dispersion coefficient. *Int. J. of Sustainable Water and Environmental Systems*, *2*(2), 77-84.
18. Nikora, V. and Roy, A.G., 2012. Secondary flows in rivers: Theoretical framework, recent advances, and current challenges. *Gravel bed rivers: Processes, tools, environments*, 3-22. https://doi.org/10.1002/9781119952497.ch1
19. Kişi, Ö., 2009. Modeling monthly evaporation using two different neural computing techniques. *Irrigation Science*, 27(5), 417-430. https://doi.org/10.1007/s00271-009-0158-z
20. Khatibi, R., Ghorbani, M.A., Kashani, M.H. and Kisi, O., 2011. Comparison of three artificial intelligence techniques for discharge routing. *Journal of Hydrology*, 403(3-4), 201-212. https://doi.org/10.1016/j.jhydrol.2011.03.007
21. Abolfathi, S., Yeganeh-Bakhtiary, A., Hamze-Ziabari, S.M., Borzooei, S., 2016. Wave runup prediction using M5' model tree algorithm. *Ocean Engineering*, 112, 76-81. https://doi.org/10.1016/j.oceaneng.2015.12.016
22. Granata, F., Papirio, S., Esposito, G., Gargano, R. and De Marinis, G., 2017. Machine learning algorithms for the forecasting of wastewater quality indicators. *Water*, 9(2), 105. https://doi.org/10.3390/w9020105
23. Jaramillo, F., Orchard, M., Muñoz, C., Antileo, C., Sáez, D. and Espinoza, P., 2018. On-line estimation of the aerobic phase length for partial nitrification processes in SBR based on features extraction and SVM classification. *Chemical Engineering Journal*, 331, 114-123. https://doi.org/10.1016/j.cej.2017.07.185
24. Ahmadi, M.H., Ahmadi, M.A., Nazari, M.A., Mahian, O. and Ghasempour, R., 2019. A proposed model to predict thermal conductivity ratio of Al_2O_3/EG nanofluid by applying least squares support vector machine (LSSVM) and genetic algorithm as a connectionist approach. *Journal of Thermal Analysis and Calorimetry*, 135(1), 271-281. https://doi.org/10.1007/s10973-018-7035-z
25. Leottau, D.L., Lobos-Tsunekawa, K., Jaramillo, F. and Ruiz-del-Solar, J., 2019. Accelerating decentralized reinforcement learning of complex individual behaviors. *Engineering Applications of Artificial Intelligence*, 85, 243-253. https://doi.org/10.1016/j.engappai.2019.06.019
26. Ghimire, S., Yaseen, Z.M., Farooque, A.A., Deo, R.C., Zhang, J., Tao, X., 2021. Streamflow prediction using an integrated methodology based on convolutional neural network and long short-term memory networks. *Scientific Reports*, 11(1), 17497. https://doi.org/10.1038/s41598-021-96751-4
27. Moosavi, S.R., Wood, D.A., Ahmadi, M.A. and Choubineh, A., 2019. ANN-based prediction of laboratory-scale performance of CO_2-foam flooding for improving oil recovery. *Natural Resources Research*, 28(4), 1619-1637. https://doi.org/10.1007/s11053-019-09459-8
28. Borzooei, S., Miranda, G. H. B., Abolfathi, S., Scibilia, G., Meucci, L., Zanetti, M. C., 2020. Application of unsupervised learning and process simulation for energy optimization of a WWTP under various weather conditions. *Water Science & Technology*, 81(8), 1541-1551. https://doi.org/10.2166/wst.2020.220
29. Yaseen, Z.M., Ali, M., Sharafati, A., Al-Ansari, N., Shahid, S., 2021. Forecasting standardized precipitation index using data intelligence models: regional investigation of Bangladesh. *Scientific Reports*, 11(1), 3435. https://doi.org/10.1038/s41598-021-82977-9

30. Afan, H.A., Allawi, M.F., El-Shafie, A., Yaseen, Z.M., Ahmed, A.N., Malek, M.A., Koting, S.B., Salih, S.Q., Mohtar, W.H.M.W., Lai, S.H. and Sefelnasr, A., 2020. Input attributes optimization using the feasibility of genetic nature inspired algorithm: Application of river flow forecasting. *Scientific Reports*, 10(1), 4684. https://doi.org/10.1038/s41598-020-61355-x
31. Tayfur, G. and Singh, V.P., 2005. Predicting longitudinal dispersion coefficient in natural streams by artificial neural network. *Journal of Hydraulic Engineering*, *131*(11), 991-1000. https://doi.org/10.1061/(ASCE)0733-9429(2005)131:11(991)
32. Toprak, Z.F., Hamidi, N., Kisi, O. and Gerger, R., 2014. Modeling dimensionless longitudinal dispersion coefficient in natural streams using artificial intelligence methods. *KSCE Journal of Civil Engineering*, 18(2), 718-730. https://doi.org/10.1007/s12205-014-0089-y
33. Parsaie, A., Emamgholizadeh, S., Azamathulla, H.M. and Haghiabi, A.H., 2018. ANFIS-based PCA to predict the longitudinal dispersion coefficient in rivers. *International Journal of Hydrology Science and Technology*, 8(4), 410-424. https://doi.org/10.1504/IJHST.2018.095537
34. Azar, N.A., Milan, S.G. and Kayhomayoon, Z., 2021. The prediction of longitudinal dispersion coefficient in natural streams using LS-SVM and ANFIS optimized by Harris hawk optimization algorithm. *Journal of Contaminant Hydrology*, *240*, 103781. https://doi.org/10.1016/j.jconhyd.2021.103781
35. Di Nunno, F. and Granata, F., 2020. Groundwater level prediction in Apulia region (Southern Italy) using NARX neural network. *Environmental Research*, 190, 110062. https://doi.org/10.1016/j.envres.2020.110062
36. Tayfur, G., 2006. Fuzzy, ANN, and regression models to predict longitudinal dispersion coefficient in natural streams. *Hydrology Research*, *37*(2), 143-164. https://doi.org/10.2166/nh.2006.0012
37. Toprak, Z.F. and Cigizoglu, H.K., 2008. Predicting longitudinal dispersion coefficient in natural streams by artificial intelligence methods. *Hydrological Processes: An International Journal*, *22*(20), 4106-4129. https://doi.org/10.1002/hyp.7012
38. Toprak, Z.F. and Savci, M.E., 2007. Longitudinal dispersion coefficient modeling in natural channels using fuzzy logic. *CLEAN–Soil, Air, Water*, *35*(6), 626-637. https://doi.org/10.1002/clen.200700122
39. Piotrowski, A.P., Rowinski, P.M. and Napiorkowski, J.J., 2012. Comparison of evolutionary computation techniques for noise injected neural network training to estimate longitudinal dispersion coefficients in rivers. *Expert Systems with Applications*, *39*(1), 1354-1361. https://doi.org/10.1016/j.eswa.2011.08.016
40. Sahay, R.R., 2013. Predicting longitudinal dispersion coefficients in sinuous rivers by genetic algorithm. *Journal of Hydrology and Hydromechanics*, *61*(3), 214. https://doi.org/10.2478/johh-2013-0028
41. Najafzadeh, M. and Tafarojnoruz, A., 2016. Evaluation of neuro-fuzzy GMDH-based particle swarm optimization to predict longitudinal dispersion coefficient in rivers. *Environmental Earth Sciences*, *75*(2), 157. https://doi.org/10.1007/s12665-015-4877-6
42. Noori, R., Ghiasi, B., Sheikhian, H. and Adamowski, J.F., 2017. Estimation of the dispersion coefficient in natural rivers using a granular computing model. *Journal of Hydraulic Engineering*, *143*(5), 04017001. https://doi.org/10.1061/(ASCE)HY.1943-7900.0001276
43. Kargar, K., Samadianfard, S., Parsa, J., Nabipour, N., Shamshirband, S., Mosavi, A. and Chau, K.W., 2020. Estimating longitudinal dispersion coefficient in natural streams using empirical models and machine learning algorithms. *Engineering Applications of Computational Fluid Mechanics*, *14*(1), 311-322. https://doi.org/10.1080/19942060.2020.1712260
44. Riahi-Madvar, H., Dehghani, M., Parmar, K.S., Nabipour, N. and Shamshirband, S., 2020. Improvements in the explicit estimation of pollutant dispersion coefficient in rivers by subset selection of maximum dissimilarity hybridized with ANFIS-firefly algorithm (FFA). *IEEE Access*, *8*, 60314-60337. https://doi.org/10.1109/ACCESS.2020.2979927
45. Noori, R., Deng, Z., Kiaghadi, A. and Kachoosangi, F.T., 2016. How reliable are ANN, ANFIS, and SVM techniques for predicting longitudinal dispersion coefficient in natural rivers?. *Journal of Hydraulic Engineering*, *142*(1), 04015039. https://doi.org/10.1061/(ASCE)HY.1943-7900.0001062

46. Ghiasi, B., Sheikhian, H., Zeynolabedin, A. and Niksokhan, M.H., 2019. Granular computing–neural network model for prediction of longitudinal dispersion coefficients in rivers. *Water Science and Technology*, *80*(10), 1880-1892. https://doi.org/10.2166/wst.2020.006
47. Sheikhian, H., Delavar, M.R. and Stein, A., 2017. A GIS-based multi-criteria seismic vulnerability assessment using the integration of granular computing rule extraction and artificial neural networks. *Transactions in GIS*, *21*(6), 1237-1259. https://doi.org/10.1111/tgis.12274
48. Yao, Y., 2004. A partition model of granular computing. In Transactions on rough sets I (232-253). *Springer*, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-27794-1_11
49. Sheikhian, H., Delavar, M.R. and Stein, A., 2015. Integrated estimation of seismic physical vulnerability of Tehran using rule based granular computing. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, *40*(3), 187. https://doi.org/10.5194/isprsarchives-XL-3-W3-187-2015
50. Khamespanah, F., Delavar, M.R., Moradi, M. and Sheikhian, H., 2016. A GIS-based multi-criteria evaluation framework for uncertainty reduction in earthquake disaster management using granular computing. *Geodesy and Cartography*, *42*(2), 58-68. https://doi.org/10.3846/20296991.2016.1199139
51. Deng, Z.Q., Singh, V.P. and Bengtsson, L., 2001. Longitudinal dispersion coefficient in straight rivers. *Journal of hydraulic engineering*, *127*(11), 919-927. https://doi.org/10.1061/(ASCE)0733-9429(2001)127:11(919)
52. Barati Moghaddam, M., Mazaheri, M. and MohammadVali Samani, J., 2017. A comprehensive one-dimensional numerical model for solute transport in rivers. *Hydrology and Earth System Sciences*, 21(1), 99-116. https://doi.org/10.5194/hess-21-99-2017
53. Kilpatrick, F.A. and Wilson, J.F., 1989. Measurement of time of travel in streams by dye tracing (Vol. 3). *US Government Printing Office*.
54. Iwasa, Y. and Aya, S., 1991. Transverse mixing in a river with complicated channel geometry. *Bulletin of the Disaster Prevention Research Institute*, *41*(3), 129-175.
55. Liu, H., 1977. Predicting dispersion coefficient of streams. *Journal of the Environmental Engineering Division*, *103*(1), 59-69. https://doi.org/10.1061/JEEGAV.0000605
56. Smith, R., 1992. 'Physics od Dispersion' coastal and estuarine pollution – methods and soloutions' technical sessions, *Scottish hydraulic study group*, One day seminar 3rd Aprill, Glasgow.
57. Taylor, G.I., 1954. The dispersion of matter in turbulent flow through a pipe. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, *223*(1155), 446-468. https://doi.org/10.1098/rspa.1954.0130
58. Taylor, G.I., 1953. Dispersion of soluble matter in solvent flowing slowly through a tube. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, *219*(1137), 186-203. https://doi.org/10.1098/rspa.1953.0139
59. Elder, J., 1959. The dispersion of marked fluid in turbulent shear flow. *Journal of fluid mechanics*, *5*(4), 544-560. https://doi.org/10.1017/S0022112059000374
60. Fischer, H.B., 1966. Longitudinal dispersion in laboratory and natural streams. *California Institute of Technology*. https://doi.org/10.7907/Z9F769HC
61. Carr, M.L. and Rehmann, C.R., 2007. Measuring the dispersion coefficient with acoustic Doppler current profilers. *Journal of Hydraulic Engineering*, *133*(8), 977-982. https://doi.org/10.1061/(ASCE)0733-9429(2007)133:8(977)
62. Papadimitrakis, I. and Orphanos, I., 2004. Longitudinal dispersion characteristics of rivers and natural streams in Greece. *Water, Air and Soil Pollution: Focus*, *4*(4), 289-305. https://doi.org/10.1023/B:WAFO.0000044806.98243.97
63. Fischer, H.B., List, J.E., Koh, C.R., Imberger, J. and Brooks, N.H., 1979. Mixing in inland and coastal waters. *Academic press*, New York.
64. Riahi-Madvar, H., Dehghani, M., Seifi, A. and Singh, V.P., 2019. Pareto optimal multigene genetic programming for prediction of longitudinal dispersion coefficient. *Water resources management*, *33*(3), 905-921. https://doi.org/10.1007/s11269-018-2139-6

65. Yao, Y.Y., 2001, October. On modeling data mining with granular computing. In *25th Annual International Computer Software and Applications Conference. COMPSAC 2001* (638-643). IEEE. https://doi.org/10.1109/CMPSAC.2001.960680
66. Yao, J.T. and Yao, Y.Y., 2002, October. Induction of classification rules by granular computing. In *International Conference on Rough Sets and Current Trends in Computing* (331-338). Springer, Berlin, Heidelberg. https://doi.org/10.1007/3-540-45813-1_43
67. Yao, Y.Y. and Zhong, N., 2002. Granular computing using information tables. *Data mining, rough sets and granular computing*, 102-124. https://doi.org/10.1007/978-3-7908-1791-1_5
68. Efron, B. and Tibshirani, R.J., 1994. An introduction to the bootstrap. *CRC press*.
69. Srivastav, R.K., Sudheer, K.P. and Chaubey, I., 2007. A simplified approach to quantifying predictive and parametric uncertainty in artificial neural network hydrologic models. *Water Resources Research*, *43*(10). https://doi.org/10.1029/2006WR005352
70. Abbaspour, K.C., Yang, J., Maximov, I., Siber, R., Bogner, K., Mieleitner, J., Zobrist, J. and Srinivasan, R., 2007. Modelling hydrology and water quality in the pre-alpine/alpine Thur watershed using SWAT. *Journal of Hydrology*, *333*(2-4), 413-430. https://doi.org/10.1016/j.jhydrol.2006.09.014
71. Bello, R., Falcón, R. and Pedrycz, W. eds., 2007. Granular computing: at the junction of rough sets and fuzzy sets (Vol. 224). *Springer*.
72. Koh, Y.S. and Rountree, N. eds., 2009. Rare Association Rule Mining and Knowledge Discovery: Technologies for Infrequent and Critical Event Detection: Technologies for Infrequent and Critical Event Detection (Vol. 3). *IGI Global.*
73. Etemad-Shahidi, A. and Taghipour, M., 2012. Predicting longitudinal dispersion coefficient in natural streams using M5′ model tree. *Journal of hydraulic engineering*, *138*(6), 542-554. https://doi.org/10.1061/(ASCE)HY.1943-7900.0000550
74. Sahay, R.R. and Dutta, S., 2009. Prediction of longitudinal dispersion coefficients in natural rivers using genetic algorithm. *Hydrology Research*, *40*(6), 544-552. https://doi.org/10.2166/nh.2009.014
75. Najafzadeh, M., Noori, R., Afroozi, D., Ghiasi, B., Hosseini-Moghari, S.M., Mirchi, A., Haghighi, A.T. and Kløve, B., 2021. A comprehensive uncertainty analysis of model-estimated longitudinal and lateral dispersion coefficients in open channels. *Journal of Hydrology*, 126850. https://doi.org/10.1016/j.jhydrol.2021.126850
76. Ghiasi, B., Jodeiri, A. and Andik, B., 2021. Using a deep convolutional network to predict the longitudinal dispersion coefficient. *Journal of Contaminant Hydrology*. 240, 103798. https://doi.org/10.1016/j.jconhyd.2021.103798
77. Sahay, R.R., 2011. Prediction of longitudinal dispersion coefficients in natural rivers using artificial neural network. *Environmental Fluid Mechanics*, *11*(3), 247-261. https://doi.org/10.1007/s10652-010-9175-y
78. Adarsh, S., 2010. Prediction of longitudinal dispersion coefficient in natural channels using soft computing techniques. *Scientia Iranica*, 17(5), 363-371.
79. Noori, R., Karbassi, A., Farokhnia, A. and Dehghani, M., 2009. Predicting the longitudinal dispersion coefficient using support vector machine and adaptive neuro-fuzzy inference system techniques. *Environmental Engineering Science*, *26*(10), 1503-1510. https://doi.org/10.1089/ees.2008.0360
80. Memarzadeh, R., Zadeh, H.G., Dehghani, M., Riahi-Madvar, H., Seifi, A., Mortazavi, S.M., 2020. A novel equation for longitudinal dispersion coefficient prediction based on the hybrid of SSMD and whale optimization algorithm. *Science of The Total Environment*, *716*, 137007. https://doi.org/10.1016/j.scitotenv.2020.137007
81. Dehghani, M., Zargar, M., Riahi-Madvar, H., Memarzadeh, R., 2020. A novel approach for longitudinal dispersion coefficient estimation via tri-variate archimedean copulas. *Journal of Hydrology*, 584, 124662. https://doi.org/10.1016/j.jhydrol.2020.124662

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Author contributions

Data collection and analysis were carried out by **B.G.**, **S.U.**, **H.S.** and **A.Z. R.N.** conceived the study conceptually. The models were ran by **R.N.** and **H.S.** The first draft of manuscript was prepared by **B.G.**, **R.N.**, **S.U.**, **C.J.** and **M.H.** The funding acquisition was made by **C.J.** and **M.H.** The analyses and results were supervised and validated by **R.N.**, **S.M.B.** and **S.A.** All authors read and approved the final version of manuscript.