Desalination and Water Treatment
www.deswater.com

doi: 10.5004/dwt.2020.25746

190 (2020) 440–448
June

Prediction of a rosette dense jet group in crossflow ambient conditions using multi-gene genetic programming

Xiaohui Yan*, Abdolmajid Mohammadian

Department of Civil Engineering, University of Ottawa, Ottawa, Canada, 161 Louis Pasteur, Ottawa Ontario, K1N6N5, Canada, emails: clarkyanxh@gmail.com (X. Yan), Majid.Mohammadian@uottawa.ca (A. Mohammadian)

Received 28 June 2019; Accepted 14 February 2020

ABSTRACT

A new approach for predicting the mixing properties of a rosette dense jet group in crossflow ambient conditions based on multi-gene genetic programming (MGGP) is presented. Explicit MGGP formulations for the dimensionless terminal rise height (y_t/dF), the dimensionless impact distance (x_i/dF), and the impact dilution (S_i/F) are developed. Experimental data are used for training and testing the models. The performances of the MGGP models are found to be superior to the existing empirical equations, which are not sufficiently accurate to be used in conceptual rosette diffuser design. A confidence analysis is also conducted for the developed models. The results demonstrate that the MGGP technique is a promising tool to evolve an explicit, accurate and relatively compact mathematical model to predict a rosette dense jet group. The developed MGGP models can facilitate accurate estimation of the primary mixing characteristic parameters of a rosette dense jet group in crossflow ambient conditions for the investigated data range, and they could be continuously and quickly improved or extended with the availability of more training data.

Keywords: Artificial intelligence; Crossflow; Dense jet; Multi-gene genetic programming; Rosette diffuser

1. Introduction

Wastewater effluents are often discharged into the receiving water through rosette diffusers [1–3]. Such diffusers are commonly used for coastal desalination and disposal of municipal wastewater, and some examples include the outfall diffusers in Melbourne, San Francisco, and Sydney [2]. Improper outfall design may result in inadequate disposal of wastewater effluents, which could, in turn, lead to serious environmental and ecological issues [4]. Therefore, it is important to predict the mixing properties of the effluents for a sound outfall design or performance assessment [5,6].

Depending on the density difference between the jets and the ambient water, a rosette jet group can be classified into two categories. A rosette buoyant jet group refers to the scenarios in which the jets have a lower density than the ambient water, while a rosette dense jet group corresponds to the cases in which the jets have a higher density than the receiving water. Compared with rosette buoyant jets, the studies on rosette dense jets are relatively few, and thus the mixing properties of rosette dense jets require further investigation and better predicting tools.

As shown in Fig. 1, a rosette diffuser consists of a cluster of upward-inclined ports that are located around a circle. The dense jets are discharged from the ports with an initial velocity of u_j to the ambient water that has a velocity of u_a. In the regions close to the nozzles, the jets move upward due to the initial momentum and spread as they mix with the ambient water. Because of the momentum dissipation and negative buoyancy, the jets reach a terminal rise height, y_t, and then start falling back toward the seafloor. The jets reach the seabed at the return point that is located x_i from the diffuser. This distance is known as the impact distance, and the dilution at this location is known as the impact dilution,

* Corresponding author.

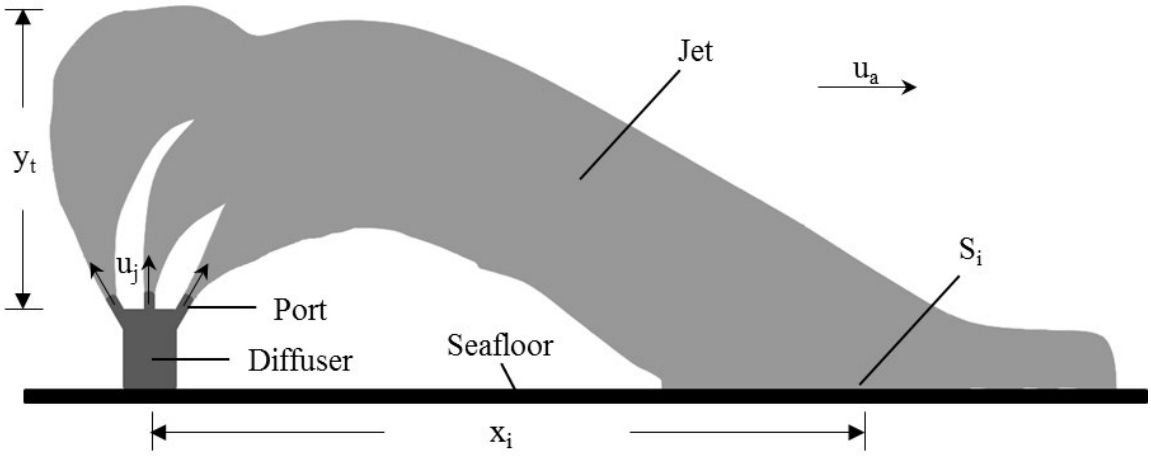

Fig. 1. Schematic diagram of a rosette dense jet group in crossflow ambient conditions.

S_i. If the jets have sufficient clearance from other jets or boundaries, such as single jets, the jets are efficiently diluted due to the entrainment of the ambient water. However, for rosette jets, there could be dynamic interactions or the Coanda effect between the individual jets, which can affect the mixing properties of the jets [7,8]. These properties can be characterized by y_t, x_i, and S_i, so predicting these jet characteristic parameters is of significant importance to a better understanding of the effluent mixing properties.

A rosette dense jet group can be studied using various methods, such as physical and numerical modeling approaches. Although these approaches can provide reliable data, they are either costly or computationally expensive, so their application in practical engineering problems is restricted. Simplified theoretical or regression-based empirical equations are quite useful in practical applications, especially for a rough estimate of the jet properties. However, these equations may contain errors due to some predetermined assumptions or the ignorance of some hidden nonlinear effects. Abessi and Roberts [3] have developed empirical models for rosette dense jets, but their *R*-squared values were below 0.90. Therefore, the existing empirical equations are not sufficiently accurate to be used in conceptual rosette diffuser design, and the subject requires further investigation.

Artificial intelligence (AI) techniques provide a new avenue for predicting engineering data. These techniques, such as an artificial neural network, adaptive neuro-fuzzy inference system, grey forecasting, genetic programming (GP), and artificial neural network-genetic algorithm, have been widely applied and validated in water resources problems in recent years [9–16]. The multi-gene genetic programming (MGGP) technique is a recent advancement of GP [17,18]. An MGGP model consists of several traditional GP genes that can capture nonlinear behavior and these genes are linearly combined. This technique effectively combines the capability of the standard GP in describing nonlinear behavior and that of the ordinary least squares methods in estimating regression coefficients and has been found to be better than both the standard GP and conventional regression methods [18,19]. Therefore, MGGP is a potentially promising tool to evolve an explicit, accurate, and relatively compact mathematical model to predict a rosette dense jet group. It is acknowledged that there are many well-established numerical and theoretical models in the field of jet mixings, such as those reported by Yan and Mohammadian [20,21] and Yan et al. [22], but developing MGGP models are still beneficial in three respects: first, the mechanisms for the mixing processes of a rosette dense jet group in crossflow ambient conditions are very complicated, and thus a well-recognized numerical and theoretical model has rarely been reported; second, an MGGP model is much more efficient than a numerical model; third, an MGGP model can be continuously improved with the availability of more experimental or observational data.

The primary objective of this work is twofold: first, to develop explicit models for the mixing characteristic parameters of a rosette dense jet group in crossflow ambient conditions, which are more accurate than the existing empirical equations for the investigated data range; and second, to demonstrate that the MGGP algorithm is a promising tool for investigating rosette dense jet groups, and thus the present study will encourage researchers or engineers to collect more data that are suitable for training the MGGP models and continuously improve or extend the models in further studies. The models were trained and tested with available experimental data. The formulations are also presented and compared with existing empirical equations. Recently, Yan and Mohammadian [23–25] have successfully utilized AI algorithms to develop compact and explicit models for laterally confined buoyant jets, vertical buoyant jets, and multiple inclined dense jets. However, the dilution processes of a rosette dense jet group in crossflow ambient conditions are quite different from and more complicated than these types of discharges, and thus further evaluating and applying the MGGP algorithm to develop explicit equations for a rosette dense jet group in crossflow ambient conditions is necessary. To the best of the authors' knowledge, this is the first work using MGGP to predict the mixing characteristic parameters of a rosette dense jet group in crossflow ambient conditions. The developed models can facilitate accurate estimation of the primary mixing characteristic parameters

for the investigated data range, and they could be continuously and quickly improved or extended with the availability of more training data.

2. Methods

2.1. Analysis of a rosette dense jet group

The most important parameters characterizing a jet are the riser height, y_t, impact distance, x_i, and impact dilution, S_i (Fig. 1). It is well known that the mixing properties of a single jet are mainly governed by the jet densimetric Froude number [26,27], F, which is defined as

$$F = \frac{u_j}{\sqrt{g'd}} \tag{1}$$

with

$$g' = g\frac{\rho_a - \rho_j}{\rho_a} \tag{2}$$

where u_j is the initial jet velocity, d is the port diameter, g is the gravitational acceleration, ρ_a is the density of ambient water, and ρ_j is the initial jet density.

The present work focuses on a rosette dense jet group in crossflow an ambient condition, which is also affected by the riser spacing, s_r, and the ambient current speed, u_a. Therefore, the dimensionless riser height, y_t/dF, the dimensionless impact distance, x_i/dF, and the dimensionless impact dilution, S_i/F, can be expressed as functions of the dimensionless riser spacing, s_r/dF and the dimensionless velocity, u_rF, as [3]

$$\frac{y_t}{F} = f\left(\frac{s_r}{dF}, u_rF\right) \tag{3}$$

$$\frac{x_i}{dF} = f\left(\frac{s_r}{dF}, u_rF\right) \tag{4}$$

$$\frac{S_i}{F} = f\left(\frac{s_r}{dF}, u_rF\right) \tag{5}$$

where u_r is defined as u_a/u_j.

Abessi and Roberts [3] have proposed comprehensive empirical Eqs. (6)–(8) for estimating these variables, which are summarized as follows:

$$\frac{y_t}{dF} = \left[0.09\left(u_rF\right) - 0.45\right]\log\left(\frac{s_r}{dF}\right) + \left[-0.36\left(u_rF\right) + 2.40\right] \tag{6}$$

$$\frac{x_i}{dF} = \left[0.28\left(u_rF\right) - 0.05\right]\log\left(\frac{s_r}{dF}\right) + \left[2.63\left(u_rF\right) + 0.63\right] \tag{7}$$

$$\frac{S_i}{F} = \left[0.15\left(u_rF\right) + 0.69\right]\log\left(\frac{s_r}{dF}\right) + \left[0.20\left(u_rF\right) + 0.53\right] \tag{8}$$

The R-squared values for theses empirical equations were below 0.90. Therefore, they are not sufficiently accurate to be used in conceptual rosette diffuser design, and thus the subject requires further investigations.

2.2. Experimental data

The experimental data of [3] were used for training and testing the MGGP models. The experiments were designed for typical diffuser configurations with F ranging from 29 to 83, s_r/dF ranging from 0.52 to 14.5, and u_rF ranging from 1 to 3.9. The model rosette in the experiments had four ports that were placed 90° apart around a circle. The diameters of the circle and the ports were 2.25 and 0.216 cm, respectively. The nozzles were inclined by 60°. A total of 41 tests were conducted, with 8 of them for the 0° scenario, namely the angle between the ambient current direction and one of the jets was 0°. The remaining tests were for the 45° scenario, namely the rosette in the 0° scenario was rotated by 45°. The experimental results showed that the 45° configuration was preferable in terms of dilution, so this work focuses on the 45° scenario. The mixing properties of a rosette dense jet group inflowing currents were assumed to be functions of s_r/dF and u_rF, so s_r/dF and u_rF were used as input parameters in the MGGP modeling for each mixing characteristic parameter. Each model was trained and tested using 33 data sets (Fig. 2). The modeling tests in this study showed that these data sets were sufficient for the establishment of reasonable MGGP models. It is acknowledged that these models can be further improved or extended in the future when more measurements become available. In fact, with the rapid development of data collection techniques, using AI techniques to develop and improve a model is becoming a mainstream trend. However, it is necessary to demonstrate the performance of an AI technique to provide information for appropriate date collection.

2.3. MGGP modeling

MGGP is a recent variant of GP that allows for multiple genes. An MGGP model consists of several traditional GP trees and these trees are linearly combined. These trees are usually considered to be genes. A typical MGGP chromosome is shown in Fig. 3, which represents the expression

$$y = \alpha\left[0.7\exp(x_1^{0.3}) + \log(x_2)\right] + \beta\left\{2.1\left|x_1^{x_2}\right| - \log\left[\sin\left(x_2\right)\right]\right\} + \gamma \tag{9}$$

where x_1 and x_2 are the input variables, α and β are the weights of the genes, and γ is the bias term. This model predicts the output (y) using two input variables (x_1 and x_2). There are some nonlinear terms in the trees, but the trees are linearly combined. The weights and bias are determined from the training data using ordinary least squares methods. Therefore, MGGP effectively combines the capability of the standard GP in describing nonlinear behavior and that of the ordinary least squares methods in estimating regression coefficients.

In MGGP, the first population is constructed by randomly creating chromosomes, and genetic operations such as crossover, mutation, and reproduction are performed to

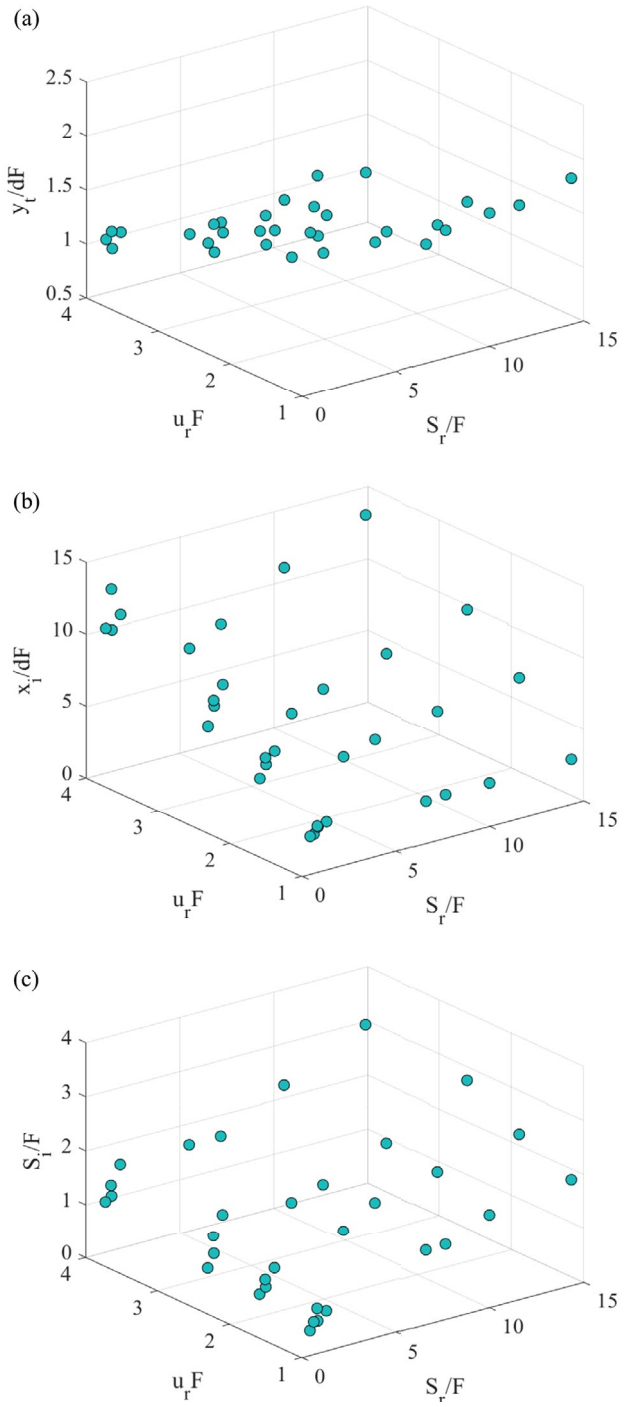

Fig. 2. 3D plots of data sets: (a) y_t/dF, (b) x_i/dF, and (c) S_i/F.

create new populations during the evolving process. The final MGGP model is a weighted linear combination of multiple trees that contain nonlinear terms. Because MGGP models allow for multiple genes, each gene only needs to contain a few tree depths (typically four or five). Therefore, the nonlinear terms are of a low order, and the evolved model can be relatively compact. The MGGP model has been widely found to be superior to other methods, such as the standard GP and conventional regression methods [18,19]. More details regarding the MGGP technique have been well documented elsewhere in the literature [17,18].

In this work, the MGGP modeling was performed using the open-source toolbox GPTIPS2 [17]. It is written in standard MATLAB and has a pluggable architecture, so it is readily modifiable and extendible. Typical parameter settings were employed in the present MGGP modeling. The population size was set to be 500. The selection tournament size was set to be 20 and 30% of tournaments to be Pareto tournaments. The maximum tree depth was set to be four, based on a sensitivity study. The program performs two runs simultaneously, and each run terminates after 60 s. The elite fraction number was set to be 0.3; namely, 30% of the models in one generation were copied to the next generation. The functional set included most of the commonly seen mathematical operations: {'times', 'minus', 'plus', 'rdivide', 'square', 'sin', 'cos', 'exp', 'mult3', 'add3', 'sqrt', 'cube', 'power', 'negexp', 'neg', 'abs', 'log'}.

3. Results and discussion

3.1. MGGP training and sensitivity analysis

The experimental data were divided into two groups: the data in 80% of the cases were used for training, and those in the remaining 20% were used for testing, which is a typical ratio of data splitting [28,29]. The data splitting was conducted in a random manner using MATLAB's random-permutation function. The model accuracy was evaluated primarily based on the root-mean-square error (RMSE) and *R*-squared (R^2) values, which can be expressed as

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(x_{\mathrm{calc}} - x_{\mathrm{obs}}\right)^2} \tag{10}$$

$$R^2 = \left\{ \frac{N\sum_{i=1}^{N}\left(x_{\mathrm{obs}}x_{\mathrm{calc}}\right) - \sum_{i=1}^{N}\left(x_{\mathrm{obs}}\right)\sum_{i=1}^{N}\left(x_{\mathrm{calc}}\right)}{\sqrt{\left[\sum_{i=1}^{N}\left(x_{\mathrm{obs}}^2\right) - \left(\sum_{i=1}^{N}x_{\mathrm{obs}}\right)^2\right]\left[\sum_{i=1}^{N}\left(x_{\mathrm{calc}}^2\right) - \left(\sum_{i=1}^{N}x_{\mathrm{calc}}\right)^2\right]}} \right\}^2 \tag{11}$$

where x_{obs} and x_{calc} denote the observed and calculated values, respectively; N is the number of data points.

Preparatory modeling was conducted to evaluate the influence of the input variables and maximum genes on a chromosome. MGGP models based on a single input variable were first tried. For example, the MGGP model for y_t/dF based on s_r/dF with a maximum number of genes of 4 gave the best result with RMSE = 0.30 and R^2 = 0.52, and that based on u_rF gave the best result with RMSE = 0.21 and R^2 = 0.77. The high error and low correlation values indicated that a single input variable cannot adequately describe the mixing characteristic parameters, and implied that the two input variables both had a clear influence on the model performance. Therefore, both input variables were utilized in subsequent analysis.

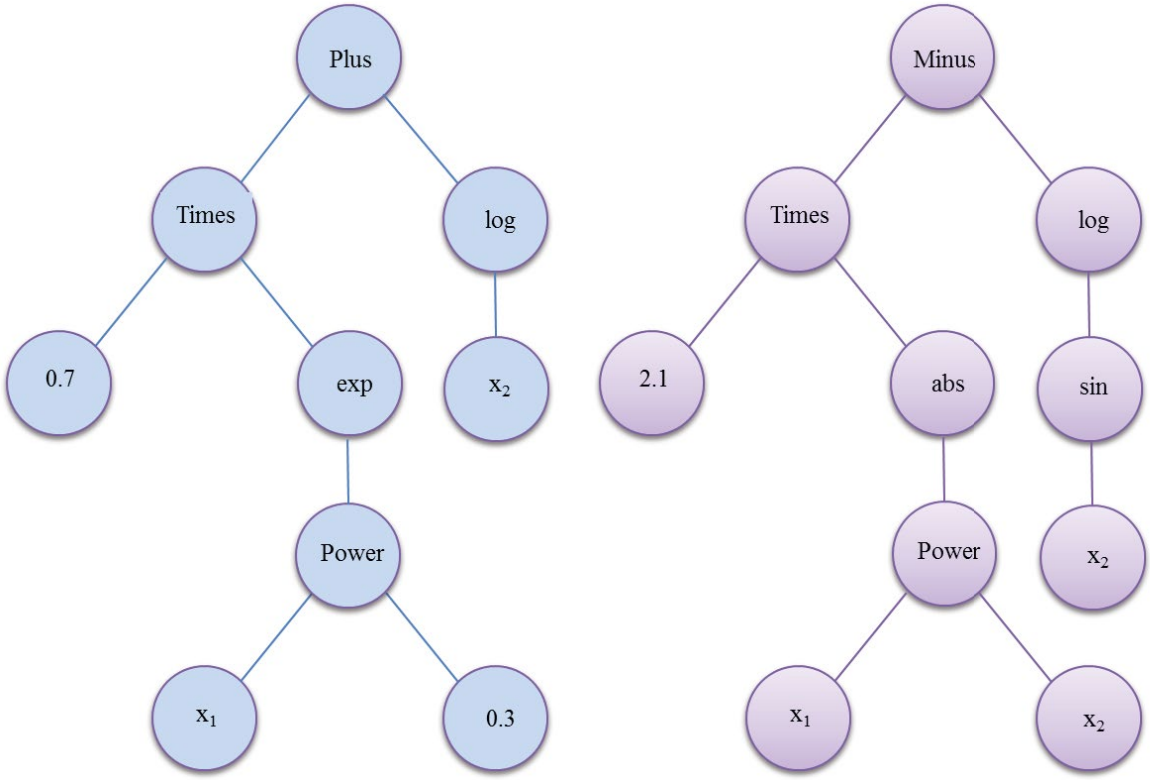

Fig. 3. Example of a multi-gene genetic programming chromosome.

The sensitivity analysis for maximum genes in a chromosome was conducted starting from training models with only one gene in a chromosome. Then, the gene number was successively increased, and the model performance was evaluated using both the training and testing data sets. For example, the MGGP model for y_t/dF with only one gene in a chromosome gave the best result with RMSE = 0.13 and R^2 = 0.91 for the training data set. However, the same model performed poorly in predicting the testing (unseen) data (RMSE = 0.22 and R^2 = 0.55), implying that the model was under-trained. Finally, a model with a maximum of four genes in a chromosome performed very well in both the training and testing periods. To control model complexity and reduce the risk of overfitting, a higher gene number was not used and the value of maximum genes in a chromosome was set to be four. The relevant sensitivity analysis results for all the mixing characteristic parameters are presented in Table 1. The results demonstrated that the models with a maximum gene number of four had a satisfactory predictive capacity (R^2 > 0.90) for all the mixing characteristic parameters.

Table 1
Sensitivity analysis results for the value of maximum genes in a chromosome

Variable	Data set	Indicator	Max Gene 3	Max Gene 4
y_t/dF	Training	RMSE	0.10	0.06
		R^2	0.95	0.98
	Testing	RMSE	0.11	0.09
		R^2	0.90	0.97
x_t/dF	Training	RMSE	0.56	0.45
		R^2	0.97	0.98
	Testing	RMSE	0.63	0.54
		R^2	0.97	0.98
S_t/F	Training	RMSE	0.10	0.12
		R^2	0.97	0.96
	Testing	RMSE	0.27	0.24
		R^2	0.89	0.91

3.2. MGGP models

A major objective of this study is to obtain explicit expressions for y_t/dF, x_t/dF, and S_t/dF as functions of $u_r F$ and s_r/dF. The MGGP program provided many models of different levels of model performance and complexity, and the best model herein was defined as the model that exhibited the best balance during the training period. The coefficients in the evolved models were originally expressed in the form of the ratio of two integers, which ensured a high accuracy. For the purpose of presenting these models, the variable-precision floating-point arithmetic function in MATLAB was used to reduce the significant digits. The developed models with fewer significant digits for the three mixing characteristic parameters are summarized in Table 2.

It can be observed that these models are relatively compact, and thus can be easily utilized. To evaluate the robustness of the developed models, various performance indices for both the training and testing data sets were calculated

Table 2
Summary of the developed MGGP-based equations

Variable	Equation
y_t/dF	$(0.039(u_rF)^{0.5})/(s_r/dF) - 0.134\sin(61.2(s_r/dF)) - ((4.73 \times 10^{12}(s_r/dF) + 4.73 \times 10^{12}(u_rF) + 3.74 \times 10^{13}))/(3.52 \times 10^{13}(s_r/dF) + 7.04 \times 10^{13}(u_rF)) - (305(s_r/dF)^{1.5})/((s_r/dF) + (u_rF) + 7.9)^3 - 0.34/(s_r/dF) - 2.19(u_rF)^{0.5} - 0.0323(s_r/dF)^{0.25} \times (\sin((s_r/dF)) - (u_rF)^2) + 5.47$
x_t/dF	$6.39(s_r/dF)(u_rF) - 1.93\cos((u_rF)) - 0.537\cos((s_r/dF)^2(u_rF)^3) - 6.2(s_r/dF)(u_rF)\exp(-\exp(-(s_r/dF))) + 2.11$
S_t/F	$0.0797(s_r/dF) + 0.0797\cos(\cos((u_rF))) + 0.446\log((s_r/dF) + (u_rF) + \cos((u_rF))) - 0.392\cos((u_rF)) + (0.355\sin(\exp(-(s_r/dF))))/\cos((s_r/dF)) - 0.042$

Note: Numerical precision reduced for display purposes.

(Table 3), including the mean bias error (MBE), mean absolute error (MAE), mean absolute percentage error (MAPE), root-mean-squared error (RMSE), normalized root-mean-squared error (NRMSE), R^2, and p-value of the analysis of variance [30,31]. The p-value for each parameter was greater than the significance level of 0.05, so the predictions were not found to be statistically different from the experimental data.

The MGGP predictions for y_t/dF were very accurate. The MBE values were 0 and 0.04 for the training and testing data sets, respectively. This indicated that the bias was effectively diminished during the training period whereas the model slightly over-predicted the overall magnitude of the unseen data. The MAE and RMSE values for the training data sets were 0.04 and 0.06, respectively, so the absolute error in MGGP training for y_t/dF was about 0.05. The MAE and RMSE values for the testing data sets were slightly higher with an average value of approximately 0.08. These error indicators demonstrated that the absolute error of the MGGP model predictions for y_t/dF was well below 0.1. The MAPE and NRMSE values indicated that the error in MGGP prediction for y_t/dF was about 3% and 5% for the training and testing data sets, respectively, which was satisfactorily accurate. The R^2 values were higher than 0.97, which demonstrated that the model can accurately predict the overall data trend. The performance indices for the unseen data set were very close to those for the training data set, demonstrating a high predictive capacity of the proposed model for y_t/dF.

The performance of the MGGP model for x_t/dF was also very good. Similar to the model for y_t/dF, the MBE value was also 0 for the training data set, indicating that the overall bias in the MGGP training was negligible. The MBE value for the testing data set was –0.29, which suggested that the model tended to under-estimate x_t/dF. The MAE and RMSE values for the training data sets were 0.36 and 0.45, respectively, so the absolute error in MGGP training for x_t/dF was about 0.4. The MAE and RMSE values for the testing data sets were 0.34 and 0.54, respectively, so the absolute error in MGGP predictions for x_t/dF in the testing period was slightly higher than 0.4. The MAPE and NRMSE values indicated that the error in MGGP prediction for x_t/dF was about 6% and 8% for the training and testing data sets. The R^2 values were both 0.98, which were satisfactorily high. Similar to the MGGP models for y_t/dF, the developed models for x_t/dF had consistently good performances during both the training and testing periods, demonstrating the high generalization capacity of these models.

For S_t/F, the developed model also exhibited an unnoticeable bias for the training data set and slightly underpredicted the overall testing data. The MAE and RMSE values for the training data sets were 0.10 and 0.12, respectively, so the absolute error in MGGP training for S_t/F was of the magnitude of 0.1. The MAE and RMSE values for the testing data sets were 0.21 and 0.24, respectively, so the absolute error in MGGP predictions for S_t/F in the testing period was of the magnitude of 0.2. The MAPE and NRMSE values indicated that the error in MGGP prediction for S_t/F was about 8% and 18% for the training and testing data sets, respectively. The R^2 values were 0.96 for the training and 0.91 for the testing data sets. These indices showed that the predictions for the testing data set had higher errors and lower

Table 3
Performance indices of the developed models

Measure of fit	y_t/dF		x_t/dF		S_t/F	
	Training	Testing	Training	Testing	Training	Testing
MBE	0.00	0.04	0.00	–0.29	0.00	–0.02
MAE	0.04	0.07	0.36	0.34	0.10	0.21
MAPE (%)	2.80	4.82	6.53	6.30	8.19	18.92
RMSE	0.06	0.09	0.45	0.54	0.12	0.24
NRMSE	0.04	0.06	0.06	0.09	0.08	0.17
R^2	0.98	0.97	0.98	0.98	0.96	0.91
p-value	1.00	0.86	1.00	0.88	1.00	0.97

correlation values than those for the training data set. A better outcome can be expected with the availability of more observational data. However, the overall predictions of the developed model for S_i/F were quite reasonable.

3.3. Model comparison

The fitting performances of the developed MGGP models are compared with the empirical equations proposed by [3] using the entire data set. The comparisons are presented in Fig. 4. The scattered points are very close to the identity line, demonstrating that the proposed MGGP models predicted y_t/dF, x_i/dF, and S_i/F very well. The empirical predictions were also accurate in general but deviated farther from the identity line at some points.

To compare the model performances quantitatively, the performance indices of the developed models and the existing empirical equations for the entire data sets were calculated and are presented in Table 4.

The p-value for each parameter was greater than the significance level of 0.05, so the predictions were not found to be statistically different from the experimental data. For y_t/dF, the proposed MGGP model reduced the absolute errors (MAE and RMSE) of the empirical equations by more than 50% and increased R^2 by about 10%. The MAPE and NRMSE values decreased from approximately 10% to 5% when the MGGP model was utilized. For x_i/dF, the proposed MGGP model reduced the absolute errors of the empirical equations by more than 60% and increased the R^2 by about 20%. The MAPE and NRMSE values decreased from approximately 17% to 7% when the MGGP model was utilized. For S_i/F, the developed MGGP model was not as accurate as those for y_t/dF and x_i/dF but it also performed better than the existing empirical equations. It reduced the absolute errors of the empirical equations by more than 30% and increased the R^2 by about 10%. The MAPE and NRMSE values decreased from approximately 15% to 10% when the MGGP model was employed. These results show that the proposed models have better generalization capacity than the existing empirical equations.

The MGGP algorithm does not require any pre-determination of the model structure, but it requires a modeler to define the input and output variables. In this study, the input and output variables for the MGGP models were determined by the well-established dimensional analysis presented in Section 2.1. The empirical equations had the same input and output variables because they were also derived based on the dimensional analysis. The final forms and coefficients of the chromosomes were obtained by the MGGP algorithm through an evolutionary process. The MGGP models were believed to be more accurate than the empirical equations because the results were closer to the experimental data, as the MGGP algorithm can further optimize the coefficients, detect some hidden relationships, and avoid improper predetermination of the model structures. It is acknowledged that there are many other advanced AI models that have the potential to predict the detailed mixing characteristics. However, the present study focused on MGGP primarily because it can provide explicit and compact models, and thus the other AI algorithms are not tested in the present study.

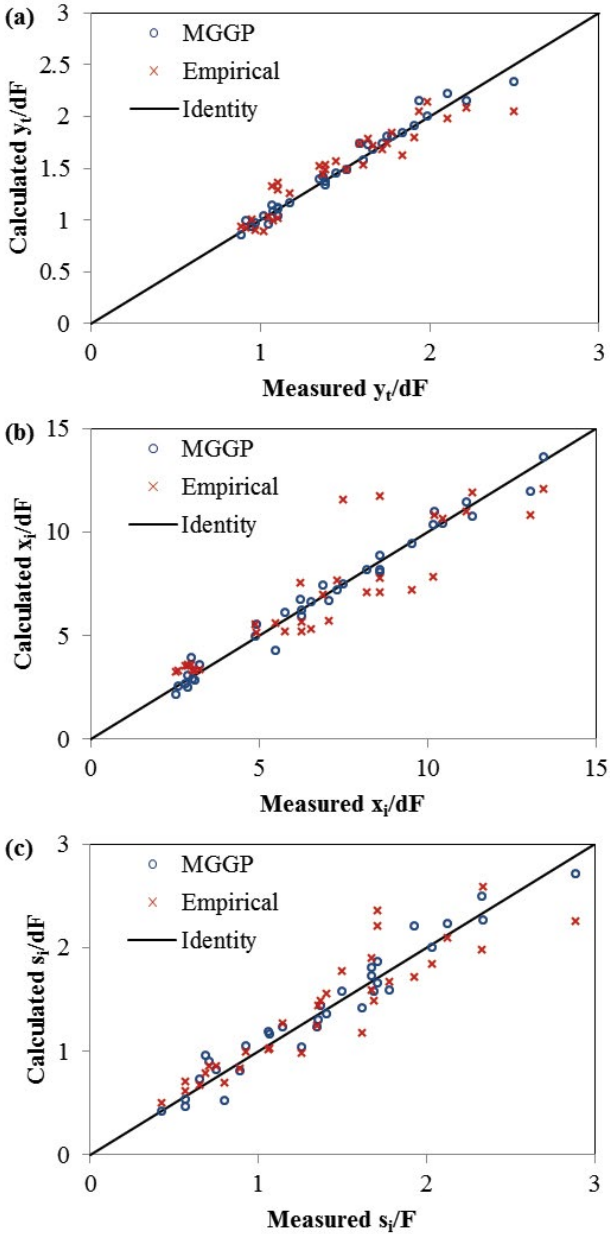

Fig. 4. Comparison of measured and calculated results: (a) y_t/dF, (b) x_i/dF, and (c) S_i/F.

3.4. Prediction confidence analysis

To assess further the capacity for generalization of the developed MGGP models, prediction confidence analyses were conducted with the assistance of MATLAB's "nlpredci" function. The function is a nonlinear prediction confidence interval function that uses a symmetric confidence interval approach. It uses the equations of [32] based on the t distribution and gives a symmetric confidence interval at every point. More information about this method is referred to in previous papers in the literature [32–34]. The results are summarized in Table 5, which shows that the mean uncertainty interval half widths were ±0.38, ±2.09, and ±0.39 for y_t/dF, x_i/dF, and S_i/F, respectively.

Table 4
Performance indices of the developed models and the existing empirical equations

Measure of fit	y_t/dF		x_t/dF		S_t/F	
	Empirical	MGGP	Empirical	MGGP	Empirical	MGGP
MBS	0.01	0.01	–0.06	–0.06	–0.01	0.00
MAE	0.12	0.05	0.98	0.36	0.19	0.13
MAPE (%)	8.00	3.18	15.09	6.48	12.86	10.46
RMSE	0.15	0.07	1.35	0.47	0.25	0.15
NRMSE	0.10	0.05	0.20	0.07	0.17	0.10
R^2	0.88	0.97	0.82	0.98	0.85	0.95
p-value	0.90	0.94	0.94	0.94	0.96	0.98

Table 5
Prediction confidence analysis of the developed MGGP models

Parameters	y_t/dF	x_t/dF	S_t/F
Mean uncertainty interval half width	±0.38	±2.09	±0.39
Mean 95% prediction error interval	(–0.37 0.38)	(–2.15 2.02)	(–0.39 0.38)
Max uncertainty interval half width	±0.41	±2.40	±0.78
Max 95% prediction error interval	(–0.40 0.42)	(–2.25 2.54)	(–0.90 0.67)

4. Conclusions

A new method based on MGGP to predict a rosette dense jet group in crossflow ambient conditions is introduced in this study. Experimental data were utilized to train and test the models to obtain y_t/dF, x_t/dF, and S_t/F as functions of s_t/dF and u_rF. The developed MGGP models were found to be capable of satisfactorily estimating y_t/dF (MAPE = 3.18%; RMSE = 0.07; R^2 = 0.97), x_t/dF (MAPE = 6.48%; RMSE = 0.47; R^2 = 0.98), and S_t/F (MAPE = 10.46%; RMSE = 0.15; R^2 = 0.95). The predictions of the developed MGGP models were clearly better than the existing empirical equations, which were not sufficiently accurate to be used in conceptual rosette diffuser design. Prediction confidence analyses were performed, and the mean uncertainty interval half widths were ±0.38, ±2.09, and ±0.39 for y_t/dF, x_t/dF, and S_t/F, respectively. Therefore, the developed models and the MGGP algorithm have been found to be useful in predicting a rosette dense jet group in crossflow ambient conditions. Model improvements can be expected with the availability of more training and testing data, so further research objectives should mainly involve the experimental and numerical studies aimed to enrich the data sets.

Acknowledgements

This work was funded by the Natural Sciences and Engineering Research Council of Canada (NSERC Discovery Grants). The first author was a recipient of a scholarship from the China Scholarship Council (CSC).

References

[1] A.C. Lai, D. Yu, J.H. Lee, Mixing of a rosette jet group in a crossflow, J. Hydraul. Eng., 137 (2011) 787–803.
[2] O. Abessi, P.J. Roberts, V. Gandhi, Rosette diffusers for dense effluents, J. Hydraul. Eng., 143 (2016) 06016029.
[3] O. Abessi, P.J. Roberts, Rosette diffusers for dense effluents inflowing currents, J. Hydraul. Eng., 144 (2017) 06017024.
[4] G.C. Christodoulou, I.G. Papakonstantis, I.K. Nikiforakis, Desalination brine disposal by means of negatively buoyant jets, Desal. Water Treat., 53 (2015) 3208–3213.
[5] N. Ahmad, T. Suzuki, Study of dilution, height, and lateral spread of vertical dense jets in marine shallow water, Water Sci. Technol., 73 (2016) 2986–2997.
[6] J.O.G. Pecly, Estimation of the dilution field near a marine outfall by using effluent turbidity as an environmental tracer and comparison with dye tracer data, Water Sci. Technol., 77 (2018) 269–277.
[7] S.J. Kwon, I.W. Seo, Experimental investigation of wastewater discharges from a Rosette-type riser using PIV, KSCE J. Civ. Eng., 9 (2005) 355–362.
[8] X. Tian, P.J. Roberts, Experiments on marine wastewater diffusers with multiport rosettes, J. Hydraul. Eng., 137 (2011) 1148–1159.
[9] A. Dashti, M. Asghari, H. Solymani, M. Rezakazemi, A. Akbari, Modeling of $CaCl_2$ removal by positively charged polysulfone-based nanofiltration membrane using artificial neural network and genetic programming, Desal. Water Treat., 111 (2018) 57–67.
[10] A.A. Tashvigh, B. Nasernejad, Soft computing method for modeling and optimization of air and water gap membrane distillation–a genetic programming approach, Desal. Water Treat., 76 (2017) 30–39.
[11] R. Hashim, C. Roy, S. Shamshirband, S. Motamedi, A. Fitri, D. Petković, K.I. Song, Estimation of wind-driven coastal waves near a Mangrove forest using adaptive neuro-fuzzy inference system, Water Resour. Manage., 30 (2016) 2391–2404.
[12] Y. Peng, X. Zhang, W. Xu, Y. Shi, Z. Zhang, An optimal algorithm for cascaded reservoir operation by combining the grey forecasting model with DDDP, Water Sci. Technol. Water Supply, 18 (2018) 142–150.
[13] A. Picos, J.M. Peralta-Hernández, Genetic algorithm and artificial neural network model for prediction of discoloration dye from an electro-oxidation process in a press-type reactor, Water Sci. Technol., 78 (2018) 925–935.

[14] X. Xia, S. Jiang, N. Zhou, X. Li, L. Wang, Genetic algorithm hyper-parameter optimization using Taguchi design for groundwater pollution source identification, Water Sci. Technol. Water Supply, 19 (2019) 137–146.
[15] A.A. Tashvigh, F.Z. Ashtiani, M. Karimi, A. Okhovat, A novel approach for estimation of solvent activity in polymer solutions using genetic programming, Calphad, 51 (2015) 35–41.
[16] A.A. Tashvigh, F.Z. Ashtiani, A. Fouladitajar, Genetic programming for modeling and optimization of gas sparging assisted microfiltration of an oil-in-water emulsion, Desal. Water Treat., 57 (2016) 19160–19170.
[17] D.P. Searson, A.H. Gandomi, Handbook of Genetic Programming Applications, Springer, Cham, 2015, pp. 551–573.
[18] M.J.S. Safari, A.D. Mehr, Multi-gene genetic programming for sediment transport modeling in sewers for conditions of non-deposition with a bed deposit, Int. J. Sediment Res., 33 (2018) 262–270.
[19] D.S. Pandey, I. Pan, S. Das, J.J. Leahy, W. Kwapinski, Multi-gene genetic programming based predictive models for municipal solid waste gasification in a fluidized bed gasifier, Bioresour. Technol., 179 (2015) 524–533.
[20] X. Yan, A. Mohammadian, Numerical modeling of vertical buoyant jets subjected to lateral confinement, J. Hydraul. Eng., 143 (2017) 04017016.
[21] X. Yan, A. Mohammadian, Numerical modeling of multiple inclined dense jets discharged from moderately spaced ports, Water, 11 (2019) 1–15.
[22] X. Yan, A. Mohammadian, X. Chen, Three-dimensional numerical simulations of buoyant jets discharged from a rosette-type multiport diffuser, J. Mar. Sci. Eng., 7 (2019) 409.
[23] X. Yan, A. Mohammadian, Multigene genetic-programming-based models for initial dilution of laterally confined vertical buoyant jets, J. Mar. Sci. Eng., 7 (2019) 246.
[24] X. Yan, A. Mohammadian, Evolutionary modeling of inclined dense jets discharged from multiport diffusers, J. Coastal Res., 36 (2019) 362–371.
[25] X. Yan, A. Mohammadian, Evolutionary prediction of multiple vertical buoyant jets in stationary ambient water, Desal. Water Treat., 178 (2020) 41–52.
[26] S. Zhang, B. Jiang, A.W.K. Law, B. Zhao, Large-eddy simulations of 45 inclined dense jets, Environ. Fluid Mech., 16 (2016) 101–121.
[27] N. Ahmad, R.E. Baddour, Density effects on dilution and height of vertical fountains, J. Hydraul. Eng., 141 (2015) 04015024.
[28] A. Guven, M. Gunal, Prediction of scour downstream of grade-control structures using neural networks, J. Hydraul. Eng., 134 (2008) 1656–1660.
[29] H. Bashiri, E. Sharifi, V.P. Singh, Prediction of local scour depth downstream of sluice gates using harmony search algorithm and artificial neural networks, J. Irrig. Drain. Eng., 144 (2018) 06018002.
[30] C.J. Willmott, K. Matsuura, Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance, Clim. Res., 30 (2005) 79–82.
[31] M.V. Shcherbakov, A. Brebels, N.L. Shcherbakova, A.P. Tyukov, T.A. Janovsky, V.A.E. Kamaev, A survey of forecast error measures, World Appl. Sci. J., 24 (2013) 171–176.
[32] G.A.F. Seber, C.J. Wild, Nonlinear Regression, John Wiley & Sons, New York, 1989.
[33] T.P. Lane, W.H. DuMouchel, Simultaneous confidence intervals in multiple regression, Am. Stat., 48 (1994) 315–321.
[34] K.D. Dolan, L. Yang, C.P. Trampel, Nonlinear regression technique to estimate kinetic parameters and confidence intervals in unsteady-state conduction-heated foods, J. Food Eng., 80 (2007) 581–593.