

Article

# Simulations of the Concentration Fields of Rosette-Type Multiport Buoyant Discharges Using Combined CFD and Multigene Genetic Programming Techniques

Xiaohui Yan <sup>1,\*</sup> , Yan Wang <sup>1</sup>, Abdolmajid Mohammadian <sup>2</sup>  and Jianwei Liu <sup>1</sup>

<sup>1</sup> School of Water Resources Engineering, Dalian University of Technology, Dalian 116024, China; wy15665890918@mail.dlut.edu.cn (Y.W.); jwliu@dlut.edu.cn (J.L.)

<sup>2</sup> Department of Civil Engineering, University of Ottawa, Ottawa, ON K1N 6N5, Canada; majid.mohammadian@uottawa.ca

\* Correspondence: yanxh@dlut.edu.cn

**Abstract:** Rosette-type diffusers are becoming popular nowadays for discharging wastewater effluents. Effluents are known as buoyant jets if they have a lower density than the receiving water, and they are often used for municipal and desalination purposes. These buoyant effluents discharged from rosette-type diffusers are known as rosette-type multiport buoyant discharges. Investigating the mixing properties of these effluents is important for environmental impact assessment and optimal design of the diffusers. Due to the complex mixing and interacting processes, most of the traditional simple methods for studying free single jets become invalid for rosette-type multiport buoyant discharges. Three-dimensional computational fluid dynamics (3D CFD) techniques can satisfactorily model the concentration fields of rosette-type multiport buoyant discharges, but these techniques are typically computationally expensive. In this study, a new technique of simulating rosette-type multiport buoyant discharges using combined 3D CFD and multigene genetic programming (MGGP) techniques is developed. Modeling the concentration fields of rosette-type multiport buoyant discharges using the proposed approach has rarely been reported previously. A validated numerical model is used to carry out extensive simulations, and the generated dataset is used to train and test MGGP-based models. The study demonstrates that the proposed method can provide reasonable predictions and can significantly improve the prediction efficiency.

**Keywords:** numerical modeling; computational fluid dynamics; multigene genetic programming; rosette-type diffusers; buoyant discharges



**Citation:** Yan, X.; Wang, Y.; Mohammadian, A.; Liu, J. Simulations of the Concentration Fields of Rosette-Type Multiport Buoyant Discharges Using Combined CFD and Multigene Genetic Programming Techniques. *J. Mar. Sci. Eng.* **2021**, *9*, 1311. <https://doi.org/10.3390/jmse9111311>

Academic Editor:  
Alessandro Antonini

Received: 22 October 2021  
Accepted: 17 November 2021  
Published: 22 November 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Estimating the flow and mixing processes of effluent discharges is crucial for reliable environmental impact assessments, sound design of treatment and outfall systems, and proper disposal of wastewater discharges [1–3]. Effluents are known as buoyant jets if they have a lower density than the receiving water, and they are often used for municipal and desalination purposes [4–6]. A key factor influencing the mixing processes of effluents is the diffuser type. In recent decades, rosette-type multiport diffusers have become popular due to their merits of being cheaper and less space-demanding than traditional diffusers (e.g., single-port diffusers) [7–9]. However, because of the complex mechanisms, the practice of modeling the mixing processes of rosette-type multiport diffusers has not yet been well established and requires further investigation.

In recent years, complementing the experimental studies [8–12], the simulation of the mixing processes of jets and plumes using fully physics-based computational fluid dynamics models has become popular because of the improvements in computing capabilities [13–18]. One shortcoming of the CFD approach is that it typically requires heavy computing resources and long simulation time, and it is thus helpful to propose a new technique that can make predictions faster with lower requirements for computing resources.

Complementing the CFD approach, machine learning techniques have been demonstrated capable of analyzing water-related phenomenon [4,7,19]. A key advantage of using these machine learning techniques compared to traditional regression-based methods is that it does not require a predefined model structure, and is thus able to avoid the errors caused by the model structure assumptions and to detect hidden relationships between input and output variables. Thus, machine learning-based models are typically superior to empirical equations derived using traditional regression-based methods. Compared to CFD models, trained machine learning-based models have been found to be much more efficient for some problems [20–23]. However, the training process for a machine learning technique typically requires a large dataset, which often does not exist, limiting its widespread usage in water-related applications.

The MGGP technique has previously been applied to modeling other types of discharges, but the mixing mechanisms for different types of discharges are significantly different, and thus the modeling of rosette-type multiport discharges requires further investigation. To the best of the authors’ knowledge, this is the first time that a combined 3D CFD and MGGP technique is proposed to develop models for rosette-type multiport buoyant discharges. It should be emphasized that most previous studies on the applications of machine learning techniques to wastewater jets only modeled some characteristics parameters, and this is believed to be the first time that entire concentration fields are modeled using a machine learning technique.

This study aims to develop a new technique of simulating rosette-type multiport buoyant discharges using combined three-dimensional computational fluid dynamics and multigene genetic programming techniques. The mixing processes of a buoyant jet discharged from a rosette-type multiport diffuser is first simulated using a CFD model. The CFD model is then validated against experimental data and employed to carry out additional computations to enrich the dataset. The extended dataset is then utilized to train a machine learning model that can predict the normalized concentration at different locations. In this study, the CFD model is established within the framework of OpenFOAM (Open Field Operation and Manipulation), and the machine learning models are developed using the multigene genetic programming (MGGP) approach [24,25]. A key outcome of the work is a well-trained machine learning-based model. The accuracy of the developed machine learning-based model is comparable to the CFD model for the studied applications, but it can provide predictions in seconds while a CFD model needs hours or even days to complete a simulation. Most of the previous CFD studies focused on the mixing processes of jets aimed at developing and evaluating the performance of different models, whereas the present CFD simulations are used to enrich the dataset.

## 2. Materials and Methods

### 2.1. The 3D CFD Model

The governing equations for the CFD model are the Reynolds-averaged Navier–Stokes equations for mixing two fluids, which can be written as [26,27]

$$\nabla \cdot \mathbf{U} = 0 \tag{1}$$

$$\frac{\partial \rho \mathbf{U}}{\partial t} + \nabla \cdot (\rho \mathbf{U} \mathbf{U}) = -\nabla \cdot (p_{rgh}) - gh \nabla \rho + \nabla \cdot (\rho \mathbf{T}) \tag{2}$$

in which

$$\rho = \alpha_1 \rho_1 + \alpha_2 \rho_2 = \alpha_1 \rho_1 + (1 - \alpha_1) \rho_2 \tag{3}$$

$$\mathbf{T} = -\frac{2}{3} \bar{\mu}_{eff} \nabla \cdot \mathbf{U} \mathbf{I} + \bar{\mu}_{eff} \nabla \mathbf{U} + \bar{\mu}_{eff} (\nabla \mathbf{U})^T \tag{4}$$

$$\bar{\mu}_{eff} = \alpha_1 (\mu_{eff})_1 + \alpha_2 (\mu_{eff})_2 \tag{5}$$

$$(\mu_{eff})_i = (\mu - \mu_t)_i \tag{6}$$

where  $t$ ,  $U$  and  $\rho$  represent time, velocity, and density, respectively. The variables  $\rho_{rgh}$  and  $h$  represent the static pressure minus hydraulic pressure and the height of the fluid column, respectively. The variable  $\alpha$ ,  $\mu$  and  $\mu_t$  are the volume fraction of the fluids, the dynamic viscosity, and turbulent viscosity, respectively. The symbol  $T$  is the viscous stresses tensor. The subscript  $i$  indicates either the jet or the ambient fluid. For example,  $\rho_1$  represents the density of the jet, and  $\rho_2$  represents the density of the ambient fluid.

The volume fraction of the fluids,  $\alpha$ , in mixing problems can be calculated using a transport equation [5], which can be expressed as

$$\frac{\partial \alpha_1}{\partial t} + \nabla \cdot (\mathbf{U}\alpha_1) = \nabla \cdot \left( \left( D_{ab} + \frac{v_t}{S_C} \right) \nabla \alpha_1 \right) \tag{7}$$

where  $D_{ab}$ ,  $v_t$ , and  $s_c$  represent the molecular diffusivity, the turbulent viscosity, and the turbulent Schmidt number, respectively.

Regarding turbulence modeling, the previous study reported by Yan et al. [6] demonstrated that the RNG (renormalization group)  $k-\epsilon$  turbulence model outperformed the standard  $k-\epsilon$  model without significantly increasing the computational costs, and thus the RNG  $k-\epsilon$  turbulence model was used in this study. Compared with the other three different turbulence closures, including the standard  $k-\epsilon$ , standard  $k-\omega$ , and  $k-\omega$  shear stress transport (SST) models, the RNG  $k-\epsilon$  model has been found by Yan et al. [5] to be the most accurate for multiple buoyant jets. Therefore, in this research, the RNG  $k-\epsilon$  model was adopted. The standard values of the model coefficients were used in this study, and they are summarized in Table 1.

**Table 1.** The values of the model constants.

| $\sigma_k$ | $\sigma_\epsilon$ | $c_{1\epsilon}$ | $c_{2\epsilon}$ | $c_\mu$ | $\eta_0$ | $\beta$ |
|------------|-------------------|-----------------|-----------------|---------|----------|---------|
| 0.71942    | 0.71942           | 1.42            | 1.68            | 0.0845  | 4.38     | 0.012   |

### 2.2. The MGGP Technique

Genetic programming is a machine learning technique that can use an evolutionary process to relate input and output variables. Different from some other “black-box” machine learning approaches, genetic programming techniques can provide explicit models, and thus the developed models can be readily retyped and used in other programs. Genetic programming has already been successfully employed in water-related applications, such as the prediction of riprap stone size under overtopping flow and the prediction of local scour depth downstream of sluice gates [28,29]. The MGGP technique is a new advancement of the genetic programming technique and was used for the present cases. The major difference between traditional genetic programming and MGGP techniques is the number of genes in a chromosome; there is only one gene in a traditional genetic programming chromosome, whereas there can be multiple genes in an MGGP chromosome. The terms in a gene are often nonlinear, but the genes in an MGGP chromosome are linearly combined. The MGGP technique generates an initial set of models at the beginning of the evolutionary process and then improves the models by processes such as mutation, reproduction, and crossover. The technique finally terminates the evolutionary process when either the number of generations or the computational time reaches a pre-defined value. More information and the advantages of the MGGP technique can be found in previous papers [4,7,19].

### 2.3. The Combined CFD–MGGP Method

The first step of the proposed modeling approach was to establish a well-validated CFD model. This part of the work has already been completed by the authors in a previous study [6], in which a CFD model was developed in OpenFOAM and validated against experimental data.

The validated CFD model is then used to perform additional simulations to enrich the dataset. One of the most important parameters influencing the mixing properties of wastewater effluents is the densimetric Froude number, which is a function of the densities of the fluids, the port diameter, and the initial velocity. In this study, 20 different additional cases were considered, which have different values of ambient water density. As the CFD model is physics-based, it is deemed reasonable to assume that the accuracy of the numerical predictions is acceptable. The enriched dataset is then employed to develop machine learning-based models using the MGGP technique. The dataset has four variables: the Fr number (the ratio of inertial force to viscous force), the x (horizontal) and y (vertical) coordinates, and the normalized concentration. The Fr number, x, and y coordinates were defined as the features (input variables), while the normalized concentration was selected as the output variable.

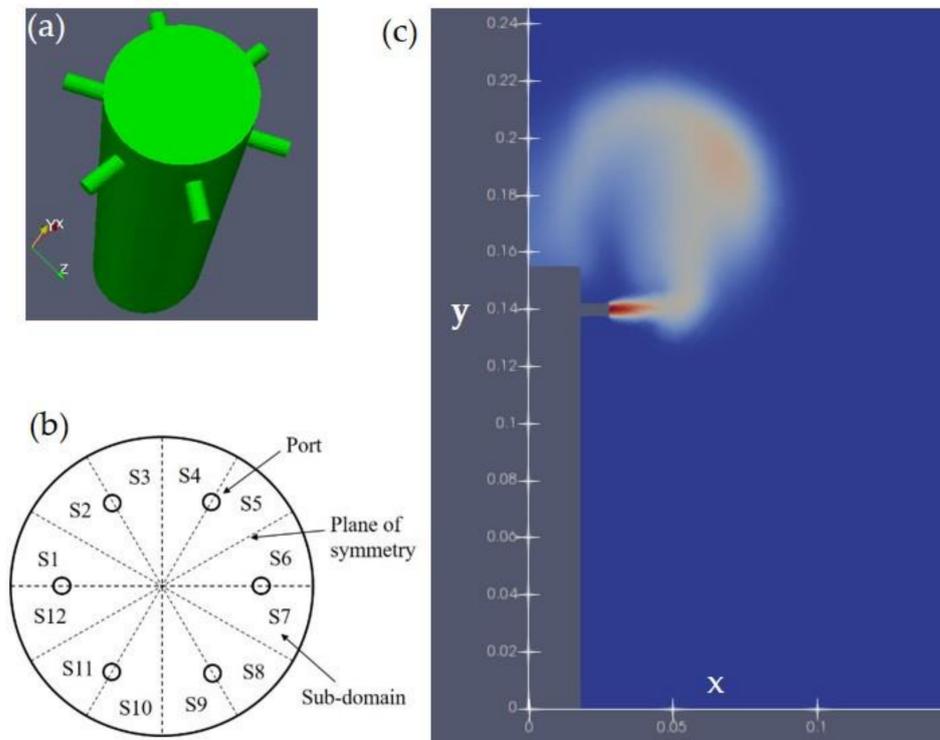
The length of the data matrix was (4040, 4). The dataset was randomly divided into two parts: 80% of the data were used for model training, and the remaining 20% of the data were used for model testing. In order to develop MGGP-based models for the concentration field of a rosette buoyant jet, the Fr number, the x and y coordinates were defined as input variables, and the normalized concentration was used as output variable. The MGGP training process was performed using GPTIPS2 (Genetic Programming Toolbox for Multigene Symbolic Regression 2) [25], which is an open-source MATLAB (MathWorks, Natick, USA) code. Sensitivity analyses were performed to determine the configurations of the training process. The number of generations was set at 1000, and each generation had a population size of 500. The tournament size, probability of the Pareto tournament, and elite fraction were set at 10, 0.3, and 0.3, respectively. The maximum number of genes was set at 20, and each gene had a maximum depth of trees of 20.

### 3. Results

#### 3.1. CFD Results

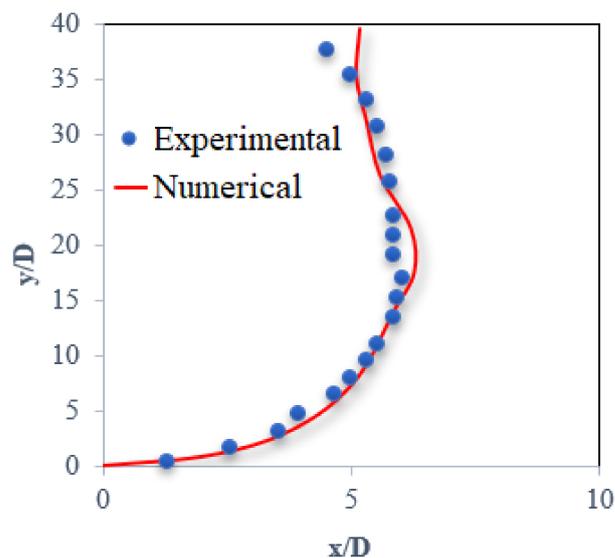
The governing equations were solved using a solver named “twoLiquidMixingFoam” on the open-source CFD platform OpenFOAM. In a pioneering study, a CFD model for rosette multiple buoyant jets has been validated by Yan et al. [6]. The computational domain and boundary conditions were primarily set according to the available experimental configurations [11]. The diffuser consisted of six ports, and the diameter of the ports was 0.0044 m. The study area was symmetric, and it was thus divided into 12 sub-domains, as illustrated in Figure 1. The model only simulated 1 of the 12 sub-domains and used the “symmetric” boundary condition available in OpenFOAM. The boundary conditions at the top surface and the outer boundaries were set to inlet–outlet, which is similar to the zero-gradient open boundary condition, but it will switch to the fixed-value boundary condition when there is any backward flow. The no-slip boundary condition with the standard wall function was employed for the bottom patch. An area of 0.06 m × 0.06 m near the jet port was selected as the area of interest because initial dilution of wastewater effluents was of the most significant importance. In the figures, the x coordinates ranged from 0.03 m to 0.09 m, and the y coordinates ranged from 0.13 m to 0.19 m. The results can be mirrored to obtain the results for the entire study domain using the software ParaView (Sandia National Laboratory, New Mexico, USA) and Tecplot (Tecplot Inc, Bellevue, USA). This approach ignores the swirling effects near the planes of the symmetry, but sensitivity studies have proved that these effects were minor, and simulating only one sub-domain can significantly reduce the computational costs. The initial values of k and  $\epsilon$  were calculated based on the jet diameter, initial velocity, and initial density, and their values were 0.000246 m<sup>2</sup>/s<sup>2</sup> and 0.002061 m<sup>2</sup>/s<sup>2</sup>, respectively. The initial velocity and initial fraction of the fluid for discharges were both set at zero. The computational mesh was generated using the open-source software Salome. An unstructured computation mesh with hexahedral cells with local refinements near the port was employed for domain discretization. Using the method reported by Yan and Mohammadian [2,13], mesh sensitivity analyses were performed. A relatively coarse grid resolution was firstly utilized, and then finer grids were tried until

convergence criterion was met. In the final mesh, the smallest grid size was 0.001 m, and the largest one was 0.005 m.



**Figure 1.** Schematic of a rosette-type diffuser (a), the sub-domains (b), and the coordinate setup (c).

The data of rosette-type buoyant jets obtained by experiment [11] and simulation were shown in Figure 2. In this figure, the comparison of the measured and simulated trajectories showed that the simulated data were in good agreement with the measured data. The root-mean-squared error (RMSE) for the normalized concentration was 0.25, the coefficient of determination (R2) was 0.98, and the normalized root-mean-squared error (NRMSE) was 0.04. Therefore, the performance of the CDF model is believed to be satisfactory, and the simulated data were assumed to be actual data.



**Figure 2.** Comparison of the experimental and simulated results. The dots = the experimental data [1]; the line = the numerical predictions.

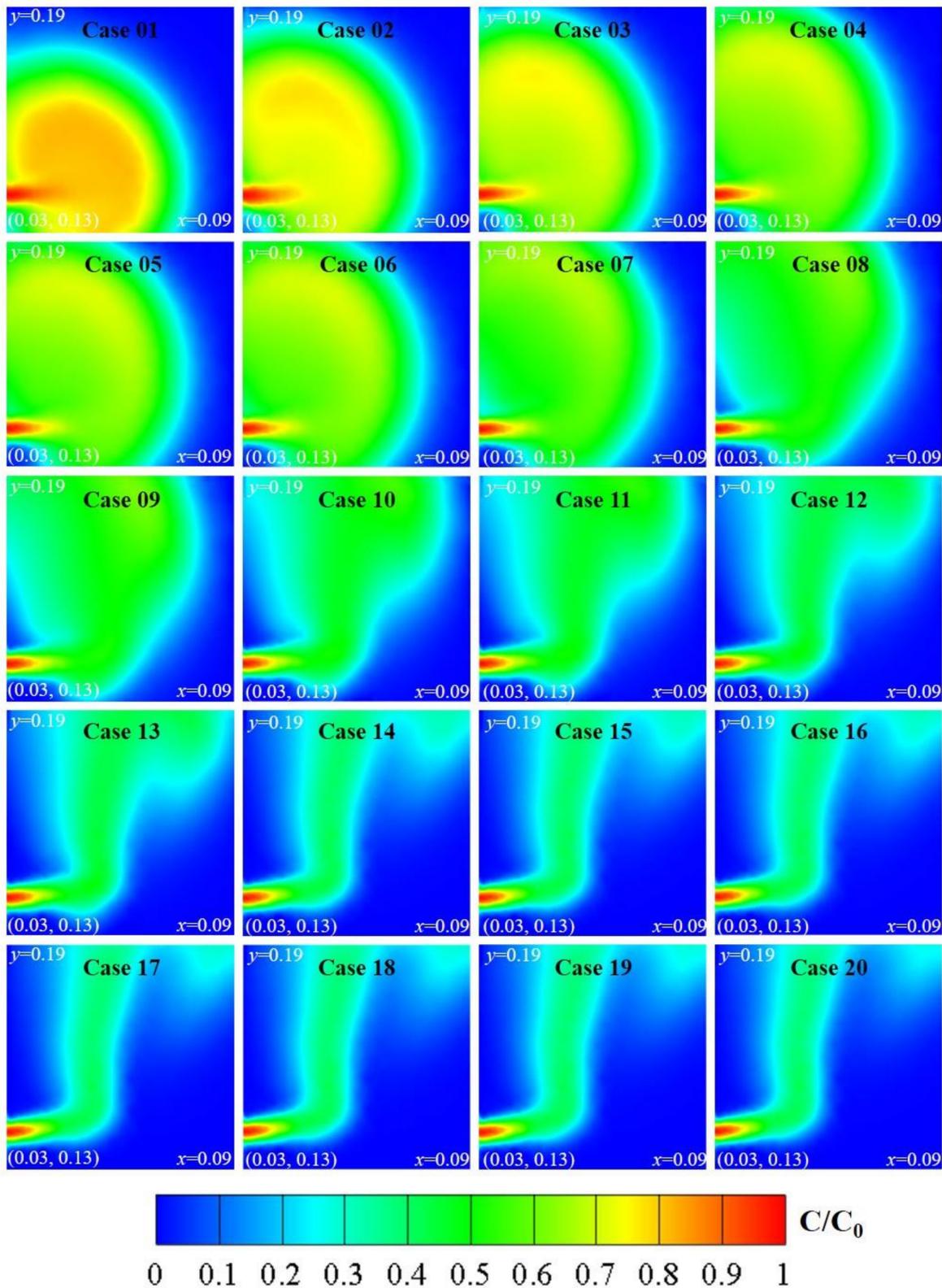
Yan et al. [6] have simulated rosette buoyant jets with different jet angles and evaluated the influences of the jet angles. Another major variable affecting the mixing properties of buoyant jets is the densimetric Froude number (Fr). In this study, the validated CFD model for rosette multiple buoyant jets was employed to perform simulations for 20 additional cases with different values of Fr. The values of ambient water density were within the reasonable ranges and determined using the “rand” function in MATLAB, and they led to Fr numbers ranging from 5.750 to 24.676. These values were randomly determined to reduce the regularity of the dataset. The intervals between the values of Fr number in different cases were small, so the current number of cases was deemed sufficient for training a machine learning-based model. A default value of 1 second was used for the computational time step, but the model automatically adjusted it based on the Courant number, which was of the value of 1. All the simulations were run up to 30 seconds, which was determined based on a sensitivity analysis; namely, longer computational duration did not significantly change the results. The detailed parameters are summarized in Table 2.

**Table 2.** Parameters for the additional computations.

| Cases   | $\rho_j$<br>(kg/m <sup>3</sup> ) | $\rho_a$<br>(kg/m <sup>3</sup> ) | $\Delta\rho$<br>(kg/m <sup>3</sup> ) | $g'$<br>(m/s <sup>2</sup> ) | D<br>(m) | U<br>(m/s) | Fr<br>(—) |
|---------|----------------------------------|----------------------------------|--------------------------------------|-----------------------------|----------|------------|-----------|
| Case 01 | 997                              | 998.3                            | 1.3                                  | 0.013                       | 0.0044   | 0.185      | 24.676    |
| Case 02 | 997                              | 999.5                            | 2.5                                  | 0.025                       | 0.0044   | 0.185      | 17.805    |
| Case 03 | 997                              | 1000.2                           | 3.2                                  | 0.031                       | 0.0044   | 0.185      | 15.743    |
| Case 04 | 997                              | 1001.1                           | 4.1                                  | 0.040                       | 0.0044   | 0.185      | 13.914    |
| Case 05 | 997                              | 1001.3                           | 4.3                                  | 0.042                       | 0.0044   | 0.185      | 13.588    |
| Case 06 | 997                              | 1001.4                           | 4.4                                  | 0.043                       | 0.0044   | 0.185      | 13.433    |
| Case 07 | 997                              | 1002.5                           | 5.5                                  | 0.054                       | 0.0044   | 0.185      | 12.022    |
| Case 08 | 997                              | 1004.8                           | 7.8                                  | 0.076                       | 0.0044   | 0.185      | 10.107    |
| Case 09 | 997                              | 1005.9                           | 8.9                                  | 0.087                       | 0.0044   | 0.185      | 9.467     |
| Case 10 | 997                              | 1009.0                           | 12.0                                 | 0.117                       | 0.0044   | 0.185      | 8.165     |
| Case 11 | 997                              | 1009.9                           | 12.9                                 | 0.125                       | 0.0044   | 0.185      | 7.879     |
| Case 12 | 997                              | 1013.5                           | 16.5                                 | 0.160                       | 0.0044   | 0.185      | 6.979     |
| Case 13 | 997                              | 1014.4                           | 17.4                                 | 0.168                       | 0.0044   | 0.185      | 6.799     |
| Case 14 | 997                              | 1020.2                           | 23.2                                 | 0.223                       | 0.0044   | 0.185      | 5.905     |
| Case 15 | 997                              | 1021.0                           | 24.0                                 | 0.231                       | 0.0044   | 0.185      | 5.808     |
| Case 16 | 997                              | 1021.5                           | 24.5                                 | 0.235                       | 0.0044   | 0.185      | 5.750     |
| Case 17 | 997                              | 1023.1                           | 26.1                                 | 0.250                       | 0.0044   | 0.185      | 5.575     |
| Case 18 | 997                              | 1023.9                           | 26.9                                 | 0.258                       | 0.0044   | 0.185      | 5.494     |
| Case 19 | 997                              | 1024.3                           | 27.3                                 | 0.261                       | 0.0044   | 0.185      | 5.454     |
| Case 20 | 997                              | 1025.9                           | 28.9                                 | 0.276                       | 0.0044   | 0.185      | 5.305     |

Note:  $\rho_j$  = the initial density of the effluent;  $\rho_a$  = the initial density of the ambient water;  $\Delta\rho = \rho_a - \rho_j$ ;  $g'$  = the reduced gravity; D = the port diameter of the diffuser; U = the initial velocity of the discharge; Fr = densimetric Froude number.

The normalized concentrations (defined by the concentration, C, divided by the initial jet concentration, C0) at the central plane (indicated by the dashed line in Figure 1b) for different cases obtained by the CFD model are presented in Figure 3. As can be seen, the differences in the mixing properties for different scenarios were very obvious, demonstrating that the influences of Fr are significant. In a non-buoyant case, the jets symmetrically spread in the vertical direction. In Case 01, the density of the effluent was slightly smaller than the ambient water, so the jet was bent slightly upward due to the buoyancy effect. With smaller values of Fr, the outlines of the buoyant jets became more obvious. In the cases with a smaller value of Fr, the density differences between the effluents and ambient water were greater, thus the effects of buoyancy were greater. Therefore, in Cases 02~08, the jets were less significantly bent upward, and buoyant jets can be clearly seen in Cases 09~20. The numerical data were then extracted using ParaView and MATLAB, which can be used for training and testing the MGPP-based models.



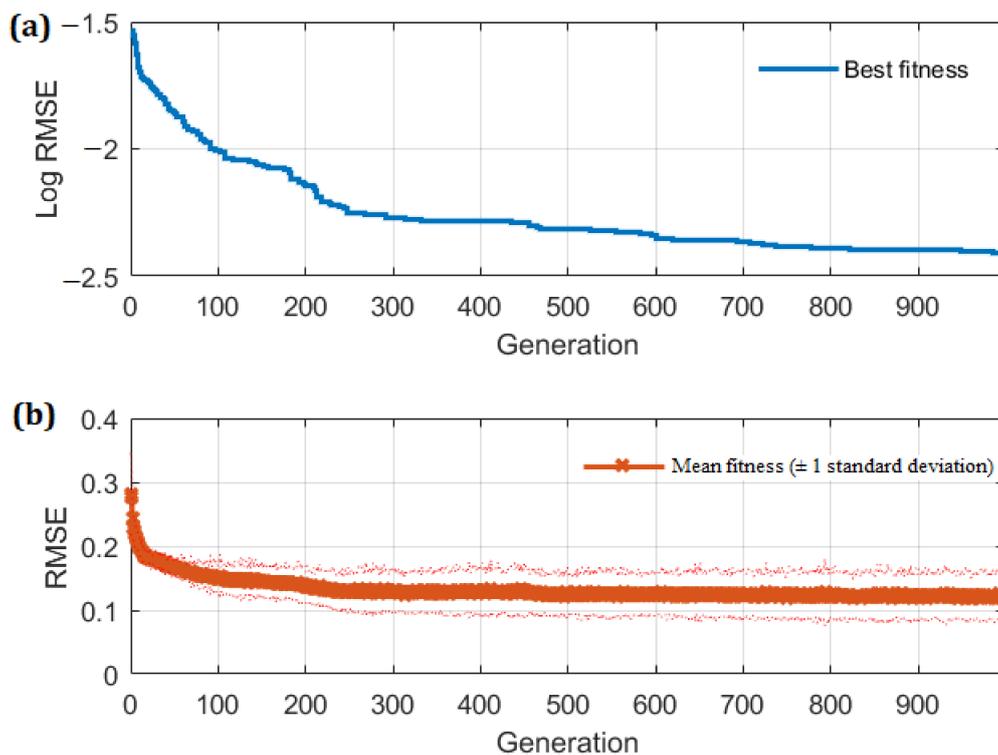
**Figure 3.** The normalized concentration field at the central plane for different cases obtained by the CFD model.

### 3.2. Results Obtained by the Combined Method

In the dataset, the mean, minimum, maximum, and median values for Fr were 9.993, 5.305, 24.676, and 8.022, respectively. The standard deviation, variance, kurtosis, and

skewness were 5.079, 25.801, 4.109, and 1.267, respectively. The mean, minimum, maximum, and median values for normalized concentration were 0.33, 0, 1, and 0.28, respectively. The standard deviation, variance, kurtosis, and skewness of normalized concentration were 0.28, 0.08, 2.09, and 0.49, respectively. This dataset was then used to train and test MGGP-based models. A total of 3232 data were assigned into the training dataset, and the remaining data were assigned into the testing dataset. The mean values for Fr and normalized concentration in the training dataset were 9.935 and 0.33, respectively; those in the testing dataset were 10.225 and 0.33, respectively.

Each generation in the evolutionary process generated 500 models, and the performances of these models varied significantly. The fitness of the MGGP models in each evolutionary step of the model training process was plotted in Figure 4. The fitness of the models was quantified by the RMSE values. In the earlier generations, the performance of the models was relatively poor. The mean RMSE value was approximately 0.3. The overall quality of the models significantly improved with the processes of mutation, crossover, and reproduction. At the generation of about 300, the mean fitness went below 0.15. As can be seen, the model errors did not change much after about 300 generations, implying that the performance cannot be substantially improved by running further steps. To be conservative, the training processes were run up to 1000 generations.



**Figure 4.** The fitness of the MGGP models in each evolutionary step: (a) the best fitness, and (b) the mean fitness and standard deviations.

The final generation contained 500 models, and these models had different levels of accuracy and simplicity. The present study used the Pareto-optimal approach to figure out the best model. The input variables (Fr, x, and y) were then supplied, and the best model was employed to predict the normalized concentration for each case. The normalized concentration at different locations for different cases obtained by the best model was utilized to re-construct the normalized concentration field at the central plane, which is shown in Figure 5. The concentration fields predicted by the combined approach did not perform very well in predicting the core of the jets in the first few cases. However, in general, the predictions matched the physics-based data very well, especially for the

influence area of the jets, which is one of the most important factors in environmental impact assessments. The jets were bent slightly upward for Cases 01~08. The outlines of the buoyant jets became more obvious in the cases with smaller values of Fr. These general observations were consistent with those obtained by the physics-based model.

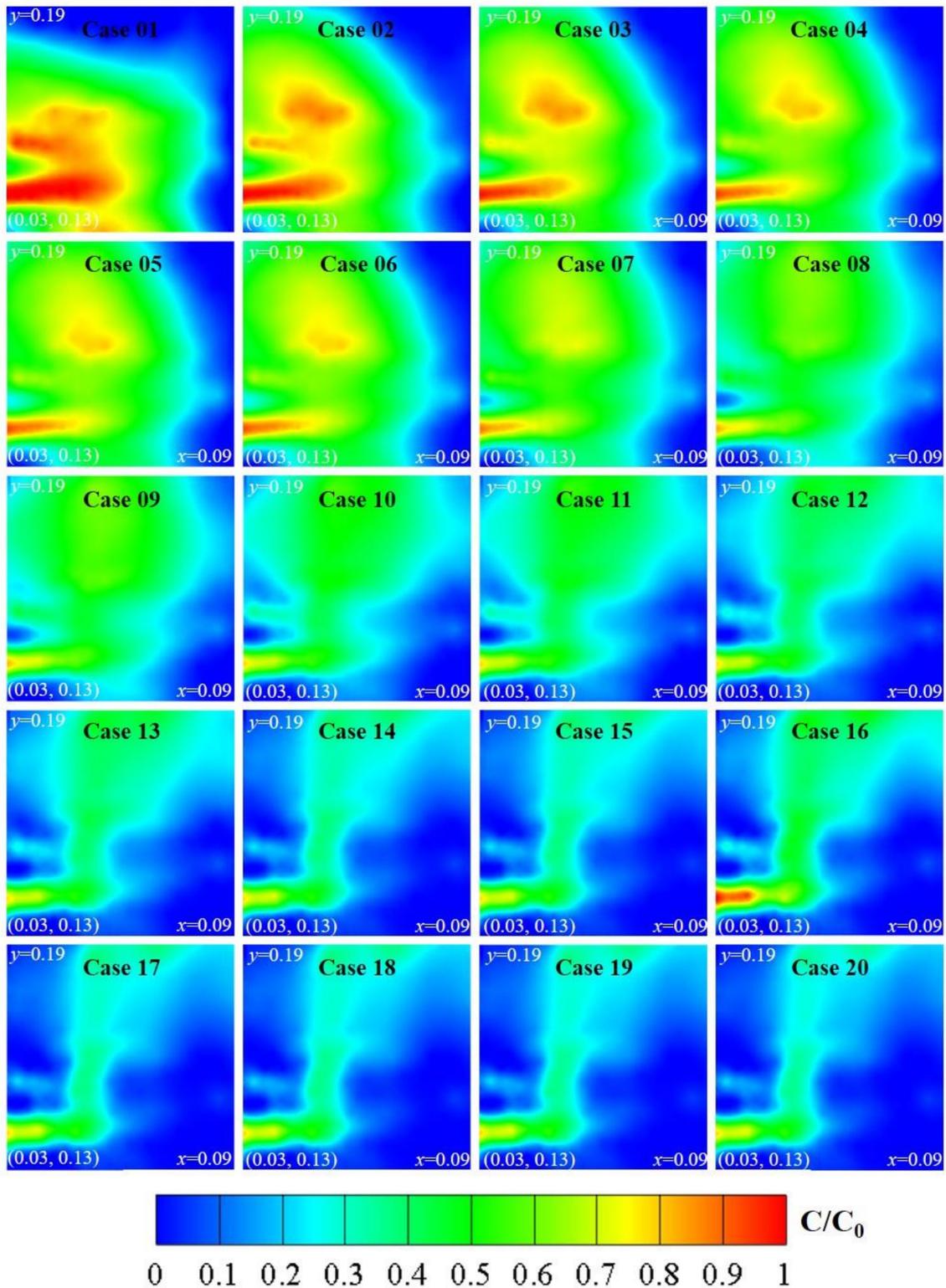
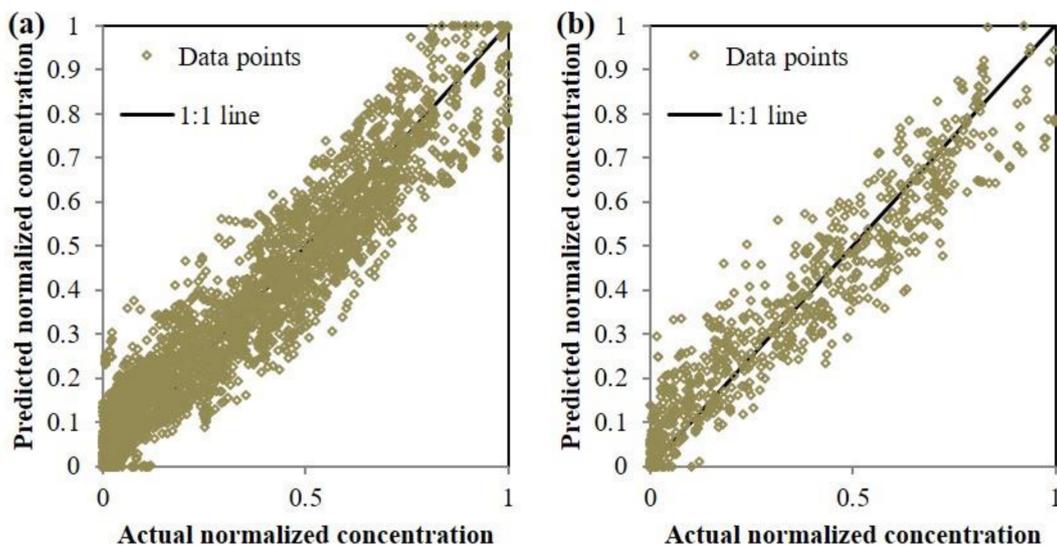


Figure 5. The normalized concentration field at the central plane for different cases obtained by the MGGP model.

The spatial distribution of the normalized concentration was also close to that provided by the physics-based model, but the results were less smooth, and there were a few regions with unexpected high values of normalized concentration. This phenomenon revealed a weakness of data-driven models: the predictions may not follow strict physical principles. However, it is reasonable to expect that the model can be continuously improved with more datasets with higher resolution and larger size. Nevertheless, the performance of the model obtained in this study was regarded as satisfactory.

The predictions for the training and testing datasets were further compared in Figure 6, and RMSE and R2 values were also calculated. In this figure, the results obtained by the physics-based model were referred to as the actual data, and those obtained by the combined model were called predicted data. The 1:1 line was also plotted in the figure. If a data point was located higher than the 1:1 line, the combined model over-estimated the data, and vice versa. As can be seen in the figure, most of the data points were located close to the 1:1 line, demonstrating the overall good performance of the combined model. The spatial distribution of the data points was relatively uniform, indicating that the model did not tend to over- or under-predict the data. To further quantify the performance of the prediction, the RMSE and R2 values for the training and datasets were calculated. RMSE and R2 values for the training dataset were 0.087 and 0.903, respectively. Those for the testing dataset were 0.088 and 0.902, respectively. The RMSE values were low and the R2 values were high, so the accuracy of the model predictions was acceptable. The fitness for the testing dataset was close to the training dataset, indicating that the risk of over-fitting was low.



**Figure 6.** Comparison of the actual and predicted normalized concentration: (a) training dataset, and (b) testing dataset.

#### 4. Discussion

The primary contribution of this study is the proposed hybrid method of modeling rosette-type multipoint buoyant discharges using combined 3D CFD and MGGP techniques. A key factor hindering the wide usage of 3D CFD techniques is the heavy computational costs. Although it is acceptable to validate a model and perform several computations, it is impractical to use a 3D CFD model to perform extensive calculations to check the influences of some parameters. Using the proposed approach, a new model was developed using the CFD data. Using the original CFD model to simulate a new case would take hours or even days, but using the developed MGGP-based model to perform a calculation would only take a few seconds. Although interpolation techniques may also be capable of providing satisfactory predictions for the current cases, the MGGP approach is believed to be more promising because (1) it can produce predictions by simply using the equations, unlike

interpolation techniques, which always require known values; and (2) the developed model can be readily extended to incorporate other factors. The present study only considered the influence of the Fr number because it primarily aimed to evaluate the proposed approach, but the model can be continuously extended in future studies to include more parameters.

Another factor limiting the application of 3D CFD models in practical water-related projects is the high requirements of disk storage. The CFD outputs include the volume fraction, the turbulent kinetic energy, turbulent energy dissipation rate, eddy viscosity, pressure, and velocity at each time step. However, in this approach, due to the fast calculation velocity, it is unnecessary to save the predicted data in the simulation progress. The size of the MATLAB script file for the final model was only 4 KB in the present study. Therefore, the proposed approach can substantially reduce the storage requirements of traditional CFD methods without significantly impacting the model accuracy.

It is acknowledged that artificial intelligence modeling typically requires a large amount of data for model training, and these methods are very sensitive to the number of input data. However, in the present study, only 20 additional simulations were performed, which provided a data matrix with a dimension of 4040 rows and 4 columns. However, the comparisons demonstrated that the limited number of data have provided a reasonable model in this study. The reasons can be summarized as follows: first, the combined effects of the input variables have been represented by the Fr number, and thus the number of features has been substantially reduced, making it much easier for MGGP to figure out the relationships between input and output variables; second, the typical range of the Fr number was narrow, so the intervals of Fr between different cases were actually sufficiently small; third, the values for different cases were determined randomly, and thus the regularity of the dataset has been reduced, which further improved the data quality. However, the predictions at some locations were notably scattered. Firstly, machine learning techniques make predictions from the data for model training instead of mechanisms, and thus it is quite common that the predictions do not follow strict physical principles. Secondly, no additional data were provided in this study, as the overall performance of the model was believed to be satisfactory, but the limited number of data has definitely lowered the prediction accuracy at some locations. Thus, a larger size of dataset would be required if a higher-performance model were desired.

## 5. Conclusions

The present study proposed a novel method of modeling rosette-type multiport buoyant discharges using combined 3D CFD and MGGP techniques. A total of 20 cases was simulated using a validated CFD model. Overall, the agreement between the predictions provided by the combined model and the physics-based data was acceptable. The jets were bent slightly upward for cases with larger values of Fr number. The outlines of the buoyant jets became more obvious in the cases with smaller values of Fr, which was consistent with the observations based on the physics-based results. The RMSE and R2 values for the training dataset were 0.087 and 0.903, respectively. Those for the testing dataset were 0.088 and 0.902, respectively. Therefore, although the predicted concentration fields exhibited some unrealistic values, the RMSE values were low and the R2 values were high, and thus the overall performance of the model was deemed satisfactory.

Using the original CFD model to simulate a new case would take hours or even days, and would require larger storage space. In contrast, the proposed approach only took a few seconds to perform calculations for a new case, and it used less memory. What is more, this is believed to be the first time that the entire concentration fields were modeled using a combined approach of CFD and machine learning techniques. This study demonstrated the capability of the proposed approach in replicating the concentration fields for rosette buoyant jets. The performance of the combined models can be further improved or extended when more data are available.

**Author Contributions:** Methodology, X.Y. and Y.W.; formal analysis, Y.W.; investigation, A.M.; writing—original draft preparation, Y.W.; writing—review and editing, A.M. and J.L.; visualization,

Y.W.; project administration, X.Y.; funding acquisition, X.Y. and A.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Fundamental Research Funds for the Central Universities (China; DUT20RC(3)096), and Natural Sciences and Engineering Research Council of Canada (NSERC Discovery Grants).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Acknowledgments:** The first author, Xiaohui Yan, is supported by the Fundamental Research Funds for the Central Universities (China; DUT20RC(3)096). The third author, Abdolmajid Mohammadian, was funded by the Natural Sciences and Engineering Research Council of Canada (NSERC Discovery Grants). We would like to thank the editors and the three reviewers for their careful reading of our manuscript and their insightful comments and suggestions.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Mohammadian, A.; Kheirkhah Gildeh, H.; Nistor, I. CFD modeling of effluent discharges: A review of past numerical studies. *Water* **2020**, *12*, 856. [\[CrossRef\]](#)
- Yan, X.; Mohammadian, A. Numerical modeling of vertical buoyant jets subjected to lateral confinement. *J. Hydraul. Eng.* **2016**, *43*, 04017016. [\[CrossRef\]](#)
- Knystautas, R. The turbulent jet from a series of holes in line. *Aeronaut. Q.* **2016**, *15*, 1–28. [\[CrossRef\]](#)
- Yan, X.; Mohammadian, A. Multigene genetic-programming-based models for initial dilution of laterally con-fined vertical buoyant jets. *J. Mar. Sci. Eng.* **2019**, *7*, 246. [\[CrossRef\]](#)
- Yan, X.; Ghodoosipour, B.; Mohammadian, A. Three-dimensional numerical study of multiple vertical buoyant jets in stationary ambient water. *J. Hydraul. Eng.* **2020**, *7*, 146. [\[CrossRef\]](#)
- Yan, X.; Mohammadian, A.; Chen, X. Three-dimensional numerical simulations of buoyant jets discharged from a rosette-type multiport diffuser. *J. Mar. Sci. Eng.* **2019**, *7*, 409. [\[CrossRef\]](#)
- Yan, X.; Mohammadian, A. Evolutionary prediction of the trajectory of a rosette momentum jet group in flowing currents. *J. Coast. Res.* **2020**, *36*, 1059–1067. [\[CrossRef\]](#)
- Abessi, O.; Roberts, P.J.W.; Gandhi, V. Rosette diffusers for dense effluents. *J. Hydraul. Eng.* **2016**, *143*, 06016029. [\[CrossRef\]](#)
- Abessi, O.; Roberts, P.J.W. Rosette diffusers for dense effluents in flowing currents. *J. Hydraul. Eng.* **2017**, *144*, 06017024. [\[CrossRef\]](#)
- Fan, W.; Bao, W.; Cai, Y.; Xiao, C.; Zhang, Z.; Pan, Y.; Chen, Y.; Liu, S. Experimental study on the effects of a vertical jet impinging on soft bottom sediments. *Sustainability* **2020**, *12*, 3775. [\[CrossRef\]](#)
- Lai, C.C.; Lee, J.H. Mixing of inclined dense jets in stationary ambient. *J. Hydro-Environ. Res.* **2012**, *6*, 9–28. [\[CrossRef\]](#)
- Lai, A.C.H.; Chan, S.N.; Law, A.W.K.; Adams, E.E. Spreading hypothesis of a particle plume. *J. Hydraul. Eng.* **2016**, *142*, 04016065. [\[CrossRef\]](#)
- Yan, X.; Mohammadian, A. Numerical modeling of multiple inclined dense jets discharged from moderately spaced ports. *Water* **2019**, *11*, 2077. [\[CrossRef\]](#)
- Yan, X.; Mohammadian, A.; Chen, X. Numerical modeling of inclined plane jets in a linearly stratified environment. *Alex. Eng. J.* **2020**, *59*, 1857–1867. [\[CrossRef\]](#)
- Kheirkhah Gildeh, H.; Mohammadian, A.; Nistor, I.; Qiblawey, H. Numerical modeling of 30° and 45° inclined dense turbulent jets in stationary ambient. *Environ. Fluid Mech.* **2015**, *15*, 537–562. [\[CrossRef\]](#)
- Zhang, S.; Law, A.W.K.; Jiang, M. Large eddy simulations of 45° and 60° inclined dense jets with bottom impact. *J. Hydro-Environ. Res.* **2017**, *15*, 54–66. [\[CrossRef\]](#)
- Lou, Y.; Jiang, H.; Han, X. Numerical simulation of two coalescing turbulent forced plumes in linearly stratified fluids. *Phys. Fluids* **2019**, *31*, 037111. [\[CrossRef\]](#)
- Glaze, D.J.; Frankel, S.H. Stochastic inlet conditions for large-eddy simulation of a fully turbulent jet. *AIAA J.* **2003**, *41*, 1064–1073. [\[CrossRef\]](#)
- Yan, X.; Mohammadian, A. Evolutionary modeling of inclined dense jets discharged from multiport diffusers. *J. Coast. Res.* **2019**, *36*, 362–371. [\[CrossRef\]](#)
- Mehr, A.D.; Nourani, V. A pareto-optimal moving average-multigene genetic programming model for rain-fall-runoff modelling. *Environ. Model. Softw.* **2017**, *92*, 239–251. [\[CrossRef\]](#)
- Bayazidi, A.M.; Wang, G.-G.; Bolandi, H.; Alavi, A.H.; Gandomi, A. Multigene genetic programming for estimation of elastic modulus of concrete. *Math. Probl. Eng.* **2014**, *2014*, 1–10. [\[CrossRef\]](#)
- De Paiva, G.M.; Pimentel, S.P.; Alvarenga, B.P.; Marra, E.; Mussetta, M.; Leva, S. Multiple site intraday solar irradiance forecasting by machine learning algorithms: MGGP and MLP neural networks. *Energies* **2020**, *13*, 3005. [\[CrossRef\]](#)

23. Safari, M.J.S.; Mehr, A.D. Multigene genetic programming for sediment transport modeling in sewers for conditions of non-deposition with a bed deposit. *Int. J. Sediment Res* **2018**, *33*, 262–270. [[CrossRef](#)]
24. Sakhaei, Z.; Nikooee, E.; Riazi, M. A new formulation for non-equilibrium capillarity effect using multi-gene genetic programming (MGGP): Accounting for fluid and porous media properties. *Eng. Comput.* **2020**. [[CrossRef](#)]
25. Searson, D.P. GPTIPS 2: An Open-Source Software Platform for Symbolic Data Mining. In *Handbook of Genetic Programming Applications*; Springer: Cham, Switzerland, 2015; pp. 551–573.
26. Holzmann, T. *Mathematics, Numerics, Derivations and OpenFOAM®*; Holzmann CFD: Loeben, Germany, 2016.
27. OpenFOAM Foundation. *OpenFOAM User Guide*; Version 4.0; The OpenCFD Foundation: London, UK, 2016.
28. Najafzadeh, M.; Tafarjnoruz, A.; Lim, S.Y. Prediction of local scour depth downstream of sluice gates using data-driven models. *ISH J. Hydraul. Eng.* **2017**, *23*, 1–8. [[CrossRef](#)]
29. Najafzadeh, M.; Rezaie Balf, M.; Tafarjnoruz, A. Prediction of riprap stone size under overtopping flow using data-driven models. *Int. J. River Basin Manag.* **2018**, *16*, 1–30. [[CrossRef](#)]