


## Article

# Scour Propagation Rates around Offshore Pipelines Exposed to Currents by Applying Data-Driven Models

Mohammad Najafzadeh <sup>1,\*</sup> and Giuseppe Oliveto <sup>2</sup> 

<sup>1</sup> Department of Water Engineering, Faculty of Civil and Surveying Engineering, Graduate University of Advanced Technology, Kerman 76315-117, Iran

<sup>2</sup> School of Engineering, University of Basilicata, Viale dell'Ateneo Lucano 10, I-85100 Potenza, Italy; giuseppe.oliveto@unibas.it

\* Correspondence: moha.najafzadeh@gmail.com or m.najafzadeh@kgut.ac.ir

**Abstract:** Offshore pipelines are occasionally exposed to scouring processes; detrimental impacts on their safety are inevitable. The process of scouring propagation around offshore pipelines is naturally complex and is mainly due to currents and/or waves. There is a considerable demand for the safe design of offshore pipelines exposed to scouring phenomena. Therefore, scouring propagation patterns must be focused on. In the present research, machine learning (ML) models are applied to achieve equations for the prediction of the scouring propagation rate around pipelines due to currents. The approaching flow Froude number, the ratio of embedment depth to pipeline diameter, the Shields parameter, and the current angle of attack to the pipeline were considered the main dimensionless factors from the reliable literature. ML models were developed based on various setting parameters and optimization strategies coming from evolutionary and classification contents. Moreover, the explicit equations yielded from ML models were used to demonstrate how the proposed approaches are in harmony with experimental observations. The performance of ML models was assessed utilizing statistical benchmarks. The results revealed that the equations given by ML models provided reliable and physically consistent predictions of scouring propagation rates regarding their comparison with scouring tests.

**Keywords:** offshore pipeline; currents; scour propagation rate; regression analysis; soft computing techniques



**Citation:** Najafzadeh, M.; Oliveto, G. Scour Propagation Rates around Offshore Pipelines Exposed to Currents by Applying Data-Driven Models. *Water* **2022**, *14*, 493. <https://doi.org/10.3390/w14030493>

Academic Editors: Michele Greco and Giovanni Martino

Received: 12 December 2021

Accepted: 4 February 2022

Published: 7 February 2022

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



**Copyright:** © 2022 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

Offshore pipeline structures are commonly employed to transport various fluids, such as oil, gas, oil-gas mixtures, and water. In offshore projects, there is a need to consider many engineering aspects, such as ecology, geo-hazards, and environmental loading [1]. Once a submarine pipeline is constructed, several physical factors needing close attention are introduced, such as the state of the seabed, seabed mobility, submarine landslides, waves, and currents [2–4]. The seabed has two various states: it is either relatively smooth or corrugated (or uneven) with high and low elevations. For an uneven seabed, the pipeline is subjected to free spans when connecting two high levels, leaving the section between these levels with unsupported points. When an unsupported section is extended, the bending stress exerted onto it may increase. From another point of view, vibrations that occur from current-related vortices must be taken into consideration. Seabed leveling and post-installation support are protective measures for unsupported pipeline spans [2–4]. A further risk to pipelines is the seabed mobility, with sand waves, in particular, being a major feature. Sand waves move with passing time. When a crest of the seabed bears a pipeline during the construction of the project, this area may find itself in a trough later during the operational lifespan of the offshore pipeline [2]. As the third physical factor, submarine landslides occur when the sediment transport rate is high. Submarine landslides take place

on steeper slopes. When the landslide process is intensified by displacement, offshore pipelines might be exposed to severe bending, with possible consequent tensile failure [3]. Other mechanical factors are related to currents and waves. High-intensity currents and waves have repercussions on the laying operations [4,5].

The above-mentioned physical factors require great attention during pipeline installation and operation because these factors can disturb the safety of the pipeline itself. Underestimating these key elements might lead to scouring processes. A large number of attempts have been made during the past three decades to understand the mechanical factors (e.g., free spans, seabed sediment mobility, geomorphology of the seabed, waves, and currents) of scouring beneath seabed pipelines [6–9]. Overall, the scouring process below pipelines has a three-dimensional flow structure in the field. Experimental observations have demonstrated that free spans are the primary cause of the scouring phenomenon. The length of free spans is controlled by various factors such as the flow conditions, soil conditions, sinking of the pipeline at the span shoulders, and sagging of the pipeline in the scour hole (e.g., [10–13]). Free spanning occurs in a suspended submarine pipeline segment where there is no joint point between the offshore pipelines and the surface of the seabed due to various causes, such as the scour of ocean currents, uneven surface of the seabed, human-made dangerous activities, and residual/thermal stresses [14].

The 3D scouring propagation around offshore pipelines exposed to currents has been limited to few studies—mainly through experimental investigations—during the last decade. Firstly, Cheng et al. [6] studied the three-dimensional scour rate of propagation along offshore pipelines after the scouring process was initiated. They found that some experimental variables (e.g., embedment depth of the offshore pipeline, live-bed motion state of seabed sediments due to currents, and flow incident angle) had meaningful impacts on the scouring propagation velocities along the pipeline. They proposed a regression-based formulation for the approximation of the scouring propagation rates. Later, Wu and Chiew [7,8] carried out experimental investigations under clear-water conditions to understand the three-dimensional mechanism of scouring propagation beneath pipelines of various diameters. They emphasized that the rate of scouring propagation was controlled by the Froude number, the Shields parameter, and the initial embedment depth of the offshore pipeline. Cheng et al. [9] studied the scour rate propagation along offshore pipelines for combined wave and current flow conditions. Finally, they presented an empirical equation with a permissible level of precision to estimate the longitudinal rate of scouring propagation.

From the above-mentioned investigations, it can be inferred that a deeper understanding of three-dimensional scouring rates would require further efforts to achieve more comprehensive empirical models with more accurate performance than the current formulas presented in the literature, which are generally obtained by regression analysis or traditional techniques. In the case of current-induced scour, conventional methods based on experimental observations at the laboratory scale have revealed that the main governing variables are the geometric properties of the pipeline, the motion state of the seabed, and the approaching flow state (e.g., [7–9]).

During recent years, various machine learning (ML) models have been utilized to predict three-dimensional scour rates at seabed pipelines with promising results. Firstly, Najafzadeh and Saberi-Movahed [15] improved the group method of data handling (GMDH) by gene-expression programming (GEP) to predict the scouring rate propagation below offshore pipelines under wave conditions. They found that GMDH-GEP provided more accurate results than nonlinear regression-based equations and artificial neural networks (ANNs). Similarly, Ehteram et al. [16] optimized the structure of an ANN with the cooling body algorithm (CBA). They concluded that the ANN-CBA model provided better performance for wave-induced scouring rates than simple ANNs. In a recent study, Najafzadeh and Oliveto [17] presented new empirical equations based on four robust ML models (i.e., multivariate adaptive regression splines (MARS), evolutionary polynomial regression (EPR), model tree (MT), and gene-expression programming (GEP)) for the pre-

diction of 3D scouring rates below pipelines exposed to regular waves. ML techniques, especially those used in the latest research by Najafzadeh and Oliveto [17], rely on their inherent advantages (e.g., reduction of complexity, fast learning process, and preparation of physical patterns of observational variables) to provide effective performance.

The literature review reveals that there is a lack of using ML techniques to provide physically consistent equations for the prediction of the scouring propagation rates beneath offshore pipelines subjected to currents. Therefore, the outline of this study is articulated in the following way: (i) experimental variables for feeding ML models are provided; (ii) ML models are implemented while controlling the setting parameters of each ML model; (iii) the ML models drive empirical equations; (iv) statistical analyses is performed to evaluate the empirical equations given by ML models; and (v) the physical consistency of ML models' performance with experimental observations is controlled. A general overview of the organization of this study is outlined in Figure 1.

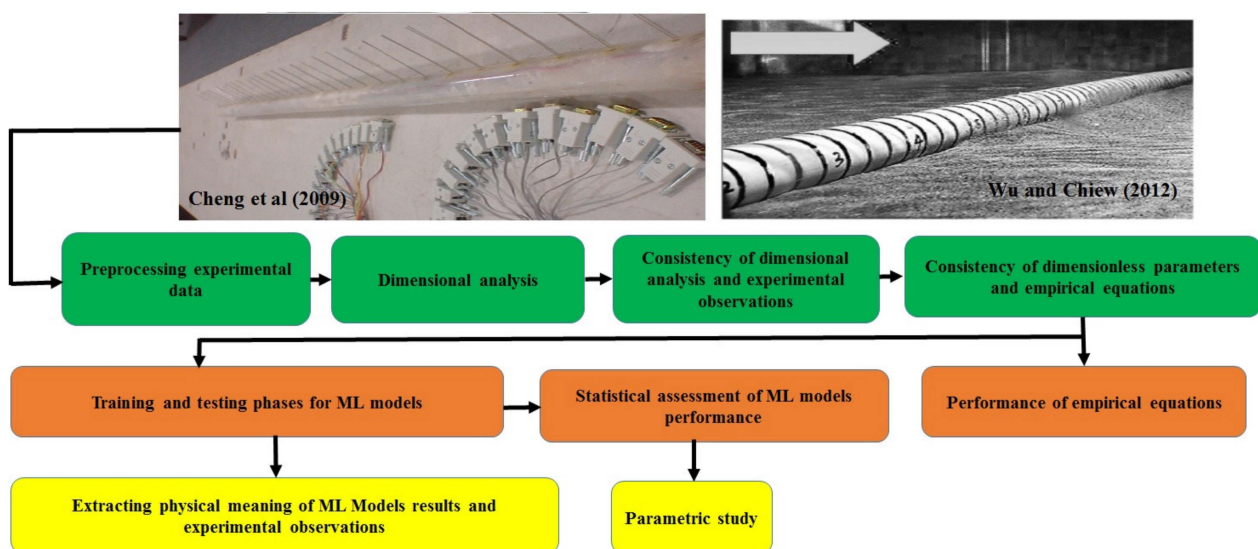


Figure 1. Flow chart of the present research work.

## 2. Overview of Databases

### 2.1. Dimensional Analysis

Based on the experimental studies from the literature, it appears that the approach flow intensity, the geometry of the offshore pipeline, the bed sediment mobility, and the physical properties of seabed sediments play a key role in scouring rate of propagation due to currents [6–8]. Specifically, the following function, ( $H_0$ ), which implies a relationship between the three-dimensional scour process and the governing variables, is assumed to be

$$H_0(V_L, u_*, e, D, U_C, d_{50}, \tan, \alpha, \rho, \rho_s, \mu, g) = 0 \quad (1)$$

where  $V_L$  is the scour rate along the pipeline,  $u_*$  is the flow shear velocity at the seabed,  $e$  is the embedment depth,  $D$  is the pipeline diameter,  $U_C$  is the flow velocity due to the current,  $d_{50}$  is the median sediment size,  $\phi$  is the angle of repose for the bed sediment,  $\alpha$  is the flow incident angle to the pipeline,  $\rho$  is the mass density of water,  $\rho_s$  is the mass density of bed sediment,  $\mu$  is the dynamic viscosity of water, and  $g$  is the acceleration due to gravity [6–8].

As mentioned in Cheng et al. [6] and Wu and Chiew [7,8], the analysis of scouring rates should be carried out considering effective dimensionless parameters. Furthermore, they investigated the physical consistency of their approaches using the experimental observations to corroborate the control of the selected variables on the scouring rate of propagation. These analyses were performed with dimensionless variables, ameliorating the scale effects related to the regression-based equations from laboratory observations. In terms of the application of ML models, the use of non-dimensional parameters in the

prediction of the rate of scouring propagation could enhance the ML models' performance compared to predictive models based on raw variables (e.g., Ehteram et al. [16]; Najafzadeh and Oliveto [17]). Therefore, the present investigation applies the  $\pi$ -theorem to identify a set of dimensionless parameters to be considered in ML models.

Among the raw variables in Equation (1), three variables ( $\rho, D, V_L$ ) were utilized as repeating variables to detect nine dimensionless parameters through the Buckingham theorem:

$$H_1(\pi_1, \pi_2, \pi_3, \pi_4, \pi_5, \pi_6, \pi_7, \pi_8, \pi_9) = 0 \tag{2}$$

where the above-mentioned parameters in Equation (2) can be specified as

$$\pi_1 = \frac{e}{D}, \pi_2 = \alpha, \pi_3 = \phi, \pi_4 = \frac{\rho_s}{\rho}, \pi_5 = \frac{gD}{U_C^2}, \pi_6 = \frac{V_L}{U_C}, \pi_7 = \frac{\rho U_C D}{\mu}, \pi_8 = \frac{D}{d_{50}}, \pi_9 = \frac{V_L}{u_*} \tag{3}$$

Before finalizing the dimensional analysis, some of the dimensionless parameters could be converted to a more attractive form, as highlighted in previous studies (Cheng et al. [6]; Wu and Chiew [8]; Najafzadeh and Oliveto [17]). The first and second parameters in Equation (3) were converted to  $1-e/D$  and  $1+\sin\alpha$ , respectively, in Cheng et al. [6]. According to Wu and Chiew [7], the fourth and fifth parameters in Equation (3) could be converted to  $\rho_s/\rho-1$  and the pipeline Froude number ( $Fr_p$ ), respectively. Moreover, the combination of dimensionless parameters in Equation (3) could be carried out in a way that more meaningful parameters are brought out. Therefore, the dimensionless parameters in Equation (3) were re-arranged as:

$$\pi'_1 = 1 - \frac{e}{D}, \pi'_2 = 1 + \sin \alpha, \pi'_3 = \tan \phi, \pi'_4 = \frac{\rho_s}{\rho} - 1, \pi'_5 = \frac{1}{\sqrt{\pi_5}} = \frac{U_C}{\sqrt{gD}} \tag{4}$$

$$\pi'_8 = \frac{(\pi_6)^2 \cdot \pi_8}{\pi'_4 \cdot (\pi_9)^2 \cdot \pi_5} = \frac{u_*^2}{g \left(\frac{\rho_s}{\rho} - 1\right) d_{50}} \tag{5}$$

$$\pi'_6 = \pi_8 \cdot (\pi'_8)^{0.5} \cdot \pi_9 \cdot \pi'_3 = \frac{V_L D \tan \phi}{\sqrt{g \left(\frac{\rho_s}{\rho} - 1\right) d_{50}^3}} \tag{6}$$

Therefore, that the functional relationship (2) can be made explicit as

$$H_2 \left( V_L^* = \frac{V_L D \tan \phi}{\sqrt{g \left(\frac{\rho_s}{\rho} - 1\right) d_{50}^3}}, 1 - \frac{e}{D}, Fr_p = \frac{U_C}{\sqrt{gD}}, \theta_C = \frac{u_*}{\sqrt{g \left(\frac{\rho_s}{\rho} - 1\right) d_{50}}}, Re_p = \frac{\rho U_C D}{\mu}, 1 + \sin \alpha \right) \tag{7}$$

where  $V_L^*$  is the dimensionless scouring rate propagation along the offshore pipeline,  $\theta_C$  is the Shields parameter due to the current, and  $Re_p$  is the pipeline's Reynolds number.

### 2.2. Description of Experimental Data

In this study, the experimental observations were obtained from three reliable studies.

Wu and Chiew [7] performed 51 experimental investigations considering uniform sand bed sediments with  $d_{50} = 0.56$  mm and geometric standard deviation ( $\sigma_g$ ) of 1.4. All experimental studies were conducted in a glass-sided flume whose length, width, and depth were 19 m, 1.6 m, and 0.45 m, respectively. The sediment recess section was 1.8 m long and 0.15 m deep and was placed 14 m downstream from the flume inlet. A total of 8 PVC pipelines, with  $D$  from 22 to 116 mm, were used. All experiments were performed under clear-water conditions and organized into four classes, each of which focused on the investigation of the impact of one of the dimensionless parameters  $e/D, Fr_p, \theta_C$ , and the ratio of the water depth ( $y$ ) on the pipeline diameter. For each class, the selected dimensionless parameter was varied, whereas all the other dimensionless parameters

remained unchanged. Scouring propagation rates were not observed in four experiments, which will therefore be excluded in the proposed modeling by ML techniques.

Cheng et al. [6] carried out 81 experiments to study the scouring propagation rates in a wave flume whose length, width, and depth were 50 m, 4 m, and 2.5 m, respectively. A concrete sandpit, 4 m long, 4 m wide, and 0.25 m deep, was constructed in the test section. A clear pipeline with a smooth surface, a diameter of 50 mm, and wall thickness of 8 mm was tested. All experiments were conducted under live-bed conditions ( $\theta_C = 0.046\text{--}0.104$ ), and the embedment depth  $e$  varied in the range from  $0.1D$  to  $0.5D$ . The  $d_{50}$  value and the relative density of sediment grains ( $S = \rho_s/\rho$ ) were 0.37 mm and 2.7, respectively. The sediment angle of repose,  $\phi$ , was kept constant and equal to  $32^\circ$ , the angle of attack  $\alpha$  ranged from 0 to  $45^\circ$ , and the dynamic viscosity of water,  $\mu$ , was equal to 0.001 Pa·s. Among the 81 experiments, piping processes occurred in 2 experiments, and, additionally, propagation scouring rates were not observed in 35 experiments. Therefore, 44 scouring tests were considered in the proposed modeling by ML techniques. For the sake of completeness, Cheng et al. [6] suggested the following relationship to estimate the scour propagation rate:

$$\frac{V_L D \tan \phi}{\sqrt{g(S-1)d_{50}^3}} = K \left[ 1 - \frac{e}{D}(1 + \sin \alpha) \right] \theta_C^{5/3} \tag{8}$$

where  $K$  is a constant that depends on  $\alpha$  values. Equation (8) relates to current conditions, and it is the only one available in the literature.

Hansen et al. [18] carried out 4 scouring experiments with  $d_{50} = 0.20$  mm and  $S = 2.65$  under clear-water conditions. They considered 2 pipeline diameters (i.e., 20 and 50 mm), and the approach flow depth, equal to 0.22 m, was kept constant during all experiments.

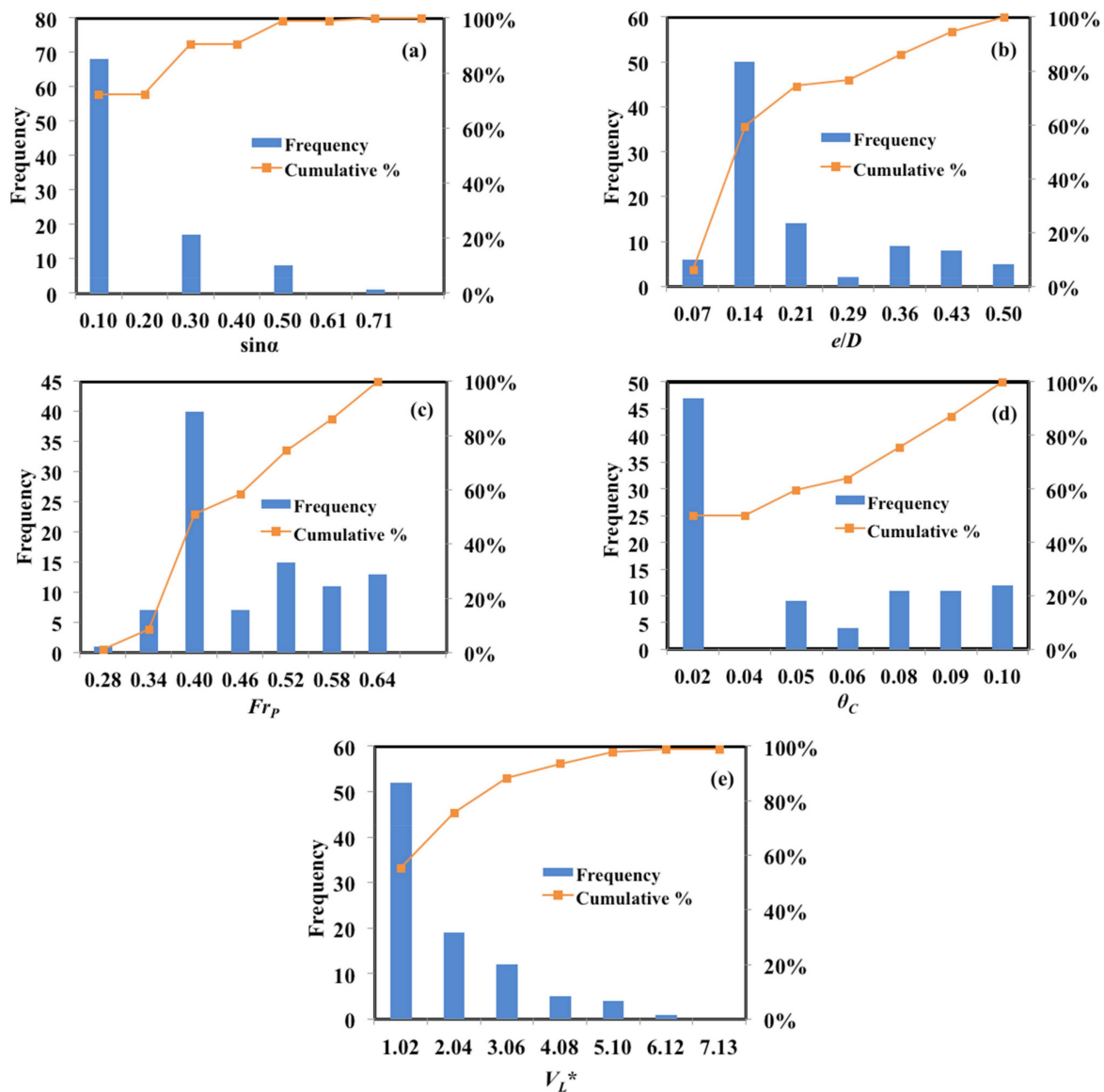
Ultimately, 95 experimental observations were obtained from Wu and Chiew [7] (47 datasets), Cheng et al. [6] (44 datasets), and Hansen et al. [18] (4 datasets), which have been considered to develop predictive models of the scour propagation rate in current conditions. Table 1 provides the statistical characterization of the variables from the above-mentioned datasets. To perform training and testing stages for the ML techniques, 75% (71 datasets) and 25% (24 datasets) were selected randomly, respectively. As seen in Table 1, the range of the Reynolds number  $Re_P$  is indicative of fully-turbulent flows, and this would imply that the effect of  $Re_P$  on the prediction of  $V_L^*$  is negligible.

**Table 1.** Statistical characterization of variables for feeding ML techniques.

| Dimensional Variables    | Minimum   | Maximum | Average   | Standard Deviation | Skewness |
|--------------------------|-----------|---------|-----------|--------------------|----------|
| $D$ (m)                  | 0.020     | 0.116   | 0.04425   | 0.012              | 1.538    |
| $e$ (m)                  | 0.0       | 0.025   | 0.00848   | 0.00643            | 1.168    |
| $d_{50}$ (m)             | 0.0002    | 0.00057 | 0.00055   | 0.0000719          | −4.7055  |
| $u^*$ (m/s)              | 0.0091    | 0.0314  | 0.02065   | 0.00624            | 0.2595   |
| $\alpha$ (degree)        | 0         | 45      | 5.454     | 10.1027            | 1.8212   |
| $U_C$ (m/s)              | 0.158     | 0.42    | 0.2817    | 0.07571            | 0.5154   |
| $y$ (m)                  | 0.09      | 0.470   | 0.2992    | 0.156              | 0.1067   |
| $V_L$ (m/s)              | 0.0000084 | 0.0056  | 0.00154   | 0.00122            | 1.2503   |
| Dimensionless Parameters | Minimum   | Maximum | Average   | Standard Deviation | Skewness |
| $e/D$                    | 0         | 0.5     | 0.1782    | 0.1256             | 1.224    |
| $\theta_C$               | 0.0091    | 0.104   | 0.04527   | 0.03221            | 0.5974   |
| $Fr_P$                   | 0.225     | 0.6321  | 0.4354    | 0.0927             | 0.4613   |
| $Re_P$                   | 5600      | 21,000  | 14,087.02 | 3917.98            | 0.40736  |
| $V_L^*$                  | 0.0000481 | 5.1082  | 0.9057    | 1.277              | 1.4825   |

Histograms for all the dimensionless parameters are illustrated in Figure 2a–e. These histograms show the frequency distributions for the explored (independent) parameters governing scouring processes, presenting succinct and beneficial details of the experimental

datasets. Incidentally, the analysis of these histogram analyses could help researchers in selecting unexplored ranges of investigation.



**Figure 2.** Frequency and cumulative relative frequency for the non-dimensional parameters considered in the present research: (a) flow attack angle  $\alpha$  to the offshore pipeline; (b) ratio of the pipeline embedded depth to pipeline diameter,  $e/D$ ; (c) approaching flow Froude number to the offshore pipeline,  $Fr_p$ ; (d) Shields parameter due to current,  $\theta_c$ ; (e) dimensionless scouring propagation rate around offshore pipeline along the longitudinal direction,  $V_L^*$ .

In Figure 2a, the  $\alpha$  parameter had a rather fragmented distribution of the frequency, with the highest fraction of the relative frequency (70%) for  $\sin \alpha = 0$ . As depicted in Figure 2b, although the distribution of the  $e/D$  parameter is fully fragmented, the pattern of the distribution is not symmetrical. Just 2 experiments were performed with  $e/D = 0.286$ , whereas about 50% of the experimental observations were carried out with  $e/D = 0.143$ . Moreover, Figure 2c shows that the approach flow Froude number,  $Fr_p$ , was not deeply explored according to a high uniform distribution, and the end tail of the histogram reaches a value of  $Fr_p$  equal to 0.636. Figure 2d,e also shows that, in the case of the Shields parameter ( $\theta_c$ ) and the dimensionless scouring propagation rate ( $V_L^*$ ), the distributions of the values are not symmetrical with maximum values for the relative frequencies around 50% and

55%, respectively. Additionally, most of the scouring tests were characterized by values of  $\theta_C$  and  $Fr_P$  approximately equal to 0.023 and 0.40, respectively. It is essential to mention that all the scouring tests were carried out utilizing various bed sediments and both states for the approach bed sediment mobility (i.e., live-bed and clear-water conditions).

### 3. Implementation of ML Models

#### 3.1. Gene-Expression Programming

GEP is known as one of the most powerful ML models; it works based on evolutionary algorithms (EAs). GEP utilizes a parse tree configuration to explore solutions, and it can provide a predictive formulation to interpret input-output systems. In addition to this, the overall formulation given by GEP includes a fair number of genetic operators [19,20]. Nevertheless, the GEP model employs a diagram of like-tree configurations with complexity, and it benefits from the merits of mathematical relationships to express a genome. The GEP model needs to assign some setting parameters, such as populations of individuals, number of generations, number of chromosomes, mutation rate, and linking functions among genes, to begin with. These factors play an important role in the goodness of the GEP performance during the training stage. GeneXproTools software was used to implement the GEP model. In the GEP performance, the validity of training and testing phases are controlled by the fitness values. In this study, root-mean-square error (RMSE) was selected as a fitness function to evaluate the GEP performance for each generation. Fitness values were scaled to 1000 by  $1000/(1 + RMSE)$ . In the GEP model, genetic operators are generally used by four strategies: optimal evolution (OE), constant fine-tuning (CFT), model fine-tuning (MFT), and sub-set selection (SSS). In this study, four alternatives of the GEP performance were considered. From the performances, the application of MFT (565.876) and OE (544.516) strategies stood at the higher goodness values in the training models in comparison with CFT (467.144) and SSS (536.205). Therefore, in the case of scouring problems below pipelines, optimal evolution methodology had promising usability; this study used the OE strategy as put forth in Najafzadeh and Oliveto [17]. The best value of the fitness function was equal to 544.516 in the training stage; the corresponding generation number obtained was 1713. Since there are four input variables for the training GEP model, the four genes can be considered the maximum number. There is no doubt that an increase in the number of genes in the structure of the GEP model causes, overall, a more complicated relationship (obtained by the GEP model). To reduce the complexity of the final equation, three genes were first considered. Then, it was found that the performance of the GEP model with three genes provided a better value of the fitness function than a GEP model developed with four genes. Table 2 indicates the setting parameters of the GEP model utilized in the optimal relationship for the prediction of the  $V_L^*$  parameter.

Equations (9)–(12), given by 4 alternatives (i.e., OE, SST, CFT, MFT) of the GEP model, were, respectively, obtained as follows:

$$V_L^* = (-\sin \alpha + 1.29145) \cdot (\theta_C - \sin \alpha + \frac{\epsilon}{D})^{5/3} + \left[ 9.3994 \cdot (2 + \sin \alpha) \cdot \theta_C^{1/3} - 0.5 \cdot (1 + \sin \alpha + \tanh 4.3693) \right]^{5/3} + \frac{3\theta_C - \frac{\epsilon}{D} - \sin \alpha}{Fr_P} \tag{9}$$

$$V_L^* = \frac{\theta_C^8 + (1 - \frac{\epsilon}{D})}{2} + \frac{2[\tanh \theta_C + \theta_C + \ln(1 - \frac{\epsilon}{D})]}{\theta_C + Fr_P} - 78.3404 \cdot \left( \frac{\sin \alpha - 0.5248}{2} - \theta_C \right) \cdot \theta_C \tag{10}$$

$$V_L^* = \left\{ 0.5 \cdot (-16.8294 - \theta_C - \sin \alpha)^{5/3} - [\tanh(1 + \sin \alpha) + 1 - \frac{\epsilon}{D}] \right\}^{1/3} + \tan^{-1}(1.3579 \cdot \theta_C) + \frac{39.2001 + (1 + \sin \alpha) \cdot Fr_P}{2} + \ln \{ \min[\tanh \theta_C, \max(\theta_C, 1 - \frac{\epsilon}{D})] - \frac{\epsilon}{D} + \sin \alpha - 12.4093 \} \tag{11}$$

$$V_L^* = \left\{ \left[ \tan^{-1} \left( 1 - \frac{e}{D} \right) - Fr_p + \frac{1}{1 + \sin \alpha} \right] \cdot \frac{15.2114 + \sin \alpha}{2} \cdot \theta_C^{5/3} \right\} + \{ [1 + 11.129 \cdot (1 - \frac{e}{D}) + \exp(2 + \sin \alpha - \frac{e}{D})] \cdot \theta_C \}^2 + 1 - \min \{ [2 \cdot (1 + \sin \alpha) + \max(\theta_C, 1 + \sin \alpha)], 9.5516 \cdot (1 + \sin \alpha) - \tan^{-1} \theta_C \} \tag{12}$$

**Table 2.** Setting parameters of the proposed structure of GEP models.

| Parameters      | Description of Parameters                   | Setting of Parameters  |
|-----------------|---|--|
| P <sub>1</sub>  | Function set                                | +, −, ×, /, Power(x <sup>2</sup> ), (1-x), Average(x <sub>1</sub> , x <sub>2</sub> ), Tanh(x), 3Rt, Ln, Min, Max |
| P <sub>2</sub>  | Linking function                            | Addition   |
| P <sub>3</sub>  | Mutation rate                               | OE (0.003), CFT (0), MFT (0), SSS (0)  |
| P <sub>4</sub>  | Inversion rate                              | OE (0.00546), CFT (0), MFT (0), SSS (0.0082)   |
| P <sub>5</sub>  | One-point and two-point recombination rates | OE (0.00277), CFT (0), MFT (0), SSS (0.0028)   |
| P <sub>6</sub>  | Gene recombination rate                     | OE (0.00277), CFT (0), MFT (0.0129), SSS (0.0028)  |
| P <sub>7</sub>  | Permutation                                 | OE (0.00546), CFT (0), MFT (0), SSS (0.0082)   |
| P <sub>8</sub>  | Maximum tree depth                          | OE (4), CFT (3), MFT (5), SSS (4)  |
| P <sub>9</sub>  | Number of genes                             | 3  |
| P <sub>10</sub> | Number of chromosomes                       | 30   |
| P <sub>11</sub> | Number of generations                       | OE (1713), CFT (961), MFT (2700), SSS (909)  |
| P <sub>12</sub> | Best fitness value                          | OE (544.516), CFT (467.144), MFT (565.876), SSS (536.205)  |

### 3.2. Multivariate Adaptive Regression Splines

Multivariate adaptive regression splines (MARS) are a globally-recognized ML model which generates mathematical expressions by the development of linear regression [21]. The MARS model generally includes second-order spline regression. Formulas are created by the cross-validation conception that can automatically control the nonlinearities of the obtained equation and interactions among variables. To generate an equation based on spline regression, the MARS model generally makes use of a number of basis functions (BFs) with their related weighted coefficients [21]. Each BF is introduced by a variable and a knot. MARS creates an expression during two phases: the forward and the backward passes. In the forward pass, the MARS model starts with a model that consists of just the intercept/bias term, which is the mean of the output parameter values. Then, the MARS model repeatedly adds BFs in pairs to the second-order regression spline model. At each step, it detects the pair of BFs that can result in the maximum reduction value in sum-of-squares residual error. This process of adding BFs continues until the variation in residual error is too small to continue or until the maximum number of BFs is met. With regards to this research, the MARS technique approximates the scouring propagation rate through the following equation:

$$\omega(r) = T_0 + \sum_{j=1}^{NBF} T_j \cdot BF_j(e/D, \theta_C, Fr_p, 1 + \sin \alpha) \tag{13}$$

in which  $T_0$ ,  $T_j$ , BF,  $r$ , and NBF are the bias, the constant coefficients related to basis functions, the basis function, the set of dimensionless parameters, and the number of basis functions, respectively. The performance of the forward pass occasionally generates an over-fitted model. To generate an expression with more efficient generalization potential, the backward pass is performed to prune the initially-extended MARS expression. This process deletes terms one by one, eliminating the lowest effective term at each stage until it obtains the best sub-model. The performance of model subsets is evaluated using the generalized cross-validation (GCV) criterion discussed in the literature [21].



In the present research, the MARS technique was run by a computer code written in MATLAB software. The MARS model was primitively created by 10 BFs and 23.5 effective parameters. To reduce the complexity level of the initially developed MARS model, the analysis of GCV was performed. In addition to this, the *k*-fold was equal to 10 such that the MARS technique was performed 10 times. The forward and backward stages were carried out for each *k*-fold value; additionally, the number of basis functions in the final MARS model and the total effective number of parameters were obtained. The definition of the 10 performances of the MARS model is provided in Table 3, in which the results during the forward and backward stages are shown. Hence, 15 BFs were obtained to predict the scour propagation rate below offshore pipelines

$$V_L^* = -231.99 - 751.4 \cdot BF_1 + 1212.7 \cdot BF_2 + 5648.3 \cdot BF_3 + 6503.5 \cdot BF_4 - 13716 \cdot BF_5 + 87.716 \cdot BF_6 - 4191.7 \cdot BF_7 + 4261 \cdot BF_8 - 4215.7 \cdot BF_9 + 4210.5 \cdot BF_{10} - 72.507 \cdot BF_{11} + 6658.5 \cdot BF_{12} - 902.35 \cdot BF_{13} + 774.61 \cdot BF_{14} - 777.01 \cdot BF_{15} \tag{14}$$

in which the BF<sub>1</sub> to BF<sub>15</sub> formulations are presented in Table 3. All four input parameters (*Fr<sub>p</sub>*, *θ<sub>C</sub>*, *e/D*, *1+sinα*) had contributions in driving Equation (14) to estimate *V<sub>L</sub><sup>\*</sup>*, as can be inferred from Table 4. The optimal value of GCV (0.5662) reduced the probability of over-parametrization of the MARS expression. This means that the MARS expression could efficiently detect input variables that have lower importance in the prediction of the scouring propagation rate. All coefficients in Equation (14) were fitted by the particle swarm optimization (PSO) algorithm, providing MSE = 0.2252 as the best results.

**Table 3.** Characterizations of MARS models during the forward and backward stages.

| <i>k</i> -Fold | Number of BFs in the Final Model | Total Effective Number of Parameters | Executive Time (s) |
|----------------|----------------------------------|--------------------------------------|--------------------|
| 1              | 11                               | 26                                   | 0.57               |
| 2              | 7                                | 16                                   | 0.69               |
| 3              | 14                               | 33.5                                 | 0.99               |
| 4              | 9                                | 21                                   | 0.71               |
| 5              | 12                               | 28.5                                 | 0.74               |
| 6              | 12                               | 28.5                                 | 0.66               |
| 7              | 11                               | 26                                   | 0.67               |
| 8              | 12                               | 28.5                                 | 0.99               |
| 9              | 6                                | 13.5                                 | 0.71               |
| 10             | 14                               | 33.5                                 | 0.58               |

### 3.3. Evolutionary Polynomial Regression

EPR, as a robust data-driven models (DDMs), is developed by the multi-objective genetic algorithm (MOGA) to be applied in four various ways: (i) data analysis from the input-output records, (ii) data modeling for both static and dynamic systems, (iii) symbolic expressions for mathematical models, and (iv) decision support for the model selection [22–24]. Overall, EPR is an integrative model that simultaneously recruits the efficacy of MOGA with numerical regression techniques for developing simple knowledge extraction of mathematical expressions [25,26].

For the prediction of the scouring propagation rate, two types of mathematical expressions, which had the ordinary structures *y* = bias + ∑*a*·*X*<sub>1</sub>·*X*<sub>2</sub>·*f*(*X*<sub>1</sub> × *X*<sub>2</sub>) and *y* = bias + ∑*a*·*X*<sub>1</sub>·*X*<sub>2</sub>·*f*(*X*<sub>1</sub>)·*f*(*X*<sub>2</sub>) were used to develop the EPR model. *X* is the vector of the input variables, *y* is the approximated output, and *a* is the set of coefficients. In this way, the following mathematical relationships were selected:

$$V_L^* = O_0 + \sum_{j=1}^m \left\{ O_j \cdot (\theta_C)^{ES(j,1)} \cdot \left(1 - \frac{e}{D}\right)^{ES(j,2)} \cdot (1 + \sin \alpha)^{ES(j,3)} \cdot (Fr_p)^{ES(j,4)} \cdot f \left[ (\theta_C)^{ES(j,5)} \cdot \left(1 - \frac{e}{D}\right)^{ES(j,6)} \cdot (1 + \sin \alpha)^{ES(j,7)} \cdot (Fr_p)^{ES(j,8)} \right] \right\} \tag{15}$$

$$V_L^* = O_0 + \sum_{j=1}^m < O_j \cdot (\theta_C)^{ES(j,1)} \cdot \left(1 - \frac{e}{D}\right)^{ES(j,2)} \cdot (1 + \sin \alpha)^{ES(j,3)} \cdot (Fr_P)^{ES(j,4)} \cdot \left\{ f \left[ (\theta_C)^{ES(j,5)} \right] \cdot f \left[ \left(1 - \frac{e}{D}\right)^{ES(j,6)} \right] \cdot f \left[ (1 + \sin \alpha)^{ES(j,7)} \right] \cdot f \left[ (Fr_P)^{ES(j,8)} \right] \right\} > \tag{16}$$

**Table 4.** Coefficients and basis functions for MARS models in scour rate prediction.

|                 | Basis Functions  |                  | Basis Functions  |
|-----------------|--|------------------|--|
| BF <sub>1</sub> | max (0, 0.1−e/D)   | BF <sub>9</sub>  | max (0, Fr <sub>P</sub> −0.34623)·max (0, 0.3−e/D)               |
| BF <sub>2</sub> | max (0, Fr <sub>P</sub> −0.34623)                                | BF <sub>10</sub> | max (0, Fr <sub>P</sub> −0.34623)·max (0, e/D−0.3)               |
| BF <sub>3</sub> | max (0, Fr <sub>P</sub> −0.34623)·max (0, 0.046−θ <sub>C</sub> ) | BF <sub>11</sub> | max (0, Fr <sub>P</sub> −0.47128)                                |
| BF <sub>4</sub> | max (0, θ <sub>C</sub> −0.018)·max (0, 0.1−e/D)                  | BF <sub>12</sub> | max (0, 0.47128−Fr <sub>P</sub> )·max (0, θ <sub>C</sub> −0.014) |
| BF <sub>5</sub> | max (0, Fr <sub>P</sub> −0.34623)·max (0, 0.054−θ <sub>C</sub> ) | BF <sub>13</sub> | max (0, θ <sub>C</sub> −0.014)                                   |
| BF <sub>6</sub> | max (0, 1−sinα)·max (0, θ <sub>C</sub> −0.021)                   | BF <sub>14</sub> | max (0, e/D−0.6)   |
| BF <sub>7</sub> | max (0, 0.1−e/D)·max (0, Fr <sub>P</sub> −0.52841)               | BF <sub>15</sub> | max (0, 0.6−e/D)   |
| BF <sub>8</sub> | max (0, 0.1−e/D)·max (0, 0.52841−Fr <sub>P</sub> )               | BF <sub>16</sub> | 0  |

In the above Equations (15) and (16),  $O_0$  is the bias term,  $m$  is the maximum value of the mathematical terms,  $O_j$  is a set of coefficients,  $f$  is a user-defined function, and the ES function is a range of exponents explored by the EPR model. Previous investigations demonstrated that the use of EPR expressions with natural logarithmic inner functions had more promising efficacy in predicting the scouring propagation rates below offshore pipelines exposed to regular waves compared to predictions obtained without inner function [17]. Therefore, six terms and logarithmic inner functions in Equations (15) and (16) were considered for the EPR model. The number of generations was 4800. This issue was automatically computed regarding many factors such as the number of input parameters, the kind of inner function, and the number of dataset rows. Tables 5 and 6 indicate the ultimate EPR expressions during training stages based on Equations (15) and (16), respectively. According to Table 5, the first model (Model #1) had the least complexity in expression with four logarithmic terms, whereas the MSE value (0.582) indicated the lowest accuracy in the prediction of scouring propagation rates. Model #5 included 6 terms with 5 logarithmic terms, providing the best results (MSE = 0.341) along with the most complicated expression. It was inferred from Table 5 that an increase in the presence of the input variables in each logarithm term could increase the precision level of approximation. For instance, the first term of Model #5 included 4 input parameters (i.e.,  $\theta_C$ ,  $1 - e/D$ ,  $1 + \sin \alpha$ ,  $Fr_P$ ); from Model #1 to Model #5, the more complicated the mathematical expression, the higher the precision level. Based on the expressions in Table 6, Model #5 had the highest level of accuracy (MSE = 0.585) in the training phase, whereas Model #1 had the lowest accuracy with an MSE of 0.747. Similarly, the complexity of terms and the number of terms played a key role in improving the efficiency of the expression returned by Equation (16). For instance, Model #2 (see Table 6) had one term with an MSE of 0.646, whereas Model #3 yielded more accurate predictions (MSE = 0.620) with two algebraic terms. Table 6 indicates that the third to fifth expressions (Model #3 to Model #5) had the same number of algebraic terms while providing various performance levels in the prediction of scouring propagation rate due to the existing various complexity levels in each expression. Generally, the EPR expressions given in Table 5 had more complexity and a number of algebraic terms in comparison with those presented in Table 6. Then, equations returned by EPR (based on Equation (15)) had more promising results than equations given in Table 6.

**Table 5.** Mathematical expressions developed by Equation (15) with logarithmic inner function.

| Model No. | Expression   | MSE   |
|-----------|--|-------|
| Model #1  | $V_L^* = 1.2869 + \frac{6.2646}{1+\sin\alpha} + 20.0234 \cdot \ln\left[(Fr_p)^{0.5}\right] + 2.4791 \cdot \ln(\theta_C)$ $+ 5.1071 \frac{(1-\frac{\epsilon}{D})^{0.5}}{Fr_p}$  | 0.582 |
| Model #2  | $V_L^* = 78.14 + \frac{6.3408}{1+\sin\alpha} + 20.0781 \cdot \ln\left[(\theta_C \cdot Fr_p)^{0.5}\right] + 6.6196(1 - \frac{\epsilon}{D}) + 33.5496 \frac{\theta_C^{0.5}}{Fr_p}$ $\cdot \ln\left[\frac{1}{(Fr_p)^{0.5}}\right] + 84.4451 \cdot \ln(\theta_C)$  | 0.460 |
| Model #3  | $V_L^* = 39.0219 + 0.12858 \cdot \frac{1-\frac{\epsilon}{D}}{\theta_C^{0.5}} \cdot \ln(\theta_C) + \frac{26.1221}{1+\sin\alpha} \cdot \ln\left[\left(\frac{1}{1+\sin\alpha}\right)^{0.5}\right] + 27.357$ $\cdot \ln\left[\left(\frac{Fr_p}{1+\sin\alpha}\right)^{0.5}\right] + 12.566 \cdot \frac{1-\frac{\epsilon}{D}}{1+\sin\alpha} \cdot \ln\left(1 - \frac{\epsilon}{D}\right) + 66.0515 \cdot \theta_C^{0.5}$ $\cdot \ln\left[\theta_C \cdot (Fr_p)^{0.5}\right] + 54.7801 \cdot \frac{[\theta_C \cdot (1+\sin\alpha)]^{0.5}}{Fr_p}$   | 0.449 |
| Model #4  | $V_L^* = 47.0887 + 0.2218 \cdot \frac{(1-\frac{\epsilon}{D}) \cdot (1+\sin\alpha)}{\theta_C^{0.5}} \cdot \ln\left[(\theta_C)^{0.5} \cdot Fr_p\right] + 50.376 \cdot \frac{Fr_p}{1+\sin\alpha}$ $\cdot \ln\left[\left(\frac{1}{1+\sin\alpha}\right)^{0.5}\right] + 30.8495 \cdot \ln\left[\left(\frac{Fr_p}{1+\sin\alpha}\right)^{0.5}\right] + 27.1979$ $\cdot \frac{1-\frac{\epsilon}{D}}{1+\sin\alpha} \cdot \ln\left[\left(1 - \frac{\epsilon}{D}\right)^{0.5}\right] + 80.6509 \cdot \theta_C^{0.5} \cdot \ln\left[\theta_C \cdot (Fr_p)^{0.5}\right]$ $+ 64.6131 \cdot \frac{[\theta_C \cdot (1+\sin\alpha)]^{0.5}}{Fr_p}$  | 0.437 |
| Model #5  | $V_L^* = 45.7795 + 0.21215 \cdot \frac{(1-\frac{\epsilon}{D}) \cdot (1+\sin\alpha)}{\theta_C^{0.5}} \cdot \ln\left[\frac{(1-\frac{\epsilon}{D})^{0.5} \cdot Fr_p}{(1+\sin\alpha)^{0.5}}\right] + 48.0925$ $\cdot \frac{Fr_p}{1+\sin\alpha} \cdot \ln\left[\left(\frac{1}{1+\sin\alpha}\right)^{0.5}\right] + 30.3275 \cdot \ln\left[\left(\frac{Fr_p}{1+\sin\alpha}\right)^{0.5}\right]$ $+ 27.0613 \cdot \frac{1-\frac{\epsilon}{D}}{1+\sin\alpha} \cdot \ln\left[\left(1 - \frac{\epsilon}{D}\right)^{0.5}\right] + 78.6329 \cdot \theta_C^{0.5}$ $\cdot \ln\left[\theta_C \cdot (Fr_p)^{0.5}\right] + 63.5902 \cdot \frac{[\theta_C \cdot (1+\sin\alpha)]^{0.5}}{Fr_p}$ | 0.341 |

**Table 6.** Mathematical expressions developed by Equation (16) with logarithmic inner function.

| Model No. | Expression  | MSE   |
|-----------|---|-------|
| Model #1  | $V_L^* = 168.4265 + \frac{\theta_C \cdot (1-\frac{\epsilon}{D})}{Fr_p \cdot (1+\sin\alpha)} \cdot \ln\left[(\theta_C + 1)^{0.5}\right] \cdot \ln\left(2 - \frac{\epsilon}{D}\right)$  | 0.747 |
| Model #2  | $V_L^* = 127.334 \cdot \frac{\theta_C \cdot (1-\frac{\epsilon}{D})}{Fr_p \cdot (1+\sin\alpha)} \cdot \frac{\ln[(\theta_C+1)^{0.5}] \cdot \ln(2-\frac{\epsilon}{D})}{\ln(2+\sin\alpha)}$   | 0.646 |
| Model #3  | $V_L^* = 41.7624 \cdot \frac{\theta_C^{0.5} \cdot (1-\frac{\epsilon}{D})}{Fr_p \cdot (1+\sin\alpha)} \cdot \frac{\ln(\theta_C+1) \cdot \ln(2-\frac{\epsilon}{D})}{\ln(2+\sin\alpha) \cdot \ln(1+Fr_p)} + 191.4399$ $\cdot \frac{\theta_C \cdot Fr_p}{(1+\sin\alpha)} \cdot \ln(\theta_C + 1)$   | 0.620 |
| Model #4  | $V_L^* = 39.5542 \cdot \frac{\theta_C^{0.5} \cdot (1-\frac{\epsilon}{D})}{Fr_p \cdot (1+\sin\alpha)} \cdot \frac{\ln(\theta_C+1) \cdot \ln(2-\frac{\epsilon}{D})}{\ln(2+\sin\alpha) \cdot \ln(1+Fr_p)} + 557.6596$ $\cdot \frac{\theta_C \cdot Fr_p \cdot (1-\frac{\epsilon}{D})}{(1+\sin\alpha) \cdot \ln(2+\sin\alpha)} \cdot \ln(1 + Fr_p) \cdot \ln(\theta_C + 1)$  | 0.596 |
| Model #5  | $V_L^* = 39.0458 \cdot \frac{\theta_C^{0.5} \cdot (1-\frac{\epsilon}{D})}{Fr_p \cdot (1+\sin\alpha)} \cdot \frac{\ln(\theta_C+1) \cdot \ln(2-\frac{\epsilon}{D})}{\ln(2+\sin\alpha) \cdot \ln(1+Fr_p)} + 775.121$ $\cdot \frac{\theta_C \cdot Fr_p \cdot (1-\frac{\epsilon}{D})}{(1+\sin\alpha)} \cdot \frac{\ln(\theta_C+1) \cdot \ln\left[\left(2 - \frac{\epsilon}{D}\right)^{0.5}\right] \cdot \ln(1+Fr_p)}{\ln(2+\sin\alpha)}$ | 0.585 |

3.4. M5 Model Tree

M5MT, as a newly-established system, is generally used to learn models that estimate values. Similar to ML techniques based on the classification concepts, such as MARS and classification and regression tree (CART) models, M5 is capable of creating tree-based expressions; these regression trees provide real values at their leaves. The trees given by M5 can provide multivariate linear expressions. The M5 version of MT efficiently solves problems with high dimensionality. Compared with MARS and CART, M5MT can interpret nonlinear systems with faster executive time performance. The main merit of M5MT over

the CART model is that trees have a smaller size (in terms of leaves and nodes) and a higher level of precision in the multi-task systems [27]. In M5MT, tree-like structures are provided by the divide-and-conquer technique. M5 splits the search space of data points into several subdivisions. Then, multilinear regression models are fitted on data points in each subdivision. M5MT is implemented based on several steps such as tree structure construction, error estimation, linear modeling, linear model simplification, pruning, and smoothing [27–29].

In this study, Weka3.9 software was utilized to develop M5MT for estimating  $V_L^*$  values around offshore pipelines. The following multilinear regression model is generally utilized in developing M5MT:

$$V_L^* = a_0 + a_1 \cdot \theta_C + a_2 \cdot \left(1 - \frac{e}{D}\right) + a_3 \cdot (1 + \sin \alpha) + a_4 \cdot Fr_P \tag{17}$$

in which  $a_0$  is the bias term, and  $a_1$  to  $a_4$  are weighting coefficients that are computed by the least-square technique.

This research used four alternatives of M5MT based on the usability of the pruning and smoothing stages in a way that if the pruning (or smoothing) stage is considered, its state is “True” (T); otherwise, it is called “False” (F). The impacts of pruning and smoothing stages on the performance of M5MT were conceptually evaluated by RMSE values during the training and testing stages. The first alternative, M5MT#1, is related to using the pruned (T) and smoothed (T) M5MT with four rules, as seen in Table 7. The Shields parameter was adjusted as a sole splitting parameter for constructing four rules. Table 7 indicates that all the input parameters were used to provide multilinear regression equations related to their rules. Table 7 demonstrates the results of M5MT for the unpruned (F) and smoothed (T) phases. Similar to Table 7, the performance of M5MT#2 yielded four rules (see Table 8), and, consequently, four multilinear regression relationships. As demonstrated in Table 8, the first rule provided only a bias term (0.0512), and additionally,  $Fr_P$  was incorporated to model a linear equation for the second rule. In the case of unpruned (F) and smoothed (T) M5MT, a list of rules and relevant multilinear regression equations M5MT#3 were provided in Table 9. As seen in Table 10, all the input parameters were incorporated into 29 driving regression equations and search spaces. Regarding unpruned (F) and unsmoothed (F) M5MT, 29 rules were obtained, and M5MT#4 consequently provided 29 bias terms without incorporating the input parameters. Details of M5MT#4 performance are given in Table 11.

**Table 7.** List of M5MT#1 details in  $V_L^*$  estimates.

| Rules of M5MT#1 Focusing on Pruning and Smoothing Phases |   |
|--|---|
| If $\theta_C \leq 0.018$ :                               |   |
|  | If $\theta_C \leq 0.013$ :LM1   |
|  | If $\theta_C > 0.013$ :LM2  |
| If $\theta_C > 0.018$ :                                  |   |
|  | If $\theta_C \leq 0.05$ :LM3  |
|  | If $\theta_C > 0.05$ :LM4   |
| Multilinear Regression Equations                         |   |
| LM1  | $V_L^* = -2.0495 + 20.7493\theta_C + 1.4622(1 - e/D) - 0.8277(1 + \sin\alpha) + 5.2496Fr_P$ |
| LM2  | $V_L^* = -2.1368 + 16.5994\theta_C + 1.4622(1 - e/D) - 0.8277(1 + \sin\alpha) + 5.7554Fr_P$ |
| LM3  | $V_L^* = 0.062 + 18.2934\theta_C + 3.8163(1 - e/D) - 2.4011(1 + \sin\alpha) - 0.3838Fr_P$   |
| LM4  | $V_L^* = -2.0857 + 24.1656\theta_C + 5.3836(1 - e/D) - 3.3503(1 + \sin\alpha) + 4.4672Fr_P$ |

**Table 8.** List of M5MT#2 details in  $V_L^*$  estimates.

| Multilinear Regression Equations |   |
|----------------------------------|---|
| LM1                              | $V_L^* = 0.0512$  |
| LM2                              | $V_L^* = -0.6047 + 2.5289Fr_P$  |
| LM3                              | $V_L^* = 0.062 + 37.1095\theta_C + 4.4068(1 - e/D) - 2.899(1 + \sin\alpha) - 10.5248Fr_P$ |
| LM4                              | $V_L^* = -1.1712 + 45.4636\theta_C + 7.2873(1 - e/D) - 4.6271(1 + \sin\alpha)$            |

**Table 9.** List of driven rules of M5MT#3 in  $V_L^*$  estimates.

---

|                                  |
|----------------------------------|
| If $\theta_C \leq 0.018$ :       |
| If $\theta_C \leq 0.013$ : LM1   |
| If $\theta_C > 0.013$ :          |
| If $Fr_p \leq 0.367$ : LM2       |
| If $Fr_p > 0.367$ :              |
| If $\theta_C \leq 0.015$ : LM3   |
| If $\theta_C > 0.015$ : LM4      |
| If $\theta_C > 0.018$ :          |
| If $\theta_C \leq 0.05$ :        |
| If $Fr_p \leq 0.361$ :           |
| If $1 - e/D \leq 0.89$ :         |
| If $1 - e/D \leq 0.82$ : LM5     |
| If $1 - e/D > 0.82$ : LM6        |
| If $1 - e/D > 0.89$ :            |
| If $1 - e/D \leq 0.94$ : LM7     |
| If $1 - e/D > 0.94$ : LM8        |
| If $Fr_p > 0.361$ :              |
| If $1 + \sin\alpha \leq 1.129$ : |
| If $\theta_C \leq 0.019$ :       |
| If $Fr_p \leq 0.425$ : LM9       |
| If $Fr_p > 0.425$ : LM10         |
| If $\theta_C > 0.019$ :          |
| If $Fr_p \leq 0.387$ : LM11      |
| If $Fr_p > 0.387$ :              |
| If $1 - e/D \leq 0.7$ : LM12     |
| If $1 - e/D > 0.7$ :             |
| If $Fr_p \leq 0.42$ : LM13       |
| If $Fr_p > 0.42$ : LM14          |
| If $1 + \sin\alpha > 1.129$ :    |
| If $1 - e/D \leq 0.85$ : LM15    |
| If $1 - e/D > 0.85$ : LM16       |
| If $\theta_C > 0.05$ :           |
| If $1 + \sin\alpha \leq 1.129$ : |
| If $1 - e/D \leq 0.75$ :         |
| If $\theta_C \leq 0.092$ :       |
| If $1 - e/D \leq 0.65$ : LM17    |
| If $1 - e/D > 0.65$ : LM18       |
| If $\theta_C > 0.092$ : LM19     |
| If $1 - e/D > 0.75$ :            |
| If $1 - e/D \leq 0.95$ :         |
| If $\theta_C \leq 0.073$ :       |
| If $\theta_C \leq 0.059$ : LM20  |
| If $\theta_C \leq 0.059$ : LM20  |
| If $\theta_C > 0.059$ : LM21     |
| If $\theta_C > 0.073$ : LM22     |
| If $1 - e/D > 0.95$ : LM23       |
| If $1 + \sin\alpha > 1.129$ :    |
| If $\theta_C \leq 0.092$ :       |
| If $1 - e/D \leq 0.85$ :         |
| If $\theta_C \leq 0.073$ : LM24  |
| If $\theta_C > 0.073$ :          |
| If $1 - e/D \leq 0.75$ : LM25    |
| If $1 - e/D > 0.75$ : LM26       |
| If $1 - e/D > 0.85$ : LM27       |
| If $\theta_C > 0.092$ :          |
| If $1 - e/D \leq 0.75$ : LM28    |
| If $1 - e/D > 0.75$ : LM29       |

---

**Table 10.** List of multilinear regression models extracted from M5MT#3 rules.

| Multilinear Regression Equations |   |
|----------------------------------|---|
| LM1                              | $V_L^* = -2.0495 + 20.7493\theta_C + 1.4622(1 - e/D) - 0.8277(1 + \sin\alpha) + 5.2496Fr_P$ |
| LM2                              | $V_L^* = -2.0955 + 16.5994\theta_C + 1.46222(1 - e/D) - 0.8277(1 + \sin\alpha) + 5.629Fr_P$ |
| LM3                              | $V_L^* = -2.0867 + 16.5994\theta_C + 1.46222(1 - e/D) - 0.8277(1 + \sin\alpha) + 5.629Fr_P$ |
| LM4                              | $V_L^* = -2.0876 + 16.5994\theta_C + 1.46222(1 - e/D) - 0.8277(1 + \sin\alpha) + 5.629Fr_P$ |
| LM5                              | $V_L^* = -1.5165 + 11.9305\theta_C + 4.4349(1 - e/D) - 1.904(1 + \sin\alpha) + 1.4209Fr_P$  |
| LM6                              | $V_L^* = -1.5153 + 11.9305\theta_C + 4.4349(1 - e/D) - 1.904(1 + \sin\alpha) + 1.4209Fr_P$  |
| LM7                              | $V_L^* = -1.3564 + 11.9305\theta_C + 4.4349(1 - e/D) - 1.904(1 + \sin\alpha) + 1.4209Fr_P$  |
| LM8                              | $V_L^* = -1.3597 + 37.1095\theta_C + 4.4068(1 - e/D) - 1.904(1 + \sin\alpha) + 1.4209Fr_P$  |
| LM9                              | $V_L^* = -0.9468 + 14.6674\theta_C + 3.2431(1 - e/D) - 1.9373(1 + \sin\alpha) + 2.3725Fr_P$ |
| LM10                             | $V_L^* = -0.9462 + 14.6674\theta_C + 3.2431(1 - e/D) - 1.9373(1 + \sin\alpha) + 2.3725Fr_P$ |
| LM11                             | $V_L^* = -1.0046 + 14.6674\theta_C + 3.2431(1 - e/D) - 1.9373(1 + \sin\alpha) + 2.5936Fr_P$ |
| LM12                             | $V_L^* = -0.9923 + 14.1337\theta_C + 3.2539(1 - e/D) - 1.9373(1 + \sin\alpha) + 2.5434Fr_P$ |
| LM13                             | $V_L^* = -0.9881 + 14.1337\theta_C + 3.2505(1 - e/D) - 1.9373(1 + \sin\alpha) + 2.5434Fr_P$ |
| LM14                             | $V_L^* = -0.9880 + 14.1337\theta_C + 3.2505(1 - e/D) - 1.9373(1 + \sin\alpha) + 2.5434Fr_P$ |
| LM15                             | $V_L^* = -0.7742 + 13.7781\theta_C + 3.2258(1 - e/D) - 2.0772(1 + \sin\alpha) + 2.3725Fr_P$ |
| LM16                             | $V_L^* = -0.7732 + 13.7781\theta_C + 3.2258(1 - e/D) - 2.0772(1 + \sin\alpha) + 2.3725Fr_P$ |
| LM17                             | $V_L^* = -3.6161 + 26.0093\theta_C + 5.6375(1 - e/D) - 2.1206(1 + \sin\alpha) + 4.4672Fr_P$ |
| LM18                             | $V_L^* = -3.6169 + 26.0093\theta_C + 5.6465(1 - e/D) - 2.1206(1 + \sin\alpha) + 4.4672Fr_P$ |
| LM19                             | $V_L^* = -3.6173 + 26.7266\theta_C + 5.5808(1 - e/D) - 2.1206(1 + \sin\alpha) + 4.4672Fr_P$ |
| LM20                             | $V_L^* = -2.6839 + 21.7296\theta_C + 4.8669(1 - e/D) - 2.1206(1 + \sin\alpha) + 4.4672Fr_P$ |
| LM21                             | $V_L^* = -2.6847 + 21.7296\theta_C + 4.8669(1 - e/D) - 2.1206(1 + \sin\alpha) + 4.4672Fr_P$ |
| LM22                             | $V_L^* = -2.7195 + 22.4079\theta_C + 4.8669(1 - e/D) - 2.1206(1 + \sin\alpha) + 4.4672Fr_P$ |
| LM23                             | $V_L^* = -2.6416 + 21.516\theta_C + 4.8669(1 - e/D) - 2.1206(1 + \sin\alpha) + 4.4672Fr_P$  |
| LM24                             | $V_L^* = -3.2052 + 24.0158\theta_C + 4.6143(1 - e/D) - 2.0809(1 + \sin\alpha) + 4.4672Fr_P$ |
| LM25                             | $V_L^* = -3.1911 + 23.9395\theta_C + 4.6105(1 - e/D) - 2.0809(1 + \sin\alpha) + 4.4672Fr_P$ |
| LM26                             | $V_L^* = -3.1909 + 23.9395\theta_C + 4.6105(1 - e/D) - 2.0809(1 + \sin\alpha) + 4.4672Fr_P$ |
| LM27                             | $V_L^* = -3.1238 + 23.2431\theta_C + 4.6001(1 - e/D) - 2.0809(1 + \sin\alpha) + 4.4672Fr_P$ |
| LM28                             | $V_L^* = -3.1293 + 21.0132\theta_C + 4.9544(1 - e/D) - 2.0809(1 + \sin\alpha) + 4.4672Fr_P$ |
| LM29                             | $V_L^* = -3.1079 + 21.0132\theta_C + 4.9544(1 - e/D) - 2.0809(1 + \sin\alpha) + 4.4672Fr_P$ |

**Table 11.** List of bias terms given by M5MT#4 in  $V_L^*$  estimates.

| Equation | $V_L^*$ | Equation | $V_L^*$ |
|----------|---------|----------|---------|
| LM1      | 0.0512  | LM16     | 0.6642  |
| LM2      | 0.2208  | LM17     | 1.2584  |
| LM3      | 0.4659  | LM18     | 2.8566  |
| LM4      | 0.3353  | LM19     | 3.4517  |
| LM5      | 0.2568  | LM20     | 3.4984  |
| LM6      | 0.5315  | LM21     | 2.3073  |
| LM7      | 2.6256  | LM22     | 4.3984  |
| LM8      | 1.8337  | LM23     | 4.6448  |
| LM9      | 0.4239  | LM24     | 0.7799  |
| LM10     | 0.6153  | LM25     | 1.5781  |
| LM11     | 0.8379  | LM26     | 1.7987  |
| LM12     | 0.5844  | LM27     | 1.8874  |
| LM13     | 1.1963  | LM28     | 1.8927  |
| LM14     | 1.3669  | LM29     | 3.5284  |
| LM15     | 0.5064  | LM30     | 0       |

Generally, once M5MT was performed with ignorance of the pruning stage, the size of the tree structure increased. Although this issue can occasionally improve the accuracy level of predictions, overfitting can be imminent. For instance, in the performance of M5MT#3, the pruning phase was excluded during the training stage, resulting in an RMSE = 0.846 for the training stage and 0.554 for the testing stage. Additionally, in M5MT#4, simultaneous exclusion of pruning and smoothing stages caused the overfitting and high growth of the model tree. RMSE values for training and testing stages were equal to 0.846 and 0.554, respectively. M5MT#1 provided better predictions of the scouring propagation rate

with an RMSE = 0.547 and 0.599 for training and testing, respectively, when compared with M5MT#2 (RMSE = 0.6886 and 0.713). It is practically permissible for complicated system analysis to select a typical alternative of M5MT in the presence of pruning and smoothing phases. In this way, M5MT#1 was selected as the superior model to estimate  $V_L^*$  in this study.

#### 4. Discussion

##### 4.1. Statistical Measures

To measure the efficacy of ML models' performance for both the training and testing stages in relation with the estimation of scouring propagation rates around offshore pipelines, the index of agreement (IOA), RMSE, mean absolute error (MAE), and scatter index (SI) have been utilized. These statistical measures are defined as:

$$IOA = 1 - \frac{\sum_{i=1}^N \left( V_{L(Pre)}^{*(i)} - V_{L(Obs)}^{*(i)} \right)^2}{\sqrt{\sum_{i=1}^N \left( \left| V_{L(Pre)}^{*(i)} - \bar{V}_{L(Obs)}^* \right| + \left| V_{L(Obs)}^{*(i)} - \bar{V}_{L(Obs)}^* \right| \right)^2}} \tag{18}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N \left( V_{L(Pre)}^{*(i)} - V_{L(Obs)}^{*(i)} \right)^2}{N}} \tag{19}$$

$$MAE = \frac{\sum_{i=1}^N \left| V_{L(Pre)}^{*(i)} - V_{L(Obs)}^{*(i)} \right|}{N} \tag{20}$$

$$SI = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N \left[ \left( V_{L(Pre)}^{*(i)} - \bar{V}_{L(Pre)}^* \right) - \left( V_{L(Obs)}^{*(i)} - \bar{V}_{L(Obs)}^* \right) \right]^2}}{\bar{V}_{L(Obs)}^*} \tag{21}$$

where  $V_{L(Obs)}^*$ ,  $V_{L(Pre)}^*$ ,  $\bar{V}_L^*$ , and  $N$  are the observed, predicted, and average values of  $V_L^*$ , respectively, and  $N$  is the number of experimental works. The most ideal value of IOA is equal to 1, whereas the worst one is zero. In addition, the RMSE, MAE, and SI values are introduced as error functions, varying from 0 to  $+\infty$ .

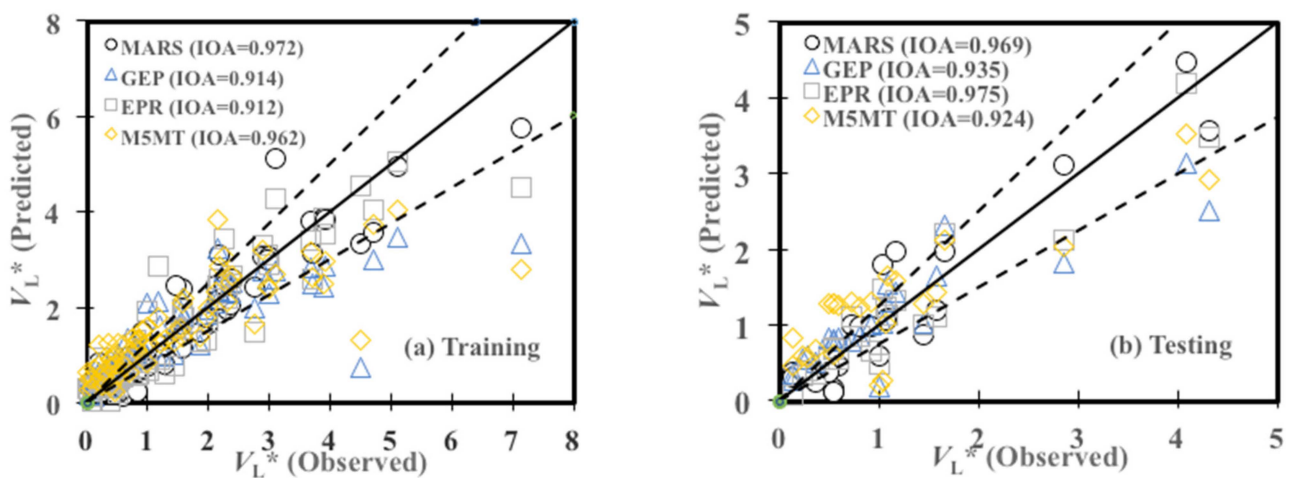
##### 4.2. Statistical Performance of ML Models

Table 12 demonstrates the performance of ML models in the estimation of scouring propagation rates ( $V_L^*$ ) in training (calibration) and testing (validation) stages. In the training stage, as seen in Table 12, the MARS expression (Equation (14)) with an RSME of 0.474 and MAE of 1.619 gave the most promising outperformance, followed by EPR (RMSE = 0.557 and MAE = 2.754), GEP (RSME = 0.836 and MAE = 3.514), and M5MT (RSME = 0.847 and MAE = 5.347). Additionally, values of IOA and SI proved the superiority of the MARS model (IOA = 0.972 and SI = 0.322) over the EPR (IOA = 0.912 and SI = 0.378), GEP (IOA = 0.914 and SI = 0.567), and M5MT (IOA = 0.962 and SI = 0.574) techniques. According to IOA, RMSE, and SI, the models GEP and M5MT provided rather the same performance in the prediction of  $V_L^*$  for the training stage. Statistical measures of testing stages indicated that EPR (see Table 5 and Model#5) provided the most successful level of performance (RMSE = 0.342 and MAE = 0.300) compared to MARS (RMSE = 0.379 and MAE = 0.449), GEP (RMSE = 0.556 and MAE = 0.490), and M5MT (RMSE = 0.599 and MAE = 0.922). Additionally, values of IOA and SI indicated that EPR expression with the natural logarithmic inner function had the most remarkable potential of estimating  $V_L^*$  in comparison with MARS (IOA = 0.969 and SI = 0.330), GEP (IOA = 0.935 and SI = 0.483), and M5MT (IOA = 0.924 and SI = 0.507).

**Table 12.** Statistical performances of the ML models considered in this study.

| ML Models | Training Stage |       |       |       |
|-----------|----------------|-------|-------|-------|
|           | IOA            | RMSE  | MAE   | SI    |
| MARS      | 0.972          | 0.474 | 1.619 | 0.322 |
| GEP       | 0.914          | 0.836 | 3.514 | 0.567 |
| EPR       | 0.912          | 0.557 | 2.754 | 0.378 |
| M5MT      | 0.962          | 0.847 | 5.347 | 0.574 |
| ML Models | Testing Stage  |       |       |       |
|           | IOA            | RMSE  | MAE   | SI    |
| MARS      | 0.969          | 0.379 | 0.449 | 0.330 |
| GEP       | 0.935          | 0.556 | 0.490 | 0.483 |
| EPR       | 0.975          | 0.342 | 0.300 | 0.296 |
| M5MT      | 0.924          | 0.599 | 0.922 | 0.507 |

Figure 3 illustrates the qualitative performance of ML techniques for both the training and testing phases. Almost all data points of the training stage in Figure 3 were concentrated on the  $\pm 25\%$  range of acceptable error. M5MT and GEP techniques significantly indicated the underestimation of  $V_L^*$  for the observed values of 4.5 and 7, whereas moderate overestimation has been attained by MARS and GEP techniques. Additionally, the scattering of testing data points in Figure 3b illustrated that M5MT and GEP techniques yielded moderate underpredictions and overpredictions for an observed  $V_L^*$  less than 2, whereas, for  $V_L^* = 2.5\text{--}5$ , almost all data points given by M5MT and GEP indicated underpredictions in comparison with EPR and MARS models.



**Figure 3.** Performance of ML models for prediction of  $V_L^*$  in the (a) training and (b) testing stages.

#### 4.3. Comparisons between ML Models and Related Works Regarding Complexity

In the present part of the investigation, the results of the ML models have been compared with those obtained by previous investigations considering various issues of ML models, such as the complexity of the general structure, the accuracy level, and the typical usability of optimization models in improving ML models' performance. In the case of the convolutional structure of the ML models, EPR expression (see Table 5 and Model #5) had a more complex mathematical structure, including six algebraic terms and natural logarithmic inner function, compared with MARS expression (Equation (14)) and multilinear regression equations by M5MT. Applying the present setting parameters in the development of the EPR model along with MOGA performance was more successful in increasing the accuracy level of the  $V_L^*$  prediction than M5MT, with its simpler expressions and low executive time performance. Furthermore, Equation (14), given by MARS, included 15 sets of second-order polynomials along with the performance of 10 forward and backward stages, providing



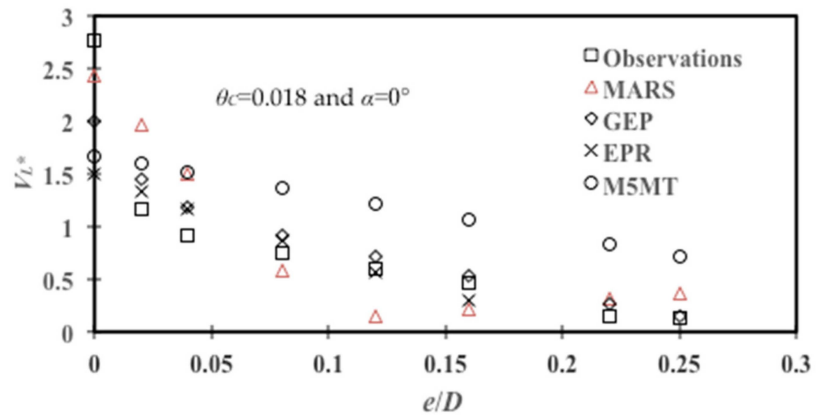
more complex expressions and higher executive time performance than the multilinear mathematical expressions developed by M5MT (see Table 7). In this case, the MARS expression had a promising application compared to all four alternatives of the M5MT models (including bias terms and linear expressions). In the GEP model, three typical inner functions (i.e., Tanh, Exp, Atan) with three algebraic terms could not lead to higher complex mathematical expressions than EPR expressions.

Since there is no related research work that has studied the application of ML models in the estimation of scouring propagation rates due to currents, the present results were only compared with previous investigations in terms of ML models' complexity. All the related works were tested on the scouring propagation data due to regular waves and live-bed sediment conditions. Najafzadeh and Saberi-Movahed [15] utilized the GEP model in the structure of the GMDH model to promisingly predict the scouring propagation rates around offshore pipelines due to waves in comparison with GMDH and GEP techniques. One of the main findings of the present study is that it is consistent with the results obtained by Najafzadeh and Saberi-Movahed [15]. Moreover, an increase in the complexity of ML models causes an increase in the precision level of GMDH (or other ML models such as EPR and GEP models). For instance, Table 5 indicates that adding logarithmic terms to the EPR expression led to more accurate results, increasing performance from Model #1 with a MSE = 0.582 to Model #5 with a MSE = 0.341. In another related research work, Ehteram et al. [16] used three optimization algorithms (i.e., CBA, WA, and PSO) to improve an ANN structure for the estimation of the scouring propagation rates. All three ANN models were less complex than the EPR expression obtained by the present study (see Table 5), and additionally, the ANN models were modeled only for wave conditions that were different from the present study. The results of this study showed that EPR, with its high degree of complexity, provided more accurate results than ANN-CBO (MAE = 0.721), ANN-WA (MAE = 0.745) and ANN-PSO (MAE = 0.814). Nevertheless, these ANN models have generally faster performance in predicting the scouring propagation rates, but the potential of MOGA in the selection of algebraic terms and exponents of variables (interactions among variables) would increase the accuracy level of ML models in this case. In Najafzadeh and Oliveto's [17] research, EPR expressions developed by the natural logarithmic inner function provided higher complexity and accuracy than the GEP model.

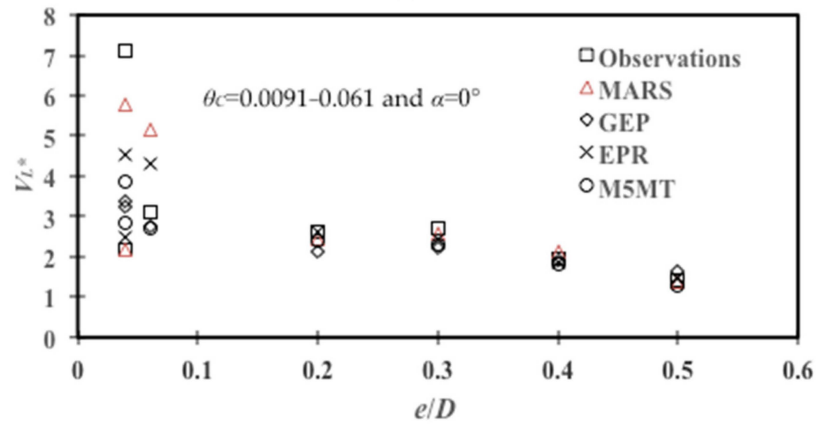
#### 4.4. Effects of the Pipeline Embedment Depth

The variation of  $V_L^*$  values versus  $e/D$  ratios is illustrated in Figure 4. These physical behaviors were studied in four levels of  $\theta_C$  and  $\alpha$  values, as seen in Figure 4a–d. For  $\theta_C = 0.018$  and  $\alpha = 0^\circ$ , Figure 4a indicates that an increase in  $e/D$  value leads to a decrease in  $V_L^*$ . All the ML models follow the decreasing trend, consistent with observed values. For instance, MARS demonstrated that  $V_L^*$  declined from 1.970 in  $e/D = 0.02$  to 0.577 in  $e/D = 0.08$ . Additionally, for the state of  $\theta_C = 0.0091$ – $0.061$  and  $\alpha = 0^\circ$ , Figure 4b illustrated a downward trend between  $V_L^*$  values and  $e/D$  ratios. In Figure 4c, the variation of  $V_L^*$  values versus  $e/D$  ratios was provided in different viewpoints. Generally, experimental observations depicted a downward trend from  $V_L^* = 2.453$  in  $e/D = 0.1$  to  $V_L^* = 1.398$  in  $e/D = 0.3$ . Afterward,  $V_L^*$  variations remained rather constant in  $e/D = 0.4$ , then had a decreasing trend up to  $e/D = 0.5$ . Figure 4c illustrated that all ML models follow the physical behavior of  $V_L^*$  versus  $e/D$  given by experimental observations for  $e/D = 0.1$ – $0.4$ , whereas the M5MT and EPR models could not follow the downward variation of  $V_L^*$  values between  $e/D = 0.4$ – $0.5$ , indicating relatively significant overpredictions. Similar to that shown in Figure 4c, MARS and EPR expressions could successfully simulate physical behaviors of  $V_L^*$  versus  $e/D$  values in  $e/D = 0.1$ – $0.2$ . As seen in Figure 4d, the M5MT and MARS models have gone through upward trends at  $e/D = 0.3$ , illustrating overpredictions. The quantitative results of ML models for different ranges of  $\theta_C$  and  $\alpha$  parameters are presented in Table 13. As seen in Table 13, the MARS technique provided the most promising performance (RMSE = 0.197 for  $\theta_C = 0.081$ – $0.104$  and  $\alpha = 15^\circ$ ) for all ranges, except  $\theta_C = 0.018$  and  $\alpha = 0^\circ$ ,

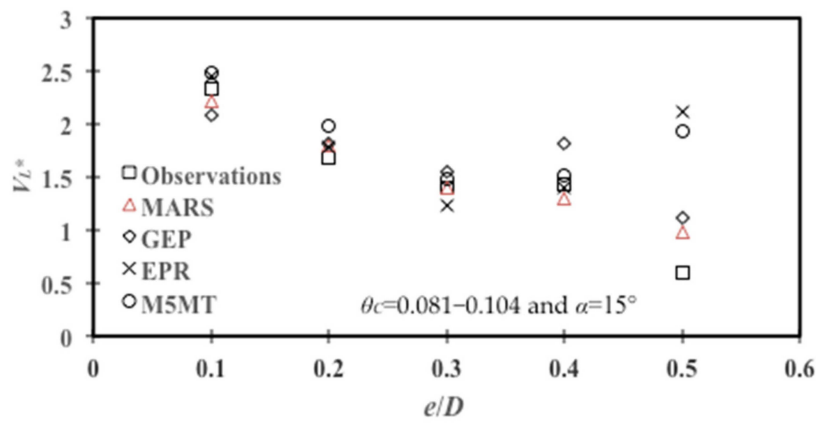
whereas M5MT indicated the lowest accuracy in the prediction of  $V_L^*$  for all ranges, except  $\theta_c = 0.046\text{--}0.104$  and  $\alpha = 30\text{--}45^\circ$ , in which GEP expression had the worst performance.



(a)

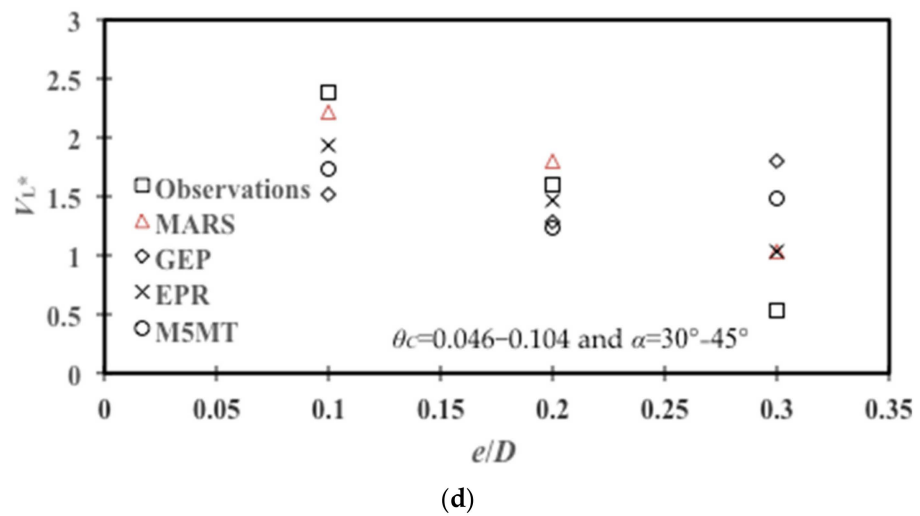


(b)



(c)

Figure 4. Cont.



**Figure 4.** Variations of dimensionless  $V_L^*$  against various  $e/D$  values at (a)  $\theta_C = 0.018$  and  $\alpha = 0^\circ$ , (b)  $\theta_C = 0.091\text{--}0.061$  and  $\alpha = 0^\circ$ , (c)  $\theta_C = 0.081\text{--}0.104$  and  $\alpha = 15^\circ$ , and (d)  $\theta_C = 0.046\text{--}0.104$  and  $\alpha = 30\text{--}45^\circ$ .

**Table 13.** RMSE values for the ML models as the various dimensionless parameters vary.

| ML Models | Variation of $e/D$  |  |   |   |
|-----------|---|--|---|---|
|           | $\theta_C = 0.018$ and $\alpha = 0^\circ$                 | $\theta_C = 0.0091\text{--}0.061$ and $\alpha = 0^\circ$ | $\theta_C = 0.081\text{--}0.104$ and $\alpha = 15^\circ$            | $\theta_C = 0.046\text{--}0.104$ and $\alpha = 30\text{--}45^\circ$ |
| MARS      | 0.431   | 0.933  | 0.197   | 0.331   |
| GEP       | 0.315   | 1.514  | 0.323   | 0.910   |
| EPR       | 0.481   | 1.100  | 0.682   | 0.398   |
| M5MT      | 0.681   | 1.767  | 0.618   | 0.698   |
| ML Models | Variation of $\theta_C$                                   |  |   |   |
|           | $e/D = 0.1$ and $\alpha = 0^\circ$                        | $e/D = 0.1\text{--}0.35$ and $\alpha = 0^\circ$          | $e/D = 0.1$ and $\alpha = 15^\circ$                                 | $e/D = 0.2$ and $\alpha = 15^\circ$                                 |
| MARS      | 0.223   | 0.183  | 0.746   | 0.191   |
| GEP       | 0.308   | 0.630  | 0.880   | 0.258   |
| EPR       | 0.185   | 0.212  | 0.490   | 0.251   |
| M5MT      | 0.681   | 0.573  | 0.690   | 0.349   |
| ML Models | Variation of $\theta_C$                                   |  |   |   |
|           | $e/D = 0.3$ and $\alpha = 15^\circ$                       | $e/D = 0.4$ and $\alpha = 15^\circ$                      | $e/D = 0.1$ and $\alpha = 30^\circ$                                 | $e/D = 0.2$ and $\alpha = 30^\circ$                                 |
| MARS      | 0.165   | 0.286  | 0.542   | 0.439   |
| GEP       | 0.248   | 0.462  | 0.532   | 1.075   |
| EPR       | 0.368   | 0.347  | 0.278   | 0.639   |
| M5MT      | 0.218   | 0.249  | 0.502   | 0.835   |
| ML Models | Variation of $Fr_p$                                       |  |   |   |
|           | $\theta_C = 0.0091\text{--}0.1045$ and $\alpha = 0^\circ$ | $\theta_C = 0.046\text{--}0.104$ and $\alpha = 15^\circ$ | $\theta_C = 0.046\text{--}0.104$ and $\alpha = 30\text{--}45^\circ$ |   |
| MARS      | 0.173   | 0.235  | 0.359   |   |
| GEP       | 1.063   | 0.376  | 0.993   |   |
| EPR       | 0.356   | 0.872  | 0.412   |   |
| M5MT      | 1.315   | 0.778  | 0.762   |   |

#### 4.5. Effects of the Shields Parameter

Figure 5 demonstrates the variation of  $V_L^*$  values versus the Shields parameter  $\theta_C$ . As seen in Figure 5a, all ML models provided overall increasing trends for  $e/D = 0.1$  and  $\alpha = 0^\circ$ .

In the case of  $e/D = 0.01\text{--}0.035$  and  $\alpha = 0^\circ$ , Figure 5b shows that all ML models detect physical behaviors of  $V_L^*$  versus  $\theta_C$  parameter in an upward way for  $\theta_C = 0.021\text{--}0.054$ ; then,  $V_L^*$  decreases in  $\theta_C = 0.064$ . Additionally, the M5MT and GEP models produced remarkable underpredictions for  $\theta_C = 0.054$ . Figure 5c depicts the variation of  $V_L^*$  values versus  $\theta_C$  for  $e/D = 0.1$  and  $\alpha = 15^\circ$ . The results of the ML models indicated that the present variation had an increasing trend. As seen in Figure 5c, for  $\theta_C = 0.104$ , ML models exhibit a certain underprediction. Similar to Figure 5c, Figure 5d–f show an increasing trend for  $e/D = 0.2\text{--}0.4$  and  $\alpha = 15^\circ$ . From Figure 5g,h, for  $e/D = 0.1\text{--}0.2$  and  $\alpha = 30^\circ$ , ML models generally provided an increasing trend for the variation of  $V_L^*$  versus the  $\theta_C$  parameter. For  $\theta_C = 0.104$ , all ML models underpredicted  $V_L^*$  for  $e/D = 0.1$  and  $\alpha = 30^\circ$  (see Figure 5g) whereas GEP and M5MT had underpredictions of  $V_L^*$  for  $e/D = 0.2$  and  $\alpha = 30^\circ$  in Figure 5h. Overall, the predicted values of  $V_L^*$  given by ML models were in good agreement with experimental observations. To allow quantitative comparisons of the ML models' performance in the four ranges of  $e/D$  and  $\alpha$ , the results of the RMSE are presented in Table 13. Table 13 indicates that EPR expression had the most successful performance for  $e/D = 0.1$  and  $\alpha = 0^\circ$  with an RMSE of 0.185,  $e/D = 0.1$  and  $\alpha = 15^\circ$  with an RMSE of 0.490, and  $e/D = 0.1$  and  $\alpha = 30^\circ$  with an RMSE of 0.278. Additionally, the MARS model had the best results for ranges  $e/D = 0.1\text{--}0.35$  and  $\alpha = 0^\circ$  (RMSE = 0.183),  $e/D = 0.2$  and  $\alpha = 15^\circ$  (RMSE = 0.191),  $e/D = 0.3$  and  $\alpha = 15^\circ$  (RMSE = 0.165),  $e/D = 0.2$  and  $\alpha = 30^\circ$  (RMSE = 0.439). Moreover, Table 13 shows that GEP had the worst performance for  $e/D = 0.2$  and  $\alpha = 30^\circ$  (RMSE = 1.075), as well as in case of  $e/D = 0.1$  and  $\alpha = 15^\circ$ , in which the GEP model still had the lowest level of accuracy (RMSE = 0.880).

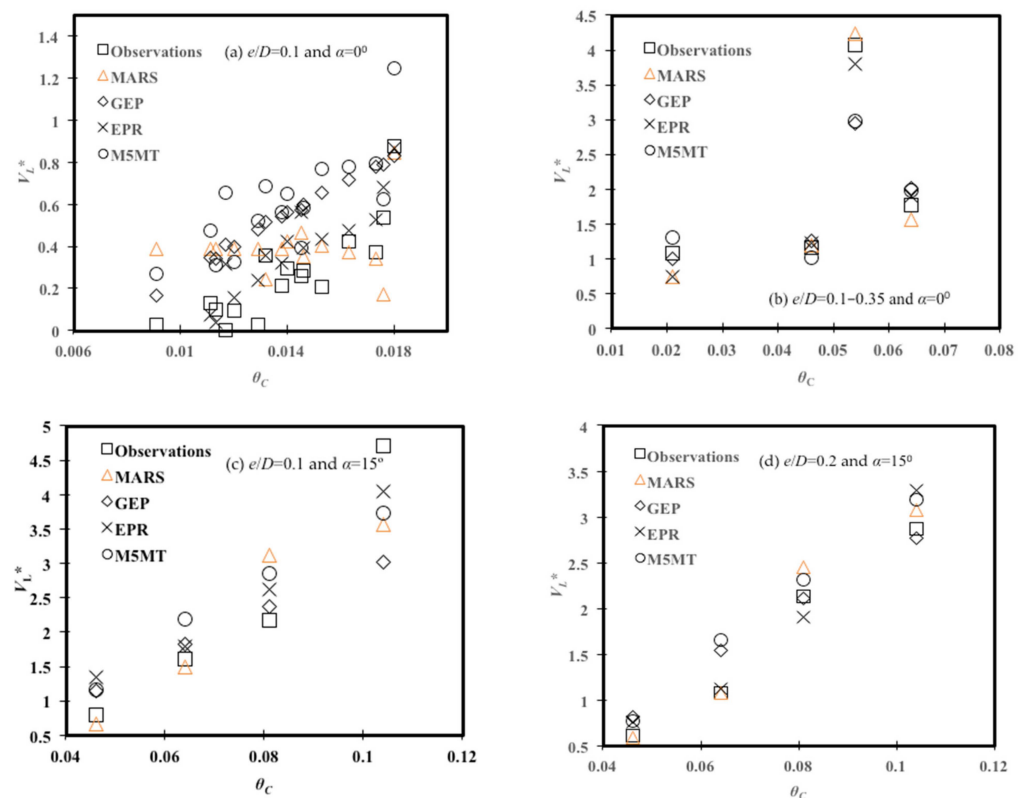
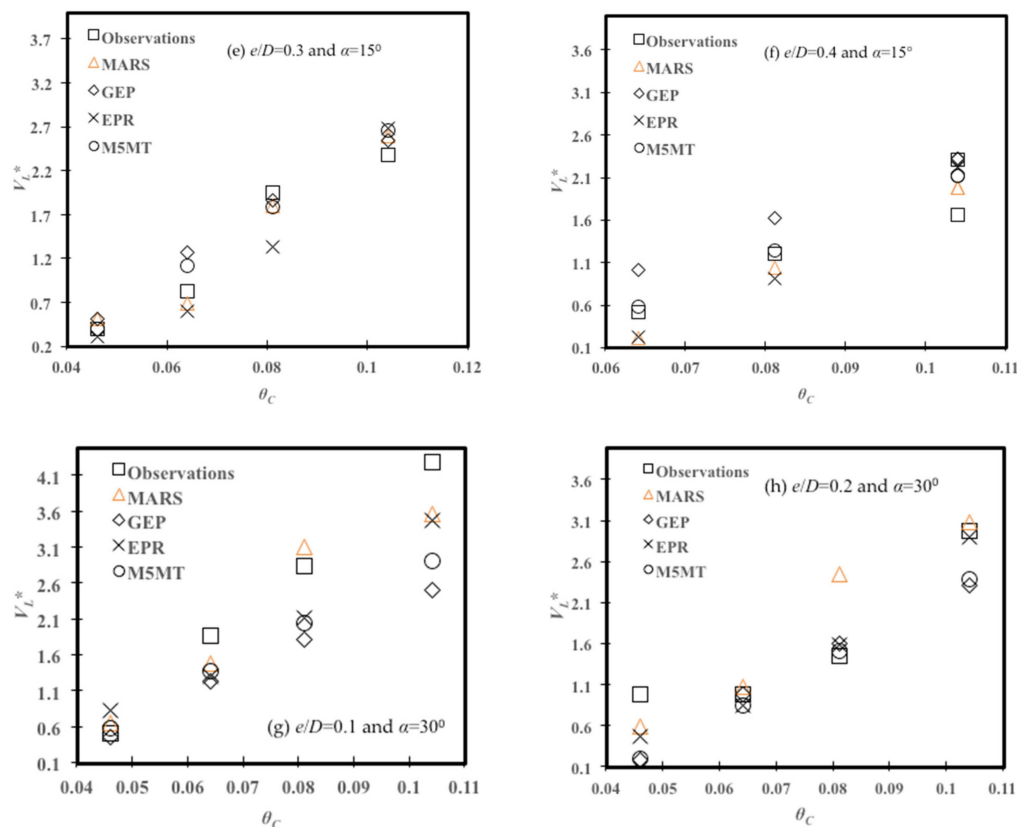


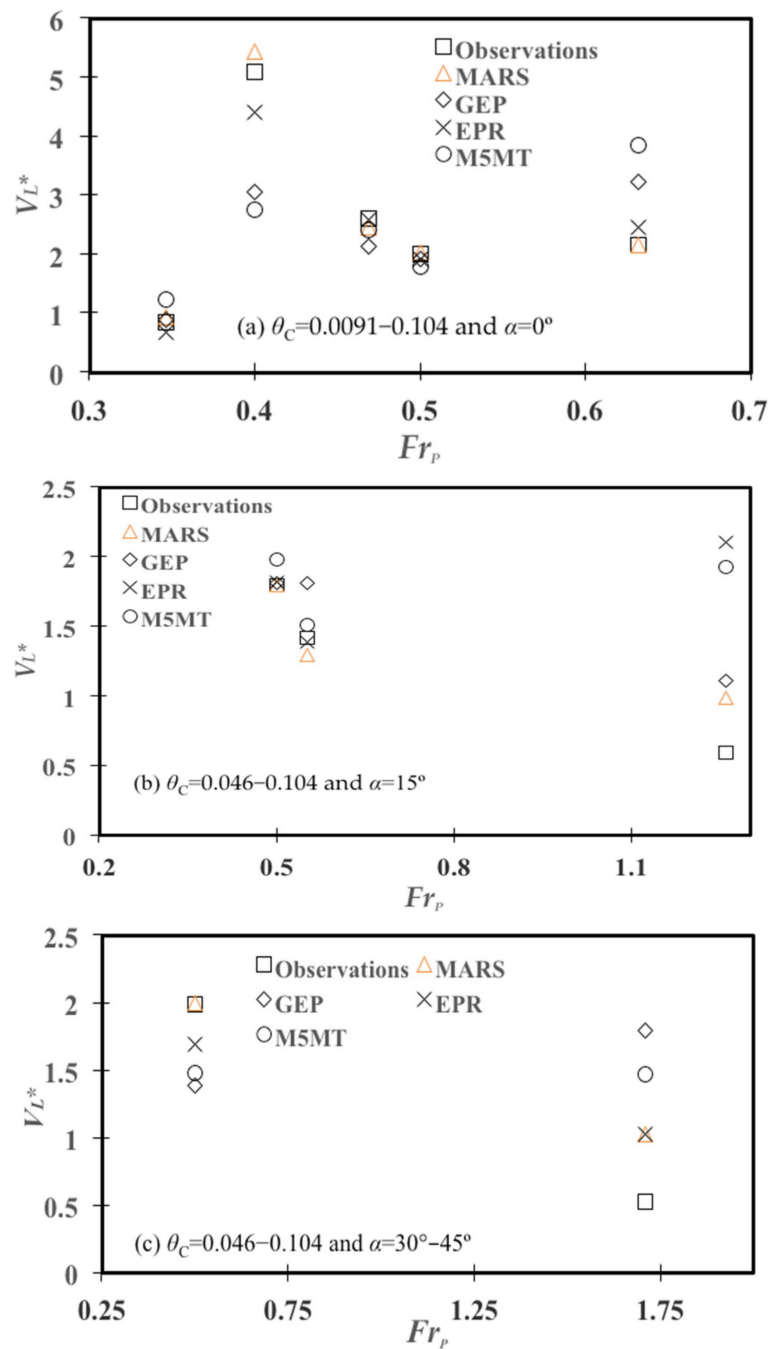
Figure 5. Cont.



**Figure 5.** Variations of dimensionless  $V_L^*$  against various  $\theta_C$  for (a)  $e/D = 0.1$  and  $\alpha = 0^\circ$ , (b)  $e/D = 0.1–0.35$  and  $\alpha = 0^\circ$ , (c)  $e/D = 0.1$  and  $\alpha = 15^\circ$ , (d)  $e/D = 0.2$  and  $\alpha = 15^\circ$ , (e)  $e/D = 0.3$  and  $\alpha = 15^\circ$ , (f)  $e/D = 0.4$  and  $\alpha = 15^\circ$ , (g)  $e/D = 0.1$  and  $\alpha = 30^\circ$ , and (h)  $e/D = 0.2$  and  $\alpha = 30^\circ$ .

4.6. Effects of the Approach Flow Froude Number

Figure 6 shows the variations of  $V_L^*$  versus  $Fr_p$  for three ranges of  $\theta_C$  and  $\alpha$ . From Figure 6a, the performance of the ML models was in good agreement with experimental observations for  $Fr_p = 0.34, 0.47,$  and  $0.5$ . Overall, the qualitative results of ML models indicated that  $V_L^*$  variations had an increasing trend for  $Fr_p = 0.346–0.4$ ; then, a downward trend was obtained for  $Fr_p = 0.4–0.5$ . Besides, values of  $V_L^*$  given by the ML techniques increased at  $Fr_p = 0.63$ . As seen in Figure 6a, all the ML models, except MARS expression, had relatively remarkable underprediction for  $Fr_p = 0.4$ , whereas M5MT and GEP techniques overpredicted  $V_L^*$  values at  $Fr_p = 0.63$ . In Figure 6b, experimental observations indicated that the variation of  $V_L^*$  values had a declining trend for  $Fr_p = 0.5–1.26$ , whereas EPR and M5MT techniques overpredicted significantly compared to GEP and MARS expressions. As illustrated in Figure 6c, MARS and EPR models indicated a downward trend for  $Fr_p = 0.5–1.71$ , whereas GEP and M5MT overpredicted significantly. In regard to the quantitative performance of ML models, the MARS model had the most successful evaluation in the prediction of  $V_L^*$  for all ranges  $\theta_C = 0.0091–0.1045$  and  $\alpha = 0^\circ$  (RMSE = 0.173),  $\theta_C = 0.046–0.104$  and  $\alpha = 15^\circ$  (RMSE = 0.235), and  $\theta_C = 0.046–0.104$  and  $\alpha = 30–45^\circ$  (RMSE = 0.359). In contrast, M5MT, EPR, and GEP yielded the lowest level of accuracy in  $\theta_C = 0.0091–0.1045$  and  $\alpha = 0^\circ$  (RMSE = 1.315),  $\theta_C = 0.046–0.104$  and  $\alpha = 15^\circ$  (RMSE = 0.872), and  $\theta_C = 0.046–0.104$  and  $\alpha = 30–45^\circ$  (RMSE = 0.993), respectively (Table 13).



**Figure 6.** Variations of dimensionless  $V_L^*$  against  $Fr_p$  for (a)  $\theta_C = 0.0091-0.1045$  and  $\alpha = 0^\circ$ , (b)  $\theta_C = 0.046-0.104$  and  $\alpha = 15^\circ$ , and (c)  $\theta_C = 0.046-0.104$  and  $\alpha = 30-45^\circ$ .

**5. Conclusions**

In the present research, various ML models based on evolutionary computing and classification concepts have been utilized to predict the scouring propagation rate around offshore pipelines exposed to currents. Effective dimensionless parameters (i.e.,  $1 - e/D$ ,  $1 + \sin\alpha$ ,  $\theta_C$ ,  $Fr_p$ ) obtained from dimensional analysis of the scouring tests were directly incorporated into the presented equations through the performance of ML models. The main findings of this research can be summarized as follows.

ML models provided explicit formulas with promising performance for the estimation of the scouring propagation rate. EPR expression included a mathematical structure with the highest degree of nonlinearity, followed by Equation (14) (given by the second-order polynomial MARS model) and the multilinear regression equation by M5MT. In

the GEP model, a kind of inner function played a key role in increasing the precision level of performance (Equation (9)) through an optimal evolutionary technique. The predictive equations resulting from the ML models are somewhat complex and have several fitting parameters, which are not so numerous when taking into account the geometrical and hydraulic conditions (even including live-bed and clear-water approaching flows) considered in this study. However, these equations have the considerable advantage of fitting the experimental data with performances that are significantly higher than the currently available literature formulas. Their structure is somewhat complex and inhibits an immediate identification of the role performed by each governing parameter (though a sensitivity analysis could be a replacement in this regard), but they could be easily implemented in numerical codes. Moreover, according to Table 1, their ranges of application are:  $0 \leq e/D \leq 0.5$ ,  $0^\circ \leq \alpha \leq 45^\circ$ ,  $0.22 \leq Fr_p \leq 0.63$ , and  $0.01 \leq \theta_C \leq 0.10$ .

The ML models considered in this study definitely perform better than the current empirical models. However, their performance differs from one model to another depending on the structure/algorithm of the model itself. This study also remarks on these differences (highlighting which model would perform better) in a physical context characterized by a limited number of data (95 experimental observations) despite the complexity of the phenomena. Statistical measures of ML models indicated that MARS could yield 15 BFs (including linear and polynomial equations) and the best performance in the training states whereas, for the testing stage, EPR expression with natural logarithmic functions resulted in the most level of precision for the prediction of  $V_L^*$ .

Quantitative and qualitative representations for variations of  $V_L^*$  values versus  $e/D$  were provided to obtain physical behaviors with variations in  $\theta_C$  and  $\alpha$  values. Generally, MARS and GEP expressions simulated well the variations of  $V_L^*$  values versus  $e/D$  ratios through a downward trend compared to EPR and M5MT techniques. These downward variations were in good agreement with experimental variations for different levels of  $\theta_C$  and  $\alpha$  (for instance,  $\theta_C = 0.018$  and  $\alpha = 0^\circ$ ;  $\theta_C = 0.0091$ – $0.061$  and  $\alpha = 0^\circ$ ).

A parametric study of  $V_L^*$  variations versus  $\theta_C$  values was carried out in various levels of  $e/D$  and  $\alpha$ . All ML models provided increasing trends for levels of  $e/D$  and  $\alpha$ . Generally, predictions of scouring propagation rates (given by ML models) were in good harmony with experiments by Hansen et al. [18], Cheng et al. [6], and Wu and Chiew [7].

Driving physical variations of  $V_L^*$  values versus the pipeline Froude number ( $Fr_p$ ) values were conceptualized for three levels of  $\theta_C$  and  $\alpha$  values. Although MARS and EPR expressions demonstrated a downward trend for some values of  $Fr_p$ , the performance of the ML models generally could follow decreasing trends for  $\theta_C = 0.0091$ – $0.104$  and  $\alpha = 0^\circ$ ,  $\theta_C = 0.046$ – $0.104$  and  $\alpha = 15^\circ$ , and  $\theta_C = 0.046$ – $0.104$  and  $\alpha = 30$ – $45^\circ$ , probably due to live-bed conditions.

ML model-performed equations exhibited physical consistency with experimental investigations, so they could result in reliable estimations of scouring propagation rates, which are utilized to consider the practical design of offshore pipelines while focusing on preventative measures of erosion and scouring. The analysis of data was carried out considering dimensionless governing parameters of great impact in sediment transport phenomena (e.g., the pipeline Froude number and the Shields parameter). This approach allows the extension of the proposed predictive equations to the field because the experimental data used in this study would appear free from scale effects. However, more experimental observations would be useful and desirable, especially those looking at the unexplored ranges highlighted in the histograms in Figure 2. The combined action of current and waves would need more attention, as would the collection of field data, which are not currently available to the authors' knowledge.

**Author Contributions:** Conceptualization, M.N. and G.O.; methodology, M.N.; software, M.N.; validation, M.N. and G.O.; formal analysis, M.N. and G.O.; investigation, M.N. and G.O.; data curation, M.N. and G.O.; writing—original draft preparation, M.N. and G.O.; writing—review and editing, M.N. and G.O.; visualization, M.N. and G.O.; supervision, M.N. and G.O. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

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

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** This research has been supported by Graduate University of Advanced Technology (Kerman-Iran) under grant number of 2594.

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

## References

1. Bai, Y.; Bai, Q. *Subsea Engineering Handbook*; Gulf Professional Publishing: Burlington, MA, USA, 2010.
2. Barrette, P. Offshore pipeline protection against seabed gouging by ice: An overview. *Cold Reg. Sci. Technol.* **2011**, *69*, 3–20. [[CrossRef](#)]
3. Palmer, A.C.; Been, K. Pipeline geohazards for Arctic conditions. In *Deepwater Foundations and Pipeline Geomechanics*; McCarron, W.O., Ed.; J. Ross Publishing: Fort Lauderdale, FL, USA, 2011; pp. 171–178.
4. Palmer, A.C.; King, R.A. *Subsea Pipeline Engineering*, 2nd ed.; PennWell Corporation: Tulsa, OK, USA, 2008.
5. Ramakrishnan, T.V. *Offshore Engineering*; Gene-Tech Books: New Delhi, India, 2008.
6. Cheng, L.; Yeow, K.; Zhang, Z.; Teng, B. Three-dimensional scour below offshore pipelines in steady currents. *Coast. Eng.* **2009**, *56*, 577–590. [[CrossRef](#)]
7. Wu, Y.; Chiew, Y.-M. Three-dimensional scour at submarine pipelines. *J. Hydraul. Eng. ASCE* **2012**, *138*, 788–795. [[CrossRef](#)]
8. Wu, Y.; Chiew, Y.-M. Mechanics of three-dimensional pipeline scour in unidirectional steady current. *J. Pipeline Syst. Eng.* **2013**, *4*, 3–10. [[CrossRef](#)]
9. Cheng, L.; Yeow, K.; Zang, Z.; Li, F. 3D scour below pipelines under waves and combined waves and currents. *Coast. Eng.* **2014**, *83*, 137–149. [[CrossRef](#)]
10. Fredsøe, J.; Sumer, B.M.; Arnskov, M.M. Time scale for wave/current scour below pipelines. *Int. J. Offshore Polar Eng.* **1992**, *2*, 13–17.
11. Sumer, B.M.; Fredsøe, J. Scour around pipelines in combined waves and current. In Proceedings of the 15th Conference on Offshore Mechanics and Arctic Engineering (OMAE), Florence, Italy, 16–20 June 1996.
12. Sumer, B.M.; Truelsen, C.; Sichmann, T.; Fredsøe, J. Onset of scour below pipelines and self-burial. *Coast. Eng.* **2001**, *42*, 313–335. [[CrossRef](#)]
13. Arya, A.K.; Shingan, B. Scour-mechanism, detection and mitigation for subsea pipeline integrity. *Int. J. Eng. Technol.* **2012**, *1*, 1–14.
14. Zhang, B.; Gong, R.; Wang, T.; Wang, Z. Causes and treatment measures of submarine pipeline free-spanning. *J. Mar. Sci. Eng.* **2020**, *8*, 329. [[CrossRef](#)]
15. Najafzadeh, M.; Saberi-Movahed, F. GMDH-GEP to predict free span expansion rates below pipelines under waves. *Mar. Georesour. Geotec.* **2019**, *37*, 375–392. [[CrossRef](#)]
16. Ehteram, M.; Ahmed, A.N.; Ling, L.; Fai, C.M.; Latif, S.D.; Afan, H.A.; Banadkooki, F.B.; El-Shafie, A. Pipeline scour rates prediction-based model utilizing a Multilayer Perceptron-Colliding Body algorithm. *Water* **2020**, *12*, 902. [[CrossRef](#)]
17. Najafzadeh, M.; Oliveto, G. Exploring 3D wave-induced scouring patterns around subsea pipelines with artificial intelligence techniques. *Appl. Sci.* **2021**, *11*, 3792. [[CrossRef](#)]
18. Hansen, E.A.; Staub, C.; Fredsøe, J.; Sumer, B.M. Time-development of scour induced free spans of pipelines. In Proceedings of the 10th Conference on Offshore Mechanics and Arctic Engineering (OMAE), Stavanger, Norway, 23–28 June 1991.
19. Ferreira, C. Gene expression programming: A new adaptive algorithm for solving problems. *Complex Syst.* **2001**, *13*, 87–129.
20. Ferreira, C. *Gene Expression Programming*, 2nd ed.; Springer: Berlin/Heidelberg, Germany, 2006.
21. Friedman, J.H. Multivariate adaptive regression splines. *Ann. Statist.* **1991**, *19*, 1–67. [[CrossRef](#)]
22. Giustolisi, O.; Savic, D.A. A symbolic data-driven technique based on evolutionary polynomial regression. *J. Hydroinf.* **2006**, *8*, 207–222. [[CrossRef](#)]
23. Giustolisi, O.; Savic, D.A. Advances in data-driven analyses and modelling using EPR-MOGA. *J. Hydroinf.* **2009**, *11*, 225–236. [[CrossRef](#)]
24. Giustolisi, O.; Doglioni, A.; Savic, D.A.; Webb, B.W. A multi-model approach to analysis of environmental analysis. *Environ. Modell. Softw.* **2007**, *22*, 674–682. [[CrossRef](#)]
25. Savic, D.A.; Giustolisi, O.; Berardi, L.; Shepherd, W.; Djordjevic, S.; Saul, A. Modelling sewer failure by evolutionary computing. In *Proceedings of the Institution of Civil Engineers-Water Management*; Thomas Telford Ltd: London, UK, 2006; Volume 159, pp. 111–118. [[CrossRef](#)]
26. Berardi, L.; Giustolisi, O.; Kapelan, Z.; Savic, D.A. Development of pipe deterioration models for water distribution systems using EPR. *J. Hydroinf.* **2008**, *10*, 113–126. [[CrossRef](#)]



27. Quinlan, J.R. Learning with continuous classes. In Proceedings of the Australian Joint Conference on Artificial Intelligence (AI'92), Hobart, Australia, 16–18 November 1992.
28. Mokhtia, M.; Eftekhari, M.; Saberi-Movahed, F. Dual-manifold regularized regression models for feature selection based on hesitant fuzzy correlation. *Knowl Based Syst.* **2021**, *229*, 107308. [[CrossRef](#)]
29. Mokhtia, M.; Eftekhari, M.; Saberi-Movahed, F. Feature selection based on regularization of sparsity based regression models by hesitant fuzzy correlation. *Appl. Soft Comput.* **2020**, *91*, 106255. [[CrossRef](#)]