applied sciences

MDPI

Article

Exploring 3D Wave-Induced Scouring Patterns around Subsea Pipelines with Artificial Intelligence Techniques

Mohammad Najafzadeh [1,*] and Giuseppe Oliveto [2]

[1] Department of Water Engineering, Faculty of Civil and Surveying Engineering, Graduate University of Advanced Technology, Kerman 76315-117, Iran
[2] School of Engineering, University of Basilicata, Viale dell'Ateneo Lucano 10, I-85100 Potenza, Italy; giuseppe.oliveto@unibas.it
* Correspondence: moha.najafzadeh@gmail.com

Featured Application: The paper presents new equations for predicting 3D scouring patterns at subsea pipelines under wave-only conditions. These equations are more reliable than conventional approaches and exhibit a clear physical consistence.

Abstract: Subsea pipelines carry oil or natural gas over long distances of the seabed, but fluid leakage due to a failure of the pipeline can culminate in huge environmental disasters. Scouring process may take place beneath pipelines due to current and/or wave action, causing pipeline suspension and leading to the risk of pipeline failure. The resulting morphological variations of the seabed propagate not only below and normally to the pipeline but also along the pipeline itself. Therefore, 3D scouring patterns need to be considered. Mainly based on the experimental works at laboratory scale by Cheng and coworkers, in this study, Artificial Intelligent (AI) techniques are employed to present new equations for predicting three dimensional current- and wave-induced scour rates around subsea pipelines. These equations are given in terms of key dimensionless parameters, among which are the Shields' parameter, the Keulegan–Carpenter number, relative embedment depth, and wave/current angle of attach. Using various statistical benchmarks, the efficiency of AI-models-based regression equations is assessed. The proposed predictive models perform much better than the existing empirical equations from literature. Even more interestingly, they exhibit a clear physical consistence and allow for highlighting the relative importance of the key dimensionless variables governing the scouring patterns.

Keywords: Evolutionary Polynomial Regression (EPR); Gene-Expression Programming (GEP); Keulegan–Carpenter number; Model Tree (MT); Multivariate Adaptive Regression Splines (MARS); scouring; subsea pipeline

Citation: Najafzadeh, M.; Oliveto, G. Exploring 3D Wave-Induced Scouring Patterns around Subsea Pipelines with Artificial Intelligence Techniques. *Appl. Sci.* **2021**, *11*, 3792. https://doi.org/10.3390/app11093792

Academic Editors: Francesco Granata and Rudy Gargano

Received: 15 March 2021
Accepted: 14 April 2021
Published: 22 April 2021

1. Introduction

Seabed pipelines are imperative to transport oil and gas from offshore platforms. Oil leakage due to failure of pipeline causes environmental disasters, which influence economic circumstances of nations. The flow structure of vortex around pipeline leads to vibration and consequently metal fatigue, which is introduced as one of the most commonly reported factors of pipeline failure. The scouring process may take place beneath pipelines when these offshore structures encounter rigorous oceanic flow (i.e., currents or waves). Longitudinal extension of the scour hole beneath the pipeline is inextricably bound to the pipeline span leading to pipeline spanning. When the extension of free span has sufficient extension in the longitudinal direction, seabed pipelines may experience severe vibration due to the existence of vortex structures (Xie and Zhu [1]).

For almost half a century, the scouring process below pipelines has been drawing meticulous attention by ocean and costal engineers in terms of safety of offshore structures. The scour hole around a pipeline is characterized by three-dimensional development due to the existence of the piping mechanism and an uneven seabed after the pipeline operation. When the scour hole grows longitudinally around the pipeline, the structure of free spans is developed. Basically, fully developed free spans are the main reason that pipelines are put into a failure state. In this case, the failure of pipelines is due to the fact that they are more exposed to structural destructions corresponding to over-stressing circumstances. The earliest experimental investigations proved that the formation of three-dimensional free spans around pipelines are efficient to augment the instability level of the pipeline. In addition to this, free spans can lead to natural self-burial into the scour hole formed below the pipeline (Sumer and Fredsøe [2]). Figure 1 shows the structure of free spans in the extension state. The propagation of the scour rate was seen in two opposite directions along the pipeline. Three-dimensional growth of the free span depends on the conditions of flow and sediment motion and pipeline geometry.

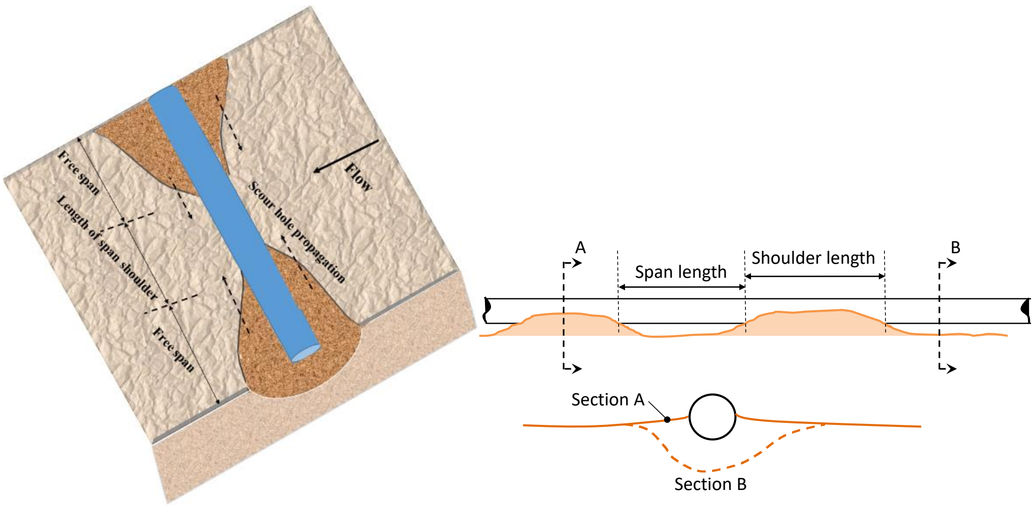

Figure 1. Schematic drawing of 3D scour hole propagation with free span expansion at a submarine pipeline (on the **left**); longitudinal and cross-section (i.e., Section A and Section B) views for a submarine pipeline prone to scour (on the **right**).

Figure 2 shows schematic diagrams of vortex structures around submarine pipelines under currents. Figure 2a illustrates three types of vortex structures. In the near vicinity of the pipeline, both vortices A and C move particles of bed sediments away from the footing area, even though movements of particles occur in opposite directions to each other. Vortex B also moves particles of bed sediment, but its acting area is constrained by Vortex C. Ultimately, a small opening is formed beneath the pipeline, which is introduced as the onset of unidirectional flow-induced scour. Figure 2b shows that as the water flows beneath the pipeline, the upstream vortex disappears, creating tunnel erosion gradually.

Figure 3 shows schematic diagrams of vortex structures around submarine pipelines under waves. As the full development of the scour hole is met, complicated flow structure of vortex shedding occurs behind the pipeline (Çevik and Yüksel [3]). As the flow structure has oscillatory pattern, a vortex system with wake intensity occurs on both sides of the pipeline. Figure 3 also shows the flow structure of lee-wake erosion, which takes place on both sides of the pipeline.

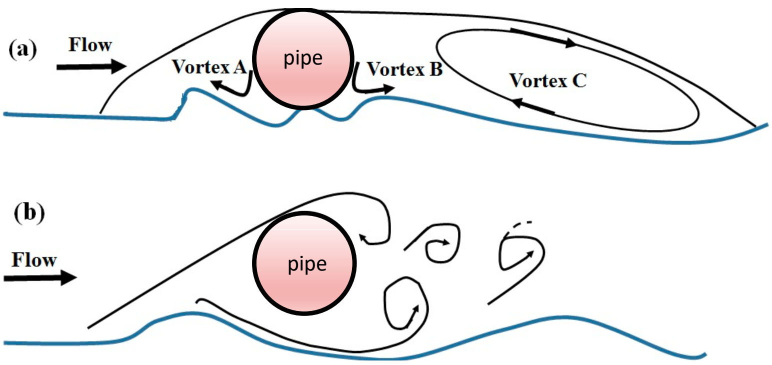

Figure 2. Schematic diagram of onset scouring process for submarine pipelines under currents. (**a**) Vortex A in front of the pipe, Vortex B in the corner downstream of the pipe, and the largest Vortex C behind the pipe; (**b**) vortex shedding that forms after the scour develops.

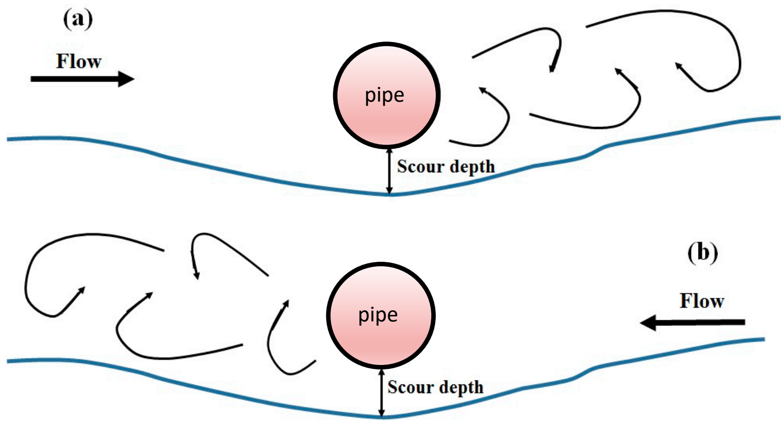

Figure 3. Schematic drawing of scouring vortices due to waves at a submarine pipeline. Wake vortexes downstream of the pipeline (**a**,**b**) due to oscillatory flow.

A large number of experimental studies have been carried out to better understand scouring mechanisms at submarine pipelines; the papers from [4–14] in the references section are some examples. Generally, all the experimental studies were carried out in three distinctive flow conditions: (1) Current, (2) regular wave, and (3) combined current and wave. Cheng et al. [6] investigated three-dimensional scour rates below a pipeline exposed to currents. They proposed an empirical equation for the prediction of longitudinal scour rates below the pipeline. Additionally, Cheng et al. [9] studied the scouring mechanism below a pipeline in two various flow conditions (i.e., combined wave and current, wave only). They proposed two regression-based equations for the prediction of the longitudinal scour rate in both flow conditions. In terms of the propagation of scour rates around pipelines, it can be inferred that there is limited knowledge on the scour hole development.

Previous attempts generally indicated that empirical equations are restricted to the range of experimental databases. Thus, these equations often do not have high potential of generalizing the prediction of scour rates in the three-dimensional space below pipelines.

In the case of experimental investigations, it may be expected to find comprehensive influences of various parameters on the scouring process below pipelines. An understanding of the free span propagation still requires in-depth investigations (Sumer and Fredsøe [2], Cheng et al. [6], Cheng et al. [9]).

With the advent of soft computing techniques, complicated mechanisms of real-life problems have become more understandable than before. In this way, there is a high demand for the presentation of robust mathematical models on the basis of Artificial Intelligent (AI) techniques in various fields of science. Over the last decade, a large number of

AI approaches have been employed prosperously to obtain an accurate estimation of the local scour depth below pipelines exposed to currents and regular waves. In fact, Artificial Neural Networks (ANNs) (Azamathulla and Zakaria [15], Etemad-Shahidi et al. [16]), Adaptive Neuro-Fuzzy Inference Systems (ANFIS) (Zanganeh et al. [17]), Support Vector Machine (SVM) (Parsaie et al. [18]), Classification and Regression Tree (CART) (Etemad-Shahidi et al. [16], Yasa and Etemad-Shahidi [19]), Linear Genetic Programming (LGP), Gene-Expression Programming (GEP) (Azamathulla and Yusoff [20], Najafzadeh and Sarkamaryan [21]), Genetic Programming (GP) (Azamathulla and Ghani [22]), Group Method of Data Handling (GMDH) (Najafzadeh et al. [23], Najafzadeh et al. [24]), Multivariate Adaptive Regression Splines (MARS) (Haghiabi [25]), and Evolutionary Polynomial Regression (EPR) (Najafzadeh and Sarkamaryan [21]) proved to be successful applications. In the case of scour rates prediction, there have been a few endeavors to estimate three-dimensional free spans via GMDH and ANN models below a pipeline exposed to regular waves (Najafzadeh and Saberi-Movahed [26], Ehteram et al. [27]).

However, few investigations have been conducted to estimate scour propagation rates at seabed pipelines by AI models. In fact, the experimental detection of three-dimensional scour rates is generally expensive due to the need of advanced sensors and other facilities. To the best of the authors' knowledge, powerful AI techniques such as MARS, GEP, and MT have not yet been applied in the prediction of scour propagation rates around seabed pipelines exposed to regular waves. As main merits, these AI techniques have potential in three main areas: (i) Presenting (non-linear) explicit equations for limited number of data; (ii) reducing the number of influential variables; and (iii) preserving the physical consistency of the generated models. The present paper moves in these directions, identifying new equations to predict 3D scour patterns at seabed pipelines depending on key dimensionless parameters. These equations would be more reliable than conventional approaches, as well as preserving physical consistency.

2. AI Predictive Techniques on Scouring below Seabed Pipeline: A Brief Review

This section provides a brief review on the literature studies in which scouring below seabed pipeline is addressed by AI techniques. The section is divided into two parts for the sake of clarity depending on whether scour is due to currents or waves.

2.1. Current-Induced-Pipeline Scour

In the case of pipeline scour due to currents, a large number of investigations was attempted to use various AI models along with preservation of scouring conceptions.

Azamathulla and Ghani [22] applied GP model to experimental datasets, which were extracted from two various flow and sediment conditions. Ultimately, they proposed a regression-based equation, which applies to both clear-water (CW) and live-bed (LB) conditions. However, two major drawbacks arise from their research work. In the first place, they should have separated the entire dataset into two basic sections depending on whether the scouring process was governed by clear-water or live-bed approach flow conditions. It is well known that these two regimes imply different scouring processes. Incidentally, Azamathulla and Ghani [22] did not consider the ratio e/D of the pipeline embedment depth e to the pipeline diameter D as a controlling variable in modeling live-bed conditions; and the reason is that e/D in the experiments made by Moncada and Aguirre [28], to which Azamathulla and Ghani [22] relate, was kept constant for live-bed conditions. The second drawback of the study by Azamathulla and Ghani [22] is that they considered the Reynolds number as the controlling variable of local scour depth, though approach flows were fully turbulent. These issues cause some concerns on the use of the equation developed by Azamathulla and Ghani [22] in practical applications. Azamathulla et al. [29] later employed the LGP technique to provide an empirical equation by still considering the same current-induced scour datasets. However, the equation given by LGP is also undermined by the same drawbacks as for Azamathulla and Ghani [22].

The same comments would apply to the study by Azamathulla and Zakaria [15], though they applied ANN models.

Zanganeh et al. [17] improved the ANFIS model by using Particle Swarm Optimization (PSO) to predict the local scour depth around seabed pipelines due to currents. However, the experimental datasets used in their study were released from a few studies with limited ranges for the governing variables. Furthermore, they did not always consider turbulent flows (i.e., flows independent from Reynolds number) during the development of the ANFIS-PSO models. Similarly, Azamathulla and Yusoff [20] proposed a regression-based formulation by means of GEP technique. It looks like also their research did not follow physical insights of scouring mechanisms in terms of sediment bed motion and turbulent flow circumstances. Najafzadeh et al. [23] applied the GMDH technique to predict the local scour for various sediment motion conditions of approach flows. They not only enhanced existing inaccuracies in the literature, but also covered fundamental misconceptions related to the physical insights of scouring processes in similar previous investigations. They concluded that the performance of GMDH was relatively more accurate than SVM technique.

Yasa and Etemad-Shahidi [19] presented the CART-model-based probabilistic framework for various levels of risk in the design of undersea pipelines. They propose a regression equation, based on experimental data, for the economic design of a pipeline due to scour depth. Haghiabi [25] used the MARS technique to provide a robust formulation for the estimation of scour depth below pipelines exposed to clear-water circumstances. He found that MARS model was more accurate than the ANN technique and empirical equations. Najafzadeh and Sarkamaryan [21] presented some regression-based equations by three robust AI models (i.e., EPR, GEP, and MT) for both clear-water and live-bed conditions. As a merit, they found that the performance of the AI models-based regression equations they considered had highly satisfying performance in comparison to empirical equations from the literature. Parsaie et al. [18] employed SVM as a kernel-based AI model. Even though the performance of SVM models indicated more accurate results than ANN and ANFIS approaches, the proposed SVM model was designed for both clear-water and live-bed conditions, though the scour mechanisms in these two conditions are completely different, as previously remarked.

In summary, the main shortcomings of the above studies lie in unifying clear-water and live-bed experimental data, though the related physical processes are intrinsically different. Moreover, several investigations consider the Reynolds number as an influential parameter on the scour depth even in the case of fully turbulent flow conditions. Indeed, regression equations given by previous AI investigations appear to be not physically consistent.

2.2. Wave-Induced-Pipeline Scour

Etemad-Shahidi et al. [16] used a new version of MT technique to predict wave-induced scour depth at underwater pipelines. They successfully extended MT to both clear-water and live-bed conditions. Najafzadeh et al. [24] later trained traditional GMDH algorithms by the Back Propagation (BP) technique. They found that the performance of the GMDH-BP model was to a higher degree of accuracy in comparison to the ANFIS model and the Multivariate Non-Linear Regression (MNLR) equation. Sharafati et al. [30] developed an efficient stochastic technique on the basis of MT to achieve reliable predictions of scour depth below pipelines exposed to waves. They concluded that their results attained highly precise levels in comparison with the deterministic models.

In recent years, investigations have been characterized by a more strongly scientific orientation. Najafzadeh and Saberi-Movahed [26] provided a new extension of GMDH by GEP to estimate three-dimensional scour rates. From their study, GMDH-GEP indicated relatively better performance in the prediction of three-dimensional free spans than GEP, GMDH, and empirical equations. Moreover, Ehteram et al. [27] applied new improvements of ANN, by Colliding Bodies' Optimization (CBO), to model the three-dimensional nature of scour below seabed pipelines. Ultimately, they concluded that the improvement of the application of CBO into ANN resulted in more accurate predictions than PSO and the Whale Algorithm (WA).

However, the chief limitations of the above investigations are the regression-based-equations, given by GMDH-GEP (Najafzadeh and Saberi-Movahed [26]), are their tree-like structures with remarkable complexity, although equations are sufficiently accurate. On the downside, the use of an improved ANN model by Ehteram et al. [27] acts as a black box, which has no potential of being consistent with experimental observations.

3. Analysis of Datasets: Dimensional Analysis and Experimental Tests

In this section, the main dimensionless parameters, which would control the three-dimensional scour processes at seabed pipelines, are first identified through the dimensional analysis, and then the experimental data considered in this study are described and discussed.

3.1. Dimensional Analysis

Three-dimensional scour rates depend on several factors, namely pipeline geometry, wave properties, embedded depth of pipeline compared to the initial level of seabed, and physical properties of seabed sediments (e.g., Cheng et al. [6], Wu and Chiew [8], Cheng et al. [9], Wu and Chiew [31]). Hence, the following functional relationship can be formulated:

$$\chi(V_H, V_R, V_L, e,\ D, U_w, T, d_{50}, \phi, \alpha, \rho, \rho_s, \mu, g) = 0 \tag{1}$$

in which V_H is the scour propagation velocity along the longitudinal (axial) direction of the pipeline; V_L and V_R are the scour propagation velocities at the left- and right-hand shoulders of the pipeline, respectively; e is the embedment depth; D is the pipe diameter; U_w is the wave orbital velocity at the seabed; d_{50} is the median grain size of the seabed sediment; T is the wave period; ϕ is the angle of repose of seabed sediments, α is the flow incident angle to the pipeline (angle of attack); ρ is the density of water; ρ_s is the density of the seabed sediment; μ is the dynamic viscosity of water; and g is the acceleration due to gravity.

Previous studies revealed that the implementation of various AI models for the prediction of scour rates below pipelines by considering dimensionless variables would lead to results with highly convincing performance in comparison to those where dimensional variables are used. What is more, the use of non-dimensional parameters can ameliorate scale effects from experimental data on the performance of AI techniques (e.g., Najafzadeh and Sarkamaryan [21], Haghiabi [25], Najafzadeh and Saberi-Movahed [26], Azamathulla et al. [29]). In this way, by using the Buckingham theorem, the non-dimensional analysis was performed. At first, ρ, D, and U_w were assigned as repeating variables, and then 11 non-dimensional parameters ($\prod_1, \prod_2, \prod_3, \ldots, \prod_{10}, \prod_{11}$) were obtained:

$$\chi_1\left(\frac{V_H}{U_w}, \frac{V_R}{U_w}, \frac{V_L}{U_w}, \frac{e}{D}, \frac{U_w T}{D}, \frac{d_{50}}{D}, \phi, \alpha, \frac{\rho_s}{\rho}, \frac{\rho U_w D}{\mu}, \frac{gD}{U_w^2}\right) = 0 \tag{2}$$

where $\prod_5$ and $\prod_{10}$ are the Kelugan–Carpenter (KC) and Reynolds numbers of the pipeline due to the regular wave (Re_w), respectively. According to Cheng et al. [9], some $\prod$-variables in Equation (2) could be rearranged and clustered as follows:

$$\Pi'_7 = tan\phi;\ \Pi'_8 = sin\alpha;\ \Pi'_9 = \frac{\rho_s}{\rho} - 1 \tag{3}$$

$$\Pi'_{11} = \frac{1}{\Pi_{11} \cdot \Pi'_9 \cdot \Pi_6} = \frac{U_w^2}{g[(\rho_s/\rho - 1)]d_{50}} \tag{4}$$

$$\Pi'_1 = \Pi'_7 \cdot \sqrt{\Pi'_{11} \cdot \left(\frac{\Pi_1}{\Pi_6}\right)^2} = \frac{V_H \cdot D \cdot tan\phi}{\sqrt{g\left[(\rho_s/\rho - 1)d_{50}^3\right]}} \tag{5}$$

$$\Pi_2' = \Pi_7' \cdot \sqrt{\Pi_{11}' \cdot \left(\frac{\Pi_2}{\Pi_6}\right)^2} = \frac{V_R \cdot D \cdot tan\phi}{\sqrt{g\left[(\rho_s/\rho - 1)d_{50}^3\right]}} \tag{6}$$

$$\Pi_3' = \Pi_7' \cdot \sqrt{\Pi_{11}' \cdot \left(\frac{\Pi_3}{\Pi_6}\right)^2} = \frac{V_L \cdot D \cdot tan\phi}{\sqrt{g\left[(\rho_s/\rho - 1)d_{50}^3\right]}} \tag{7}$$

in which Π_1', Π_2', and Π_3' can be indicated with V_H^*, V_R^*, and V_L^*, respectively, and Π_{11}' is the Shields' parameter θ_w, which plays a key role in scour propagation below pipelines.

Therefore, the functional relationship (2) can be reduced as:

$$\chi_2\left(V_H^*, V_R^*, V_L^*, \frac{e}{D}, KC, sin\alpha, Re_w, \theta_w\right) = 0 \tag{8}$$

In fact, Cheng et al. [9] found that the Reynolds number of pipeline did not influence the scouring process, and Re_w can be removed from Equation (8). Furthermore, the variables $sin\alpha$ and e/D can be finished as $1 + sin\alpha$ and $1 - e/D$ to avoid the mathematical structure of the scour equation failing for $\alpha = 0°$ or $e = 0$ (Cheng et al. [9]). Finally, Equation (8) can be written as:

$$\chi_3\left(V_H^*, V_R^*, V_L^*, 1 - \frac{e}{D}, KC, 1 + sin\alpha, Re_w, \theta_w\right) = 0 \tag{9}$$

The functional relationship (9) indicates that four dimensionless parameters would control the three dimensionless scour rates V_H^*, V_R^*, and V_L^*.

3.2. Experimental Data

The availability of experimental data on scouring process at subsea pipelines is very limited. In this study, as can be generally observed in a number of recent articles on the topic under consideration, the experimental datasets extracted from Cheng et al. [9] were collected to develop the selected AI approaches. The just-mentioned authors performed three-dimensional scour experiments in a wave flume 50 m long, 4 m wide, and 2.5 m deep. A sandpit 4 m long, 4 m wide, and 0.25 m deep was used as test section. The transitions from the original flume bed to the test section and from the test section back to the original flume bed were achieved through two 1:20 concrete slopes. The upstream end of the sandpit was 21.5 m from the flow inlet, and the downstream end was 14.5 m from the flow outlet. A pipeline with a smooth surface, 50 mm diameter, and 8 mm wall thickness was tested. The experiments were performed for four various flow attack angles (α) 0°, 15°, 30°, and 45°. Furthermore, d_{50}, ϕ, and ρ_s/ρ were 0.37 mm, 32°, and 2.7, respectively. Even though d_{50} was very small, the viscosity effect could not be significant at the interface flow-sediment bed. Particles of bed sediment were in motion during the experiments (live-bed conditions), and the corresponding Shields' parameter due to wave condition was typically either 0.18 or 0.30 (Cheng et al. [9]). A total of 125 tests were conducted, which included 60 wave-only tests and 65 tests under combined wave and current conditions. In the case of wave-only conditions, scour propagation was not observed in the 16 tests. The results of the other three tests were obtained in the state of onset of scour in multiple locations. In addition to this, in three tests, the scour propagation process was only observed on the upstream span shoulder, and these observations cannot be used to feed AI models because the scour propagation should be observed in three dimensions: Longitudinal direction of pipeline, and the left and right sides of pipeline (Cheng et al. [9]). In this way, there are 22 experiments where pieces of information are not applied for modeling the scour propagation in the wave-only conditions. Table 1 shows the ranges of experimental datasets. From 38 experiments, 75% (29 datasets) and 25% (9 datasets) were randomly selected to perform training (or calibration) and testing (or validation) tests for AI models, respectively. The three mentioned scour rates are output variables, while the other non-dimensional parameters are inputs feeding the selected AI models.

Table 1. Basic statistical properties of the variables under study for AI modeling.

Dimensional Variables	Minimum	Maximum	Average	Standard Deviation
e[mm]	5.0	20.0	11.97	5.445
U_w[m/s]	0.3	0.5	0.40	0.068
T[s]	1.5	2.0	1.79	0.193
V_H[mm/s]	1.6	6.5	3.32	1.289
V_R[mm/s]	1.4	5.7	3.32	1.215
V_L[mm/s]	1.5	6.1	3.32	1.210
Dimensionless Variables	**Minimum**	**Maximum**	**Average**	**Standard Deviation**
$sin\alpha$[-]	0.10	0.71	0.32	0.263
e/D[-]	0.10	0.2143	0.0893	0.0893
KC[-]	8.70	18.00	14.68	3.698
θ_w[-]	0.18	0.30	0.27	0.051
V_H^*[-]	1.27	5.16	2.64	1.025
V_R^*[-]	1.19	4.85	2.64	0.961
V_L^*[-]	1.11	4.53	2.64	0.965

Histograms for all the non-dimensional (independent) variables are depicted in Figure 4a–d. These histograms show the frequency distributions for the parameters governing the scouring processes, providing brief and effective details on the data properties. Incidentally, this analysis might turn out to be useful in further laboratory investigations.

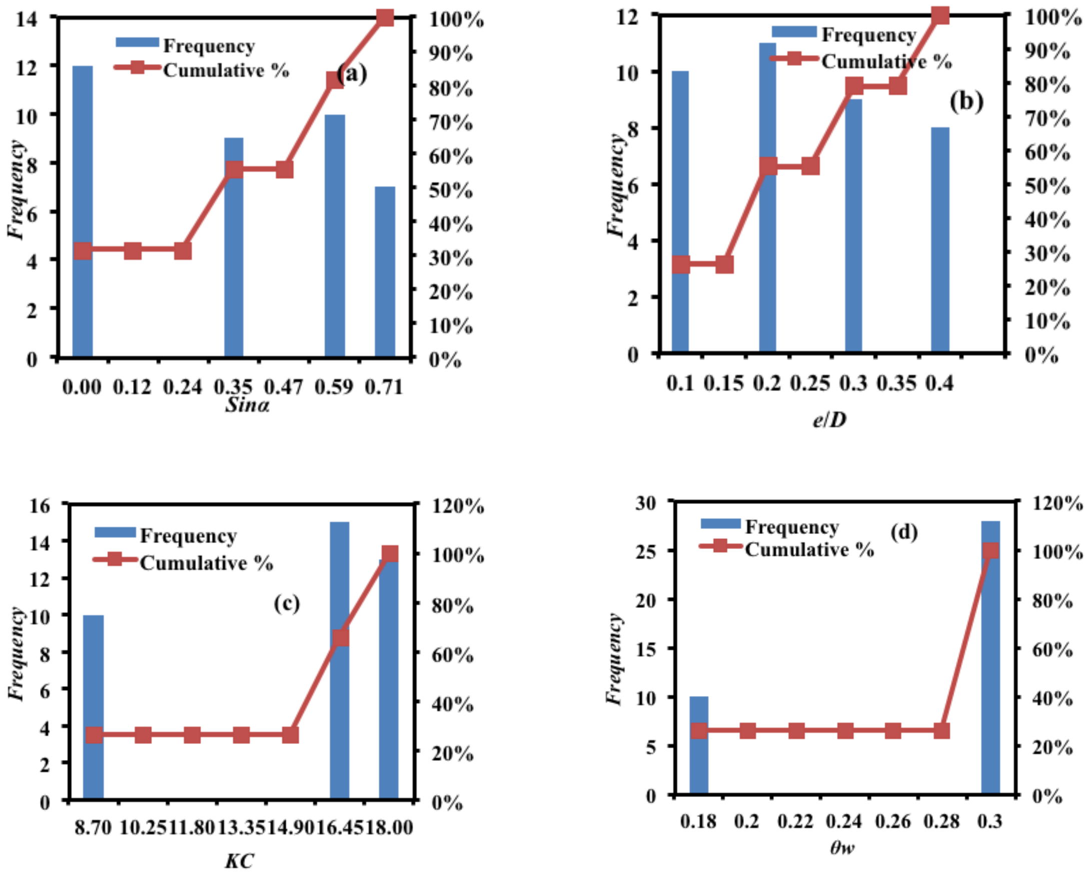

Figure 4. Frequency and cumulative relative frequency for the independent non-dimensional variables.

Figure 4a shows that the frequency of the flow attack angle, α, exhibits a rather fragmented pattern with a significant percentage of the cumulative relative frequency

(31.58%) devoted to $\alpha = 0$. As observed in Figure 4b, the frequency distribution of e/D is relatively symmetric and the tail of the distribution inclines to $e/D = 0.4$, whereas the majority of experimental observations (around 55%) were collected for $e/D \leq 0.2$. Figure 4c also illustrates that the Kelugan–Carpenter number, KC, was not exhaustively examined, with very fragmented distribution (i.e., $KC = 8.7,\ 15.8,$ and 18.0). In addition, the Shields' parameter, θ_w, was not systematically explored, with most of the experiments characterized by θ_w around 0.3 (Figure 4d). It is important to emphasize that all the experiments were performed using only one bed sediment and in live-bed conditions. This represents a limitation for the current available datasets, and further investigations on clear-water regime and various sediment beds would be of interest.

For the sake of completeness, the tests of Cheng et al. [9] considered in this study are summarized in Appendix A (Table A1).

4. Intelligent Computing Methods: Brief Descriptions and Implementations

The AI models considered benefit from two intrinsic advantages: The first merit is to drive regression-based equations along with highlighting the physical meaning of the experimental observations; the second remarkable merit is associated with the up-scaling of datasets. In this way, the dimensionless structure of equations given by the proposed AI models would be applicable at various scales from laboratory to field studies (e.g., Parsaie et al. [18], Najafzadeh and Saberi-Movahed [26], Ehteram et al. [27], and Sharafati et al. [30]).

4.1. Gene-Expression Programming (GEP)

Gene-Expression Programming (GEP), as an advanced platform of Genetic Algorithm (GA), employs populations of individuals and chooses the most suitable of them based on fitness. In fact, GEP puts genetic variation to use by at least one genetic operator (Ferreira [32]). The intrinsic disparity between GA and GEP techniques arises from the basic structure of the individuals. In GA, the individuals consist of chromosomes whose structure has linear strings with a fixed length, whereas in GEP, the individuals are converted into a coded form as linear strings whose length does not vary. In GEP, chromosomes, introduced as the genome, need to be represented as expression trees with a nonlinear mathematical structure. In addition, GEP employs characters related to linear chromosomes with strings consisting of genes structurally configured as a head and a tail. The structure of chromosomes needs to be modified using a wide range of genetic operators. In fact, efficiency of overall expression trees is inextricably bound with the appropriate selection of genetic operators (e.g., Ferreira [32]). Mutation rate, number of chromosomes, head size of chromosome, gene transposition, number of generations, and gene recombination are introduced as important genetic operators, which are used to set the optimum solution by Evolution Techniques (ETs). All the ETs in GEP are linked together by using one of the arithmetic operations.

The most favorable equations have been obtained for the prediction of the three dimensionless scour rates around subsea pipelines. Desirable genetic operations to extract the relationships with the highest level of precision via optimal evolution technique are listed in Appendix B, Table A2.

In this study, GEP models have been developed via GeneXproTools5 platform. During their development, the values of best-fit function were 1687, 1123, and 1322 for the estimation of V_H^*, V_R^*, and V_L^*, respectively. The best formulations are given in Appendix B (i.e., Equations (A1)–(A3)).

4.2. Evolutonary Polynomial Regression (EPR)

The EPR model is capable of effectively exploring the mathematical expressions that can be fitted to a dataset. The selection of a congruent mathematical structure is closely connected to the quality of existing pieces of information under the specific modeling phenomenon. There are seven general mathematical expressions to model phenomena.

One of the most frequently used expressions is known as (Giustolisi and Savic [33] and Berardi et al. [34]):

$$\Gamma = \xi_0 + \sum_{j=1}^{m} \xi_j \cdot (\eta_1)^{ES(j,1)} \cdot (\eta_2)^{ES(j,2)} \cdot \ldots \cdot (\eta_c)^{ES(j,c)} \cdot z\left[(\eta_1)^{ES(j,c+1)}\right] \cdot z\left[(\eta_2)^{ES(j,c+2)}\right] \cdot \ldots \cdot z\left[(\eta_c)^{ES(j,2c)}\right] \quad (10)$$

in which m is the maximum number of mathematical terms, ξ_0 is the bias term, ξ_j is a collection of coefficients, η is the input vector for a given problem, Γ is the output vector estimated by EPR, c is the number of elements in input vectors, z is a specific function with disparate mathematical structure, which needs to be defined by computer programmers, and ES is the exponent range which is specified by the user (Giustolisi and Savic [35] and Savic et al. [36]).

The typical inner function z is selected on the basis of prior knowledge about the phenomenon. There are three typical regression techniques (i.e., dynamic regression, statistical regression, and classification) that the EPR structure may use. Dynamic regression is applied for time-dependent datasets while the statistical regression is an appropriate approach for time-independent phenomena. In the classification modeling, a static system includes the output vector, which must be an integer value, and the input vectors that need to be arranged in classes. GA optimizes weighting coefficients used in Equation (10) by determining a fair number of generations, which is inextricably bound with some items such as number of input variables, exponent range of algebraic terms, and number of mathematical terms.

The development stage of EPR model begins with the integration of the Multi-Objective Genetic Algorithm (MOGA), which has been employed to optimally obtain the general EPR expressions [see Equation (10)] in terms of coefficients and exponents. The ES vectors including ± 2, ± 1.5, ± 1, ± 0.5, and 0 were used to develop Equation (10). The maximum number of terms has been fixed equal to 5 and, additionally, 3600 generations were produced for developing each regression equation returned by EPR. During each performance of the EPR model, several regression equations were obtained among which the best one, in terms of accuracy criterion (Mean Squared Error), was selected. Therefore, three regression-based equations with natural logarithmic functions have been achieved for the estimation of triple vectors of scour rates by EPR MOGA-XL software. These equations are given in Appendix B (i.e., Equations (A4)–(A6)).

4.3. Multivariate Adaptive Regression Spline (MARS)

In the early 1990s, an adaptive regression technique was introduced, based on spline conception, to build a non-parametric and stepwise regression equation; this model was called Multivariate Adaptive Regression Spline (MARS). This regression model appears to benefit immensely from the high potential of improving, over conventional regression-based-equations, the feeding number of observations between 50 and 1000. Between the principles of the MARS technique, the number of observations to develop a formulation is in the range of 50–1000. In the current study, the available 38 experimental observations are not well-suited in the applicability of the MARS technique, leading to an overfitting process. To eradicate this problem, K-folds (dataset partitioning) are considered within the performance of training and testing stages. In fact, 10 folds are set to fix the regression equation by MARS. Furthermore, the MARS technique has the capability to be a crucial enhancement in dealing with datasets with high dimensionality between 3 and 20. It provides a hierarchical structure using a set of basis functions (BFs), which are selected in a stepwise manner. Basically, the MARS model is capable of simulating a relationship between input variables and target values by an approximation function of the type (Friedman [37]):

$$\Omega(x_1, x_2, x_3, \ldots, x_O) = \omega_0 + \left[\omega_1 \cdot BF_1(X) + \omega_2 \cdot BF_2(X) + \omega_3 \cdot BF_3(X) + \ldots + \omega_Q \cdot BF_Q(X)\right] \quad (11)$$

where Ω, ω_0, O, and Q are the approximation function, constant coefficient, number of input variables, and number of BFs, respectively. In addition, $\omega = (\omega_1, \omega_2, \ldots, \omega_Q)$ is the set of weighting coefficients associated with the basis functions and $X = (x_1, x_2, \ldots, x_O)$ is the vector of input variables.

The programming computer codes of the MARS technique have been provided in MATLAB2015 software. In this research, the MARS models (i.e., Equations (A7)–(A9) in Appendix B) were developed for the prediction of V_H^*, V_L^*, and V_R^*. The numbers of BFs were set to 6, 7, and 4 for V_H^*, V_L^*, and V_R^*, respectively, depending on the standard deviation of each BF and the total influential number of parameters associated with the BFs.

4.4. M5 Model Tree (M5MT)

The M5 technique, as an efficient classification system, is frequently used to train models that estimate values. M5 generally consists of tree-like models, which are mathematically similar to multivariate-linear equations. It has the capability to deal with high dimensionality of datasets efficiently. Furthermore, the main merit of M5 technique is that model trees are much smaller and more accurate in comparison with the Regression Tree (RT). Tree models are basically built by means of divide-and-conquer technique (Quinlan [38]). The initial stage in constructing a model tree is to calculate the standard deviation (SD), as an error criterion, for the observed values in a training dataset. For a given sample of training dataset, a multivariate linear technique is fitted at each node of the model tree. M5 is limited to the input variables, which are referenced by linear equations somewhere in the node sub-tree. As each linear equation is constructed at each node, the mathematical shape of the linear equation can be eliminated to reach the optimum predicted error. Generally, M5 applies the Greedy Search (GS) algorithm in order to remove input variables (or attributes) that have marginal contributions to the development of the linear equation. Furthermore, processing the smoothing stage can enhance the accuracy level of each linear equation at the node.

M5MT model provides a collection of rules leading to linear regression equations. In the case of 3D-scour rate modeling, Weka3.9, as data-mining software, was considered. The linear equation being driven from M5 rules is expressed as:

$$V_{H,R,L}^* = a_0 + a_1 \cdot (1 + sin\alpha) + a_2 \cdot \left(1 - \frac{e}{D}\right) + a_3 \cdot KC + a_4 \cdot \theta_w \tag{12}$$

with a_0, a_1, a_2, a_3, and a_4 as constant coefficients of the linear equation.

In the process of M5 development, Tables A4 and A5 in the Appendix B summarize, respectively, the M5 rules and the associated linear equations returned by the Model Tree for the prediction of V_H^*. It can be seen from these tables that 12 rules were provided and, additionally, all 4 input variables were used as splitting parameters. In other words, all the inputs were important in setting the 12 formulations. In the case of V_L^* estimation, as it can be seen in Tables A6 and A7 in Appendix B, 12 rules were obtained from M5 analysis with splitting of the parameters $1 + \sin\alpha$, $1 - e/D$, and KC. Interestingly, the 12 relationships presented in Table A7 indicated that the Shields' parameter, θ_w, due to regular waves, has no role in predicting the scour rates at the left hand of the pipeline. Finally, as indicated in Tables A8 and A9 in Appendix B, 13 rules in terms of linear regression equations were extracted from M5 to predict V_R^*. Similarly, here too it was ascertained that the Shields' parameter does not control, at least significantly, the scour rate propagation at the right hand of the pipeline.

5. Results and Discussion

5.1. Performance Indices

The proposed models are evaluated using five performance measures (index): Index of Agreement (IOA), Sum of Squared Errors (SSE), Mean Absolute Error (MAE), Average Discrepancy Ratio (ADR), and Scatter Index (SI):

$$IOA = 1 - \frac{\sum_{j=1}^{N} \left[V_{Est}^*(j) - V_{Obs}^*(j)\right]^2}{\sum_{j=1}^{N} \left[\left|V_{Est}^*(j) - \overline{V_{Obs}^*}\right| + \left|V_{Obs}^*(j) - \overline{V_{Obs}^*}\right|\right]^2} \tag{13}$$

$$SSE = \left[\frac{\sum_{j=1}^{N} \left(V_{Est}^{*}(j) - V_{Obs}^{*}(j) \right)^{2}}{N} \right] \tag{14}$$

$$MAE = \left[\frac{\sum_{j=1}^{N} \left| V_{Est}^{*}(j) - V_{Obs}^{*}(j) \right|}{N} \right] \tag{15}$$

$$ADR = \frac{1}{N} \sum_{j=1}^{N} \frac{V_{Est}^{*}(j)}{V_{Obs}^{*}(j)} \tag{16}$$

$$SI = \frac{\sqrt{(1/N) \sum_{j=1}^{N} \left(\left(V_{Est}^{*}(j) - \overline{V_{Est}^{*}} \right) - \left(V_{Obs}^{*}(j) - \overline{V_{Obs}^{*}} \right) \right)^{2}}}{(1/N) \sum_{j=1}^{N} V_{Obs}^{*}(j)} \tag{17}$$

in which V_{Est}^{*} and V_{Obs}^{*} are the computed and observed scour rate values, respectively, N is the number of experimental observations, and $\overline{V_{Est}^{*}}$ and $\overline{V_{Obs}^{*}}$ are the average values for computed and observed scour rate, respectively. The Index of Agreement (IOA), varying between 0 and 1, expresses the Mean Square Error (MSE) ration to the potential error. The IOA of 1 shows the most favorable agreement whereas for IOA=0, no agreement is found for the AI model under study. Furthermore, IOA is capable of detecting additive and proportional differences among experimental and computed averages and variances. The SSE is able to measure how well the performance of AI models in 3D scour rates prediction matches the corresponding observed scour rates. Smaller values of SSE are indicative of a more reliable prediction, and additionally, a value of zero expresses the most favorable accuracy level of the model under study. MAE is a statistical criterion to measure error values between predicted and observed values; a value of zero indicates the best performance of the model. ADR can be an index to demonstrate differences between estimated and predicted values in terms of over prediction (or under prediction). The most desired value of ADR is 1, indicating that the AI model stands at the highest level of accuracy. Results of AI models demonstrate over prediction for ADR > 1 and under prediction for ADR < 1. Ultimately, the error criterion of SI is indicative of presenting Root Mean Squares Difference (%) in regard to average observation. The model with SI = 0 indicates the best performance in terms of accuracy.

5.2. Performance of Data-Driven Models under Study

In this subsection, the performance of the considered AI models is discussed. In the case of M5MT, we will refer to the more complete equations marked with #1 in the Table A5, Table A7, and Table A9. Table 2 shows training and testing results for prediction of V_{H}^{*}. By a quick glance, the values of statistical benchmarks illustrate that the EPR model predicted non-dimensionless V_{H}^{*} parameters with the highest accuracy (IOA = 0.9799, SSE = 0.0030, and MAE = 0.0550) in the training stage compared to the other AI models, followed by MARS (IOA = 0.9468, SSE = 0.0068, and MAE = 0.0826), M5MT (IOA = 0.9468, SSE = 0.0612, and MAE = 0.0886), and GEP (IOA = 0.9244, SSE = 0.0111, and MAE = 0.1056). Furthermore, values of SI (0.0599) and ADR (1.0026) given by the EPR model displayed this trend. All the ADR values, which were approximately close to 1, indicated highly satisfying performance for AI models. In the case of the testing stage, it is surprising that the EPR model was found to fail in the prediction of V_{H}^{*} even though it had the best performance in the training stage. Hence, the assessment of statistical benchmarks indicated that MARS had highest most accuracy level (IOA = 0.9506, SSE = 0.1156, and MAE = 0.0912), followed by M5MT (IOA = 0.9430, SSE = 0.1323, and MAE = 0.0644), GEP (IOA = 0.9084, SSE = 0.2143, and MAE = 0.0893), and EPR (IOA = 0.8237, SSE = 0.4124, and MAE = 0.1254). According to Table 2, SI values demonstrated the superiority of the M5MT technique. From ADR values, all AI models showed relatively marginal under prediction (ADR < 1) in the testing phase.

Table 2. Statistical performances of the AI models in the estimation of V_H^*.

AI Models	Training Stage				
	IOA	SSE	MAE	SI	ADR
M5MT	0.9468	0.0612	0.0886	0.0974	1.0090
GEP	0.9244	0.0111	0.1056	0.1158	1.0039
EPR	0.9799	0.0030	0.0550	0.0599	1.0026
MARS	0.9468	0.0068	0.0826	0.1016	1.0068
AI Models	**Testing Stage**				
	IOA	**SSE**	**MAE**	**SI**	**ADR**
M5MT	0.9430	0.1323	0.0644	0.0644	0.9682
GEP	0.9084	0.2143	0.0893	0.0893	0.9621
EPR	0.8237	0.4124	0.1254	0.1988	0.9260
MARS	0.9506	0.1156	0.0912	0.1048	0.9476

The performance of AI models for the estimation of V_H^* in training and testing stages is illustrated in Figure 5. In Figure 5, data are fitted to ±25% error lines. Most values of V_H^* predicted by the AI models in the training stage were in the range of an acceptable error bound. Similarly, Figure 5b shows well-matched predictions versus experimental observations, though a relatively significant under prediction is seen in the observed value of V_H^* = 3.8 for M5MT, GEP, and EPR models.

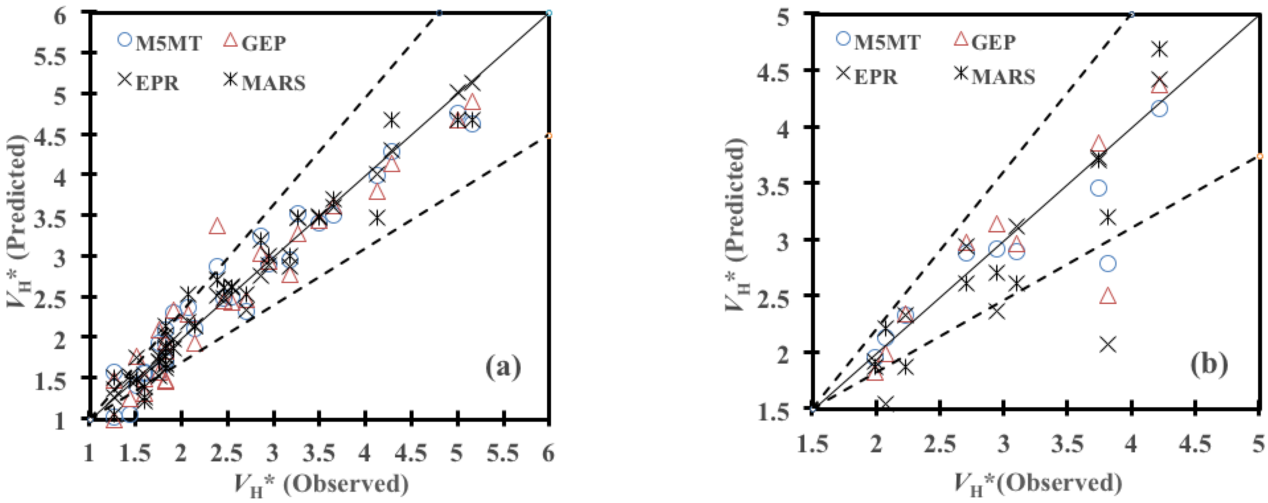

Figure 5. Performance of AI models in the prediction of V_H^* for: (**a**) Training and (**b**) testing stage.

In Table 3, statistical assessments of training and testing results in the prediction of V_L^* are given. It is inferred that the MARS technique (i.e., Equation (A9)) estimated V_L^* with the highest level of precision (IOA = 0.9950, SSE = 0.0157, and MAE = 0.0374) in the training phase in comparison with M5MT (IOA = 0.9877, SSE = 0.0387, and MAE = 0.0672), EPR (IOA = 0.9883, SSE = 0.0369, and MAE = 0.0754), and GEP (IOA = 0.9680, SSE = 0.1001, and MAE = 0.1027). SI (0.0504) and ADR (0.9993) obtained by the MARS model were also indicative of its superiority when compared to the other AI models. Table 3 also indicates that M5MT and EPR models had approximately the same performance in the prediction of V_L^* according to the statistical benchmarks (i.e., IOA, SSE, and SI). Moreover, it can be inferred from the ADR values that all AI models had convincing efficiency; specifically, the MARS model experienced a very marginal under prediction whereas the other AI approaches showed insignificant over predictions. Results of the testing stage demonstrated that the GEP-based formulation (i.e., Equation (A2)) had the best performance (IOA = 0.9911, SSE = 0.0384, and MAE = 0.0544) in the prediction of V_L^*. M5MT

(IOA = 0.9896, SSE = 0.0449, and MAE = 0.0644), and MARS (IOA = 0.9828, SSE = 0.0738, and MAE = 0.0870) stood at the second and third place in terms of precision level, respectively. Ultimately, EPR (i.e., Equation (A5)) provided V_L^* predictions with the lowest level of accuracy in terms of SSE (0.1322) and MAE (0.0944). SI values confirm an unconvincing performance associated with EPR (SI = 0.1502). ADR values indicated a very marginal under estimation (ADR < 1) for the GEP model.

Table 3. Statistical performances of the AI models in the estimation of V_L^*.

AI Models	Training Stage				
	IOA	**SSE**	**MAE**	**SI**	**ADR**
M5MT	0.9877	0.0387	0.0672	0.0772	1.0277
GEP	0.9680	0.1001	0.1027	0.1274	1.0242
EPR	0.9883	0.0369	0.0754	0.0771	1.0085
MARS	0.9950	0.0157	0.0374	0.0504	0.9993
AI Models	**Testing Stage**				
	IOA	**SSE**	**MAE**	**SI**	**ADR**
M5MT	0.9896	0.0449	0.0644	0.0481	1.0520
GEP	0.9911	0.0384	0.0544	0.0611	0.9987
EPR	0.9693	0.1322	0.0944	0.1052	1.0624
MARS	0.9828	0.0738	0.0870	0.0680	1.0729

For the sake of qualitative comparisons, Figure 6 illustrates the comparison between predicted and observed values of V_L^*. In the case of the training stage, Figure 6a shows very compromising efficiency with almost all predictions inside ±25% error lines. In the case of the testing stage, the EPR model remarkably overpredicted with the observed value of $V_L^* = 4.52$ (Figure 6b).

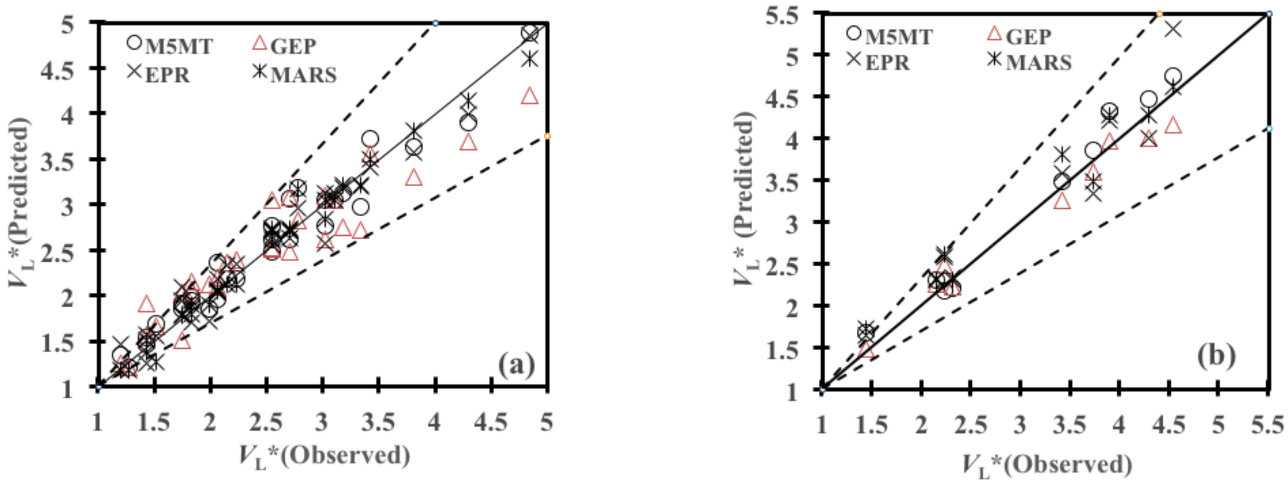

Figure 6. Performance of AI models in the prediction of V_L^* for: (**a**) training and (**b**) testing stage.

Table 4 shows statistical benchmarks associated with the training and testing phases for prediction of V_R^*. In the training phase, statistical criteria demonstrated that the EPR-base relationship [Equation (A6)] predicted V_R^* values with the highest level of accuracy (i.e., IOA = 0.9913, SSE = 0.0345, and MAE = 0.1030) than those obtained by other AI models. Furthermore, SI values given by Table 4 indicate superiority of EPR model over other AI techniques. According to IOA, SSE, and SI criteria, the MARS-based formulation (Equation (A9)) placed second in precision (IOA = 0.9831, SSE = 0.0670, and SI = 0.103), followed by M5MT (IOA = 0.9801, SSE = 0.0788, and SI = 0.1046) and GEP (IOA = 0.976, SSE = 0.0953, and SI = 0.0964). In addition to this, ADR values illustrated that all AI models had compromising performance with values of ADR roughly close to 1. These values are also indicative of very small overpredictions. The assessment of performance for the testing phase indicated that the MARS model (Equation (A9)) had the most accurate prediction in terms of IOA (0.9670), SSE (0.0855), and MAE (0.1086) in comparison to other AI models. Additionally, the value of SI for MARS was indicative of the most successful performance. M5MT with IOA of 0.9435, SSE of 0.1581, and MAE of 0.1356 came in second. The performance of GEP model (Equation (A3)) with IOA = 0.9329 and SSE = 0.1892 was comparatively at the same accuracy level as for EPR (Equation (A6)) (IOA = 0.9356 and SSE = 0.1816). Additionally, MAE and SI values proved similarity of performance between GEP and EPR models. Moreover, MARS model had more significant underprediction (ADR = 0.9578) than GEP (ADR = 1.0020) with very marginal overestimation. ADR analysis also demonstrated that M5MT and EPR provided low underprediction and fairly relative overprediction, respectively.

Table 4. Statistical performances of the AI models in the estimation of V_R^*.

AI Models	Training Stage				
	IOA	SSE	MAE	SI	ADR
M5MT	0.9801	0.0788	0.0728	0.1046	1.0163
GEP	0.9760	0.0953	0.1199	0.1147	1.0294
EPR	0.9913	0.0345	0.1030	0.0692	1.0113
MARS	0.9831	0.0670	0.1030	0.0964	1.0050
AI Models	**Testing Stage**				
	IOA	**SSE**	**MAE**	**SI**	**ADR**
M5MT	0.9435	0.1581	0.1356	0.1601	0.9955
GEP	0.9329	0.1892	0.1525	0.1755	1.0020
EPR	0.9356	0.1816	0.1500	0.1713	1.0503
MARS	0.9670	0.0855	0.1086	0.1601	0.9578

In terms of illustrative comparisons in training phase, Figure 7a indicated that GEP and MARS techniques underpredicted, with an observed value of $V_R^* = 1.11$.

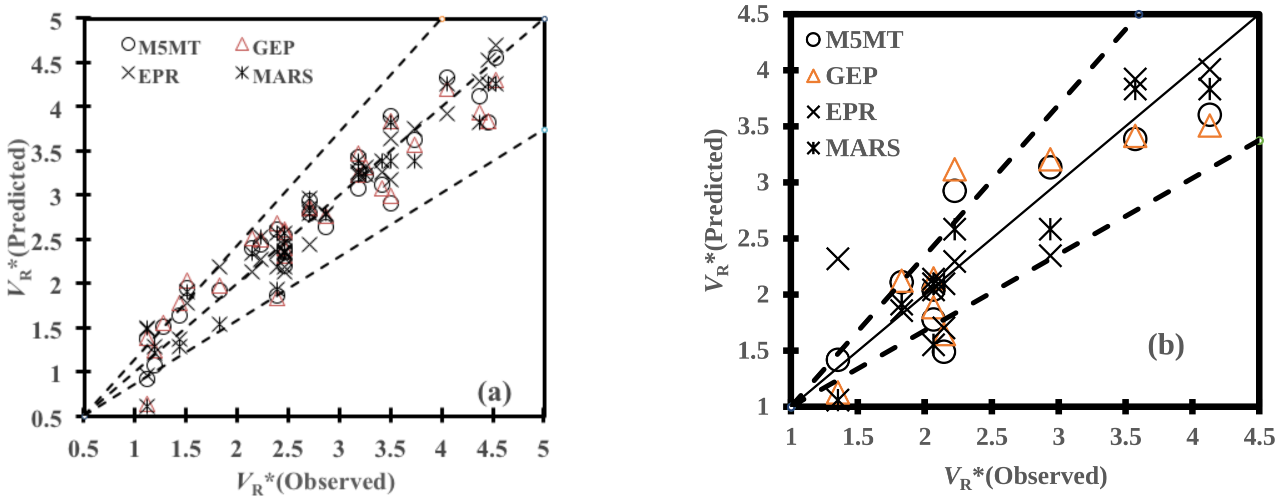

Figure 7. Performance of AI models in the prediction of V_R^* for: (**a**) Training and (**b**) testing stage.

As it can be seen in Figure 7a, the scatter plot for training phase shows highly prosperous performance with almost all data inside ±25% error ranges. Relatively important overprediction for M5MT and GEP was seen in Figure 8b at $V_R^* = 2.22$. MARS and GEP techniques underpredicted for $V_R^* = 1.35$, whereas EPR provided significant overprediction.

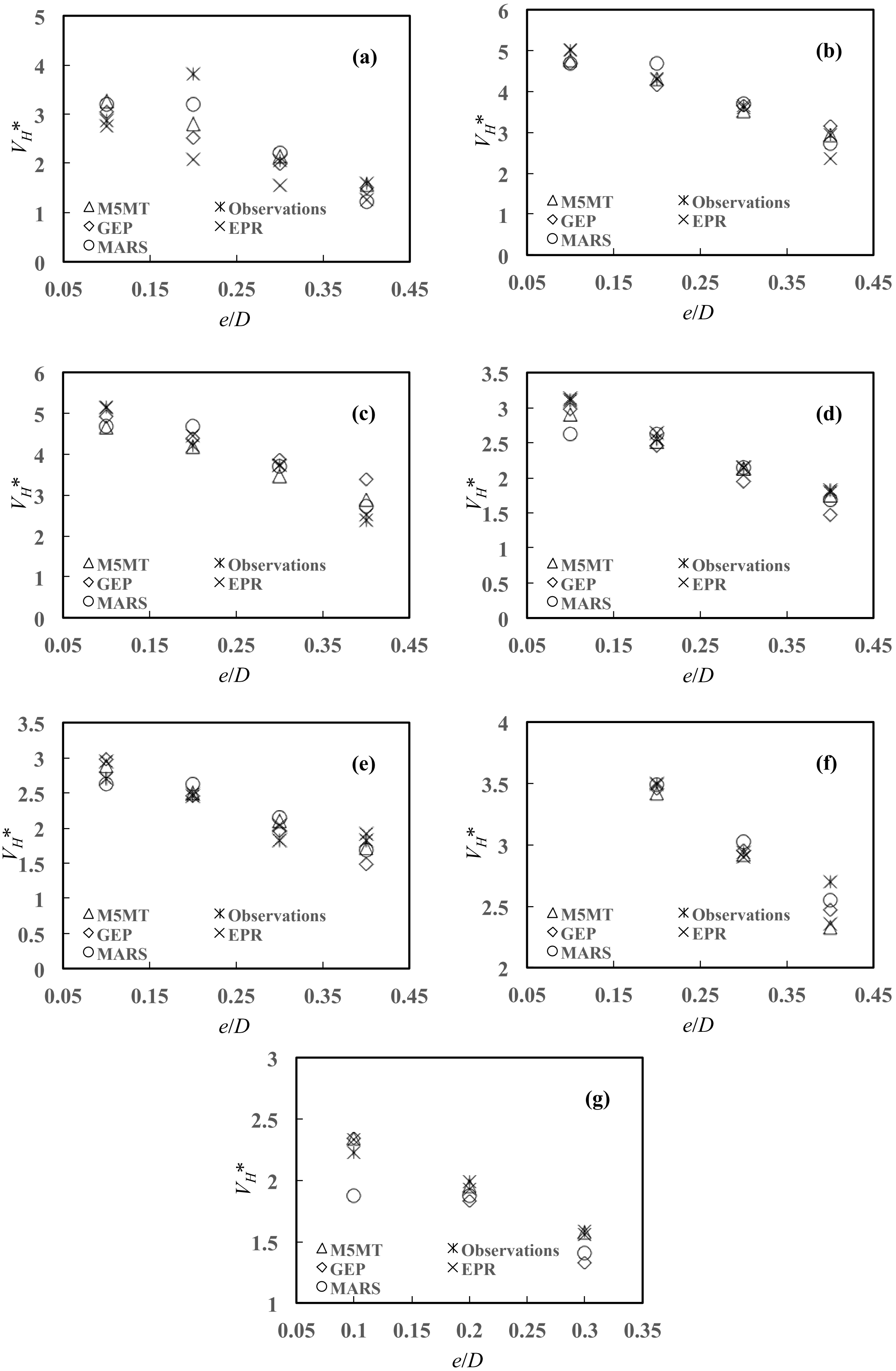

Figure 8. Variations of V_H^* against e/D at: (**a**) KC = 8.7 and α = 0°, (**b**) KC = 15.8 and α = 0°, (**c**) KC = 18.0 and α = 0°, (**d**) KC = 15.8 and α = 30°, (**e**) KC = 18.0 and α = 30°, (**f**) KC = 18.0 and α = 15°, and (**g**) KC = 15.8 and α = 45°.

5.3. Driving Physical Meaning of AI Results

In this section, the physical performance of the AI models is controlled by analyzing 3D scour propagation rates versus e/D and KC values.

5.3.1. Effects of e/D on V_H^*

Figure 8 illustrates the variations of V_H^* as a function of e/D. In fact, this type of variation was assessed for all the considered AI models at various levels of KC and α.

Figure 8a shows that the experimental values of V_H^*, for KC = 8.7 and $\alpha = 0°$ increase from 2.77 at e/D = 0.1 to 3.81 at e/D = 0.2 and then descend towards 1.51 at e/D = 0.4. This non-monotonic trend is well captured by the MARS model, the results of which are more in harmony with the observations in comparison to the remaining models. All the AI models show underprediction at e/D = 0.2. Quantitatively, the EPR model exhibits significant underprediction with SSE = 0.843 in comparison to GEP (SSE = 0.435), M5MT (SSE = 0.300), and MARS (SSE = 0.161). According to Figure 8b, all the AI models were in satisfying agreement with the observations in the case of KC = 15.8 and $\alpha = 0°$. Generally, AI models predict a linear downward trend between V_H^* and e/D, as observed experimentally. Similarly, Figure 8c, which relates to the case of KC = 18.0 and $\alpha = 0°$, shows an even more pronounced linear decreasing trend between V_H^* and e/D, quite well captured by all the AI models. EPR had the most satisfying performance (SSE = 0.018), while the GEP model showed a significant overprediction at e/D = 0.4. Figure 8d, which relates to the case KC = 15.8 and $\alpha = 30°$, exhibiting more alleviated values of V_H^* when compared to those in Figure 8b. This is due to the higher value of α (equal to 30° here while equal to 0° in Figure 8b); it remains the observed downward trend between V_H^* and e/D. Moreover, in this case, the considered AI models perform satisfactorily by assessing the effect of the angle α properly, though the MARS model appears more resilient for e/D from 0.1 to 0.2. Moreover, MARS and GEP models indicated marginal underestimations at e/D = 0.1 and 0.4. More specifically, EPR with SSE of 0.003 shows the most compromising efficiency compared to M5MT (SSE = 0.011), GEP (SSE = 0.0468), and MARS (SSE = 0.0631). Figure 8e, which relates to the case KC = 18.0 and $\alpha = 30°$, clearly shows the effect of the angle α when compared to Figure 8c, while it is less significant in realizing the effect of KC when compared to Figure 8d (on the other hand, KC in Figure 8d is equal to 15.8 and slightly differs from KC = 18.0 in Figure 8e). It is confirmed that V_H^* decreases with increasing α as well as decreasing with increasing normalized embedment depth e/D (especially for the highest values of KC). AI models are still in harmony with these observed trends. Quantitatively, EPR with an SSE of 0.028 had the most compromising efficiency compared to M5MT (SSE = 0.031), MARS (SSE = 0.040), and GEP (SSE = 0.054). Further effects of the wave angle of attack α on V_H^* are highlighted in Figure 8g, which relates to the case KC = 15.8 and $\alpha = 45°$, when Figure 8b,d is considered for comparison. In this analysis, significant underpredictions were found for the MARS model (SSE = 0.054) at e/D = 0.1 and for GEP (SSE = 0.035) and MARS (SSE = 0.054) at e/D = 0.4. For the sake of completeness, Table 5 below shows the performance index SSE for all AI models by varying the normalized embedment depth e/D. Overall, it appears that the M5MT model had the best performance compared to the other AI models, though it is not always quite so straightforward.

Table 5. SSE values for the considered AI models when predicting V_H^* at various e/D ratios.

KC and α Values	SSE Values for AI Models			
	M5MT	GEP	EPR	MARS
$KC = 8.7$ and $\alpha = 0°$	0.300	0.435	0.843	0.161
$KC = 15.8$ and $\alpha = 0°$	0.018	0.040	0.081	0.078
$KC = 18.0$ and $\alpha = 0°$	0.147	0.276	0.018	0.142
$KC = 18.0$ and $\alpha = 15°$	0.049	0.018	0.040	0.010
$KC = 15.8$ and $\alpha = 30°$	0.011	0.047	0.003	0.063
$KC = 18.0$ and $\alpha = 30°$	0.031	0.054	0.028	0.040
$KC = 15.8$ and $\alpha = 45°$	0.005	0.035	0.006	0.054

5.3.2. Effects of e/D and KC on V_R^*

Figure 9a–d shows the variations of V_R^* against e/D at given values of KC and for $\alpha = 0°$, according to AI models. All AI models exhibit a downward trend when interpreting the variations of V_R^* with e/D increasing from 0.1 to 0.4. For example, M5MT simulates that V_R^* decreases linearly with increasing e/D, as shown in Figure 9a. In addition, for a constant value of e/D, V_R^* increases with increasing KC, with a tendency towards an asymptotic value. For example, V_R^* increases sharply from 2.94 for KC = 8.7 to 4.56 for KC = 18, at e/D = 0.1 as shown in Figure 10a. This trend was generally repeated in Figure 9b–d.

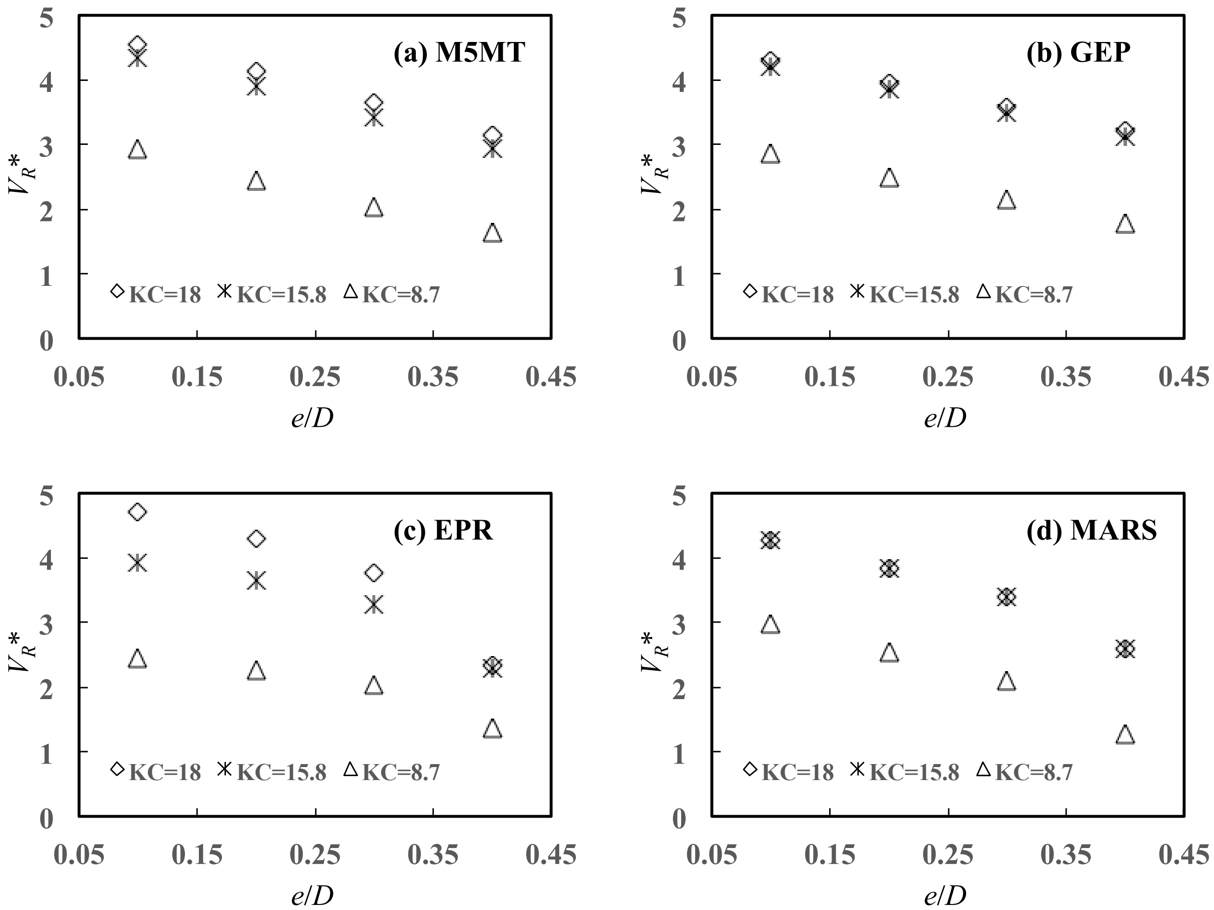

Figure 9. Variations of V_R^* against e/D at KC from 8.7 to 18.0 and α = 0° for: (**a**) M5MT, (**b**) GEP, (**c**) EPR, and (**d**) MARS.

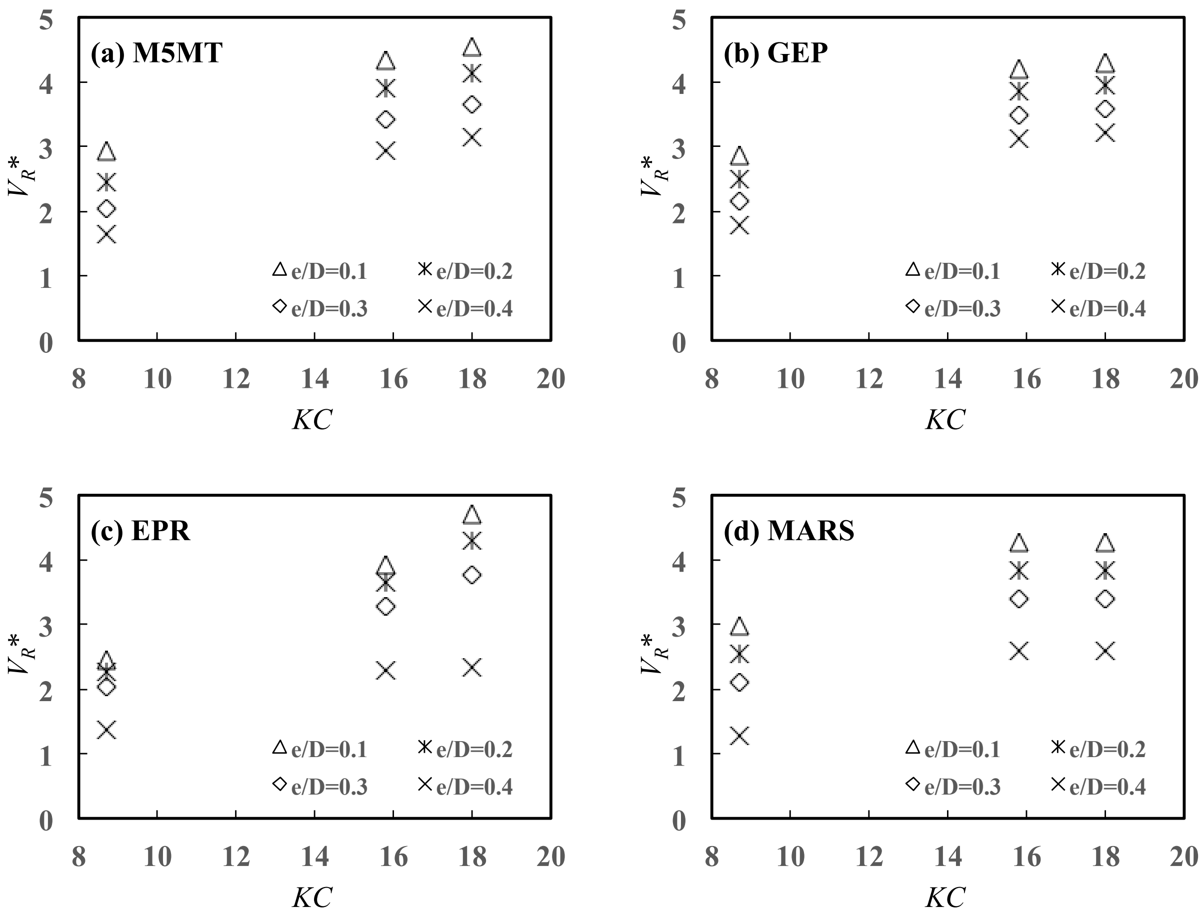

Figure 10. Variations of V_R^* against KC at e/D from 0.1 to 0.4 and α = 0° for: (**a**) M5MT, (**b**) GEP, (**c**) EPR, and (**d**) MARS.

Analogously, Figure 10a–d shows the variations of V_R^* against KC at given values of e/D, according to AI models. All AI models exhibit an upward trend when interpreting the variations of V_R^* with KC increasing from 8.7 to 18. For example, M5MT simulates that V_R^* increases from 2.4 for KC = 8.7 to 4.56 for KC = 18, at e/D = 0.2. However, in general, the trend is not linear but approximately asymptotic. In addition, for a constant value of KC, V_R^* decreases as e/D increasing. For example, V_R^* values decline from 4.344 for e/D = 0.1 to 2.927 for e/D = 0.4, at KC = 15.8, as shown in Figure 10d.

Moreover, statistical performance of AI techniques for the prediction of V_R^* with consideration of e/D and KC variations at α = 0° is presented in Table 6. It was found that all the AI techniques had the best performance for better understanding the general pattern of V_R^* versus e/D at KC = 8.7. For instance, the GEP model indicated the most satisfying efficiency at KC = 8.7 with an SSE of 0.0622 than results predicted at KC = 15.8 (SSE = 0.2602) and KC = 18 (SSE = 0.0827). For KC = 8.7, EPR was found more compromisingly for the assessment of V_R^* with SSE = 0.0168 in comparison with M5MT (SSE = 0.0388), MARS (SSE = 0.0483), and GEP (SSE = 0.0622). In the case of KC = 15.8, EPR had the most successful efficiency while for KC = 18, the M5MT technique stood at the highest level of accuracy for simulation of V_R^* variations versus e/D ratios. In Table 6, it was inferred that AI models indicated the best performance to perceive variations of V_R^* versus KC values for different e/D ratios. MARS (SSE = 0.0544), M5MT (SSE = 0.0244), and EPR (SSE = 0.0035) had the most accurate efficiency at e/D = 0.3 while for e/D = 0.1, the GEP technique had the most prosperous performance (SSE = 0.0349). GEP showed variations of V_R^* versus at e/D = 0.1, with the highest level of accuracy (SSE = 0.0349) whereas for e/D = 0.2, EPR was determined as the most precise model. On the other hand, EPR (SSE = 0.0035) and MARS (SSE = 0.0920) stood at the highest level of accuracy at e/D = 0.3 and e/D = 0.4, respectively.

Table 6. SSE values for the considered AI models when predicting V_R^* at various e/D ratios and KC values.

AI Models	SSE Values for e/D Varying		
	KC=8.7	KC=15.8	KC=18.0
M5MT	0.0388	0.2038	0.0271
GEP	0.0622	0.2602	0.0827
MARS	0.0483	0.0844	0.1490
EPR	0.0168	0.0132	0.0969

AI Models	SSE Values for KC Varying			
	e/D=0.1	e/D=0.2	e/D=0.3	e/D=0.4
M5MT	0.0480	0.0948	0.0244	0.1924
GEP	0.0349	0.1298	0.0416	0.3338
MARS	0.0629	0.1662	0.0544	0.0920
EPR	0.0355	0.0106	0.0035	0.1197

5.3.3. Effects of e/D and KC on V_L^*

Figure 11a–d shows the variations of V_L^* against e/D at given values of KC, and for $\alpha = 0°$, according to AI models. All AI models exhibit a downward trend when interpreting the variations of V_L^* with e/D increasing from 0.1 to 0.4. For example, M5MT simulates that V_L^* decreases linearly with increasing e/D, as shown in Figure 11a. In addition, for a constant value of e/D, V_L^* increases with increasing KC, with a tendency towards an asymptotic value. For example, V_L^* increases sharply from 2.776 for KC = 8.7 to 4.482 for KC = 18, at e/D = 0.2.

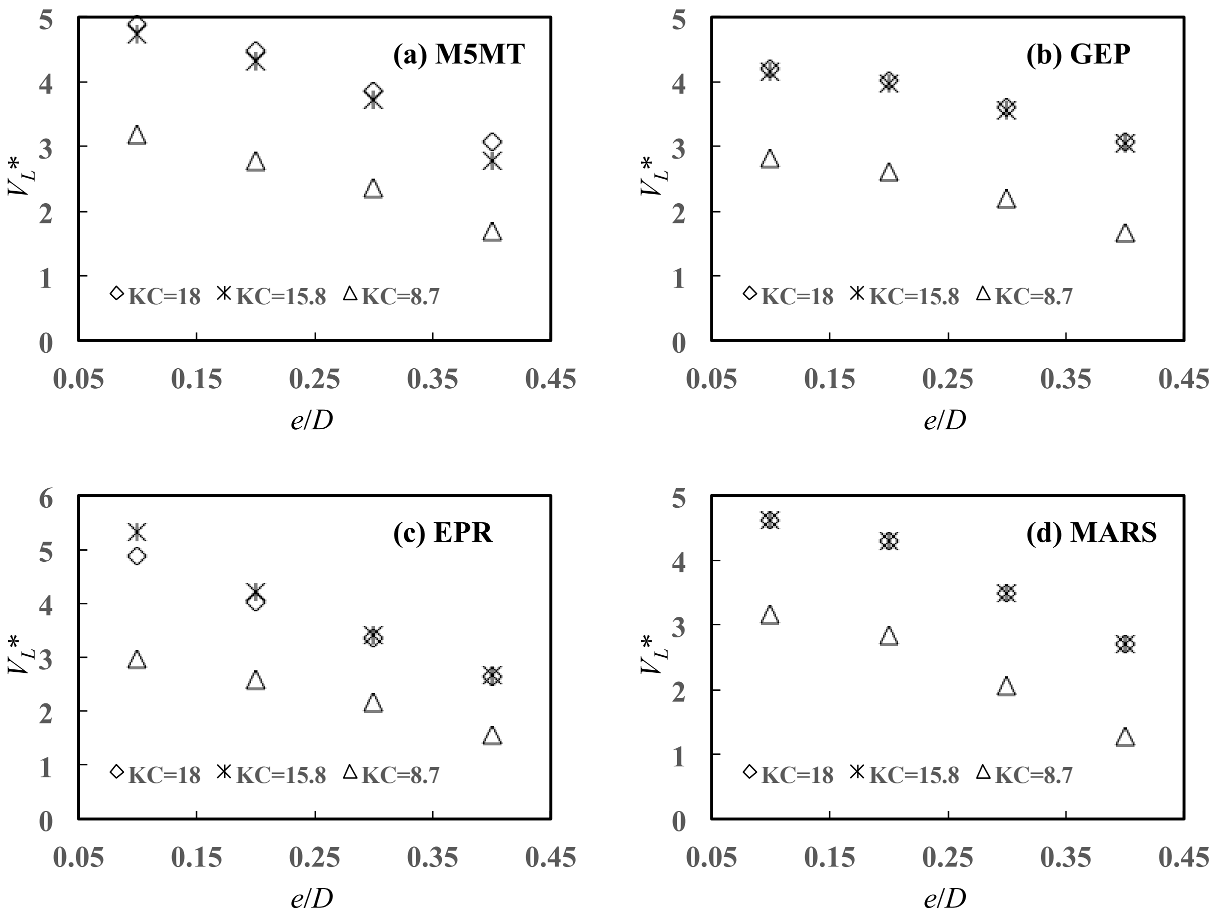

Figure 11. Variations of V_L^* against e/D at KC from 8.7 to 18.0 and $\alpha = 0°$ for: (**a**) M5MT, (**b**) GEP, (**c**) EPR, and (**d**) MARS.

Analogously, Figure 12a–d shows the variations of V_L^* against KC at given values of e/D, according to AI models. All AI models exhibit an upward trend when interpreting the variations of V_L^* with KC increasing from 8.7 to 18. For example, GEP simulates that V_L^* increases from 2.83 for KC = 8.7 to 4.207 for KC = 18, at e/D = 0.1. However, in general, the trend is not linear but approximately asymptotic. In addition, for a constant value of KC, V_L^* decreases with e/D increasing. For example, V_L^* plummets from 4.168 for e/D = 0.1 to 3.047 for e/D = 0.4, at KC = 15.8, as shown in Figure 12b.

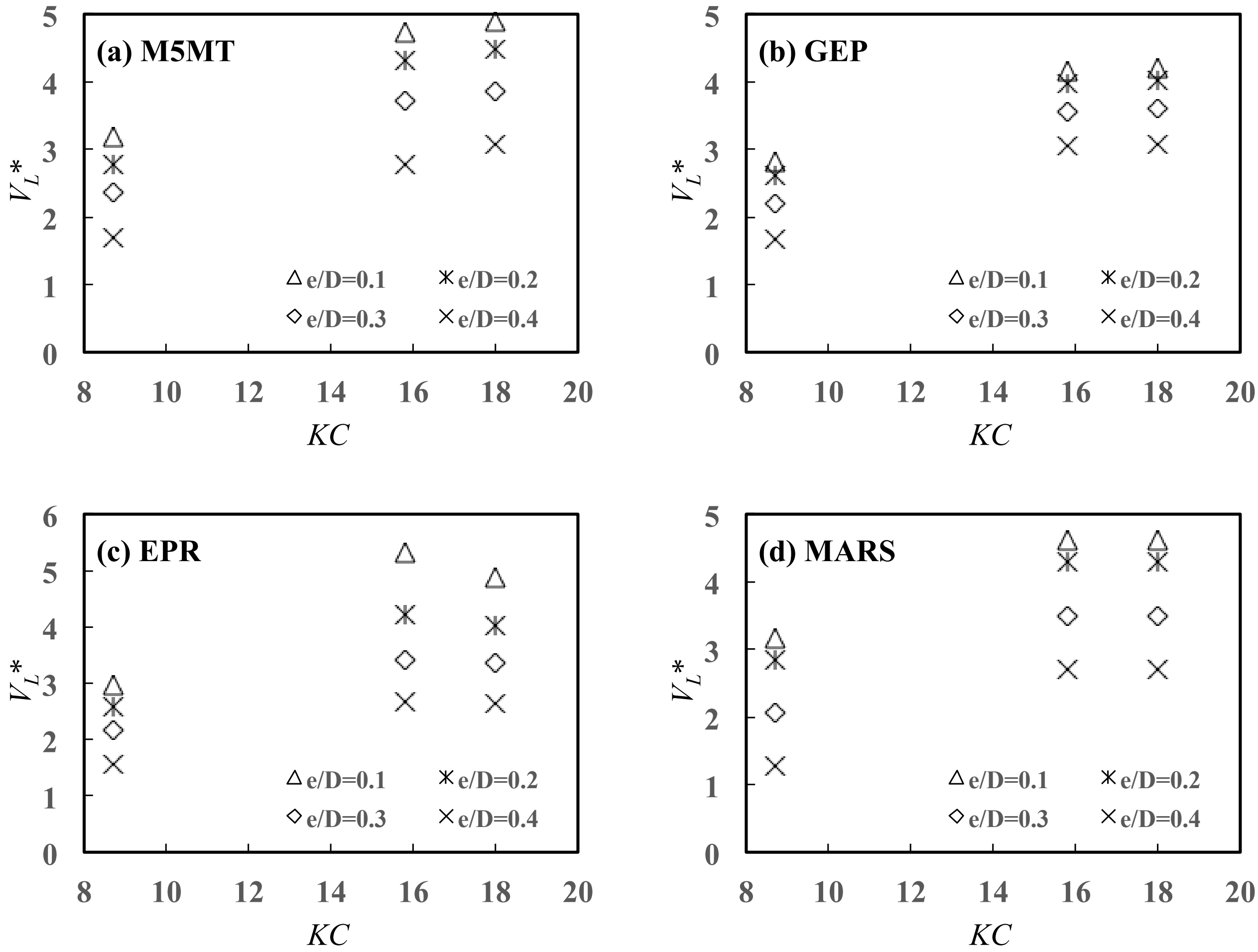

Figure 12. Variations of V_L^* against KC at e/D from 0.1 to 0.4 and α = 0° for: (**a**) M5MT, (**b**) GEP, (**c**) EPR, and (**d**) MARS.

According to Table 7, the GEP model had the best performance (SSE = 0.0502) at KC = 8.7 while all other AI models indicated the most accurate prediction of V_L^* values for KC = 18. The EPR technique predicted variations of V_L^* versus e/D with highly satisfying performance at KC = 15.8 (SSE = 0.0494) and KC = 18 (SSE = 0.0272). Table 7 also indicates that M5MT (SSE = 0.0658), EPR (SSE = 0.0207), and GEP (SSE = 0.0188) had the most favorable performance at e/D = 0.3 while MARS (SSE = 0.00633) stood at the highest level of efficiency at e/D = 0.4. EPR model indicated more successful efficiency for e/D = 0.1 (SSE = 0.0720) and e/D = 0.2 (SSE = 0.0617) than the remaining AI models. What is more, GEP raised the most well-matched predictions for the variations of V_L^* versus KC with SSE = 0.0188 at e/D = 0.2, while EPR had the best performance (SSE = 0.00633) at e/D = 0.4.

Table 7. SSE values for the considered AI models when predicting V_L^* at various *e/D* ratios and *KC* values.

AI Models	SSE Values for *e/D* Varying		
	KC=8.7	*KC*=15.8	*KC*=18.0
M5MT	0.0874	0.0958	0.0488
GEP	0.0502	0.1037	0.1627
MARS	0.0618	0.1847	0.0563
EPR	0.0606	0.0494	0.0272

AI Models	SSE Values for *KC* Varying			
	e/D=0.1	*e/D*=0.2	*e/D*=0.3	*e/D*=0.4
M5MT	0.0725	0.0959	0.0658	0.0751
GEP	0.1804	0.0803	0.0188	0.1425
MARS	0.2189	0.1260	0.0526	0.0063
EPR	0.0720	0.0617	0.0207	0.0286

6. Conclusions

This study aimed to explore 3D scouring patterns around subsea pipelines under wave-only conditions. To this purpose, the experimental work by Cheng et al. [9] was considered. Experimental data were collected in a wave flume 50 m long, 4 m wide, and 2.5 m deep. A total of 60 wave-only tests were considered, though some of these tests (i.e., 22 tests) have proved to be of no use. An almost uniform siliceous sand with d_{50} equal to 0.37 mm and specific gravity of 2.70 was used for the mobile bed. Experiments were conducted under live-bed conditions with the Shields' parameter θ_w ranging from 0.18 to 0.30, the Keulegan–Carpenter number KC from 8.7 to 18.0, the initial pipeline embedment depth, e, from 0.1D to 0.5D, and the flow incident angle, α, relative to the pipeline varying from 0° to 45°. Four robust data-driven models were applied to predict the 3D scour propagation around undersea pipelines, namely Gene-Expression Programming (GEP), Evolutionary Polynomial Regression (EPR), Multivariate Adaptive Regression Spline (MARS), and M5 Model Tree (M5MT). Generally, the following conclusions can be drawn from the current investigation:

1. The considered predictive methodologies indicated two main potentialities: (i) Providing non-linear regression equations with a high degree of complexity (as naturally seen in the scouring process around submarine pipelines) when developed by a limit volume of datasets; (ii) selecting the effective variables (i.e., e/D, KC, α, θ_w) on the scour rates estimation under a full automated manner.
2. All AI models applied in this study provided explicit relationships with satisfying performance for the prediction of 3D scour rates. However, the performance indices in the training and testing phases overall revealed that the MARS technique is most the eligible one. Therefore, the scour Equations (A7)–(A9), given in Appendix B, associated with the Basis Functions (BFs) explicitly stated in Table A3, are recommended in the prediction of V_H^*, V_L^*, and V_R^*, respectively.
3. The physical consistency of the AI models was controlled by analyzing 3D scour propagation rates versus the normalized embedment depth e/D, the flow incident angle α, and Keulegan–Carpenter number KC. It was confirmed that scour propagation rates decrease with increasing e/D, decrease with increasing the angle of attack α, and increase with KC number. Conversely, it was ascertained the Shields' parameter does not control, at least significantly, the scour rate propagation at the right- and left-hand pipeline shoulders.
4. A comparison with the conventional approach by [9] for the prediction of V_H^* would prove the superiority of the proposed scour equations. For example, by applying Equation (8) in [9], the performance index IOA is 0.842, significantly lower than the values of IOA for all the considered AI techniques (i.e., 0.9670 for MARS, 0.9435 for M5MT, 0.9356 for EPR, and 0.9329 for GEP).

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.

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

List of Notations and Acronyms

Notations

a_{i}	constant coefficient in M5MT model
c	number of elements in input vectors in EPR model
D	pipeline diameter
d_{50}	median grain size of the seabed sediment
e	pipeline embedment depth
ES	exponent range in EPR model
g	acceleration due to gravity
KC	Kelugan -Carpenter number
m	maximum number of mathematical terms in EPR model
N	number of experimental observations
O	number of input variables in MARS model
Q	number of Basis Functions in MARS model
Re_w	pipeline Reynolds number
T	wave period
U_w	wave orbital velocity at the seabed
V_H	scour propagation velocity along the longitudinal (axial) direction of the pipeline
V_L	scour propagation velocity at the left-hand shoulder of the pipeline
V_R	scour propagation velocity at the right-hand shoulder of the pipeline
V^*_{Est}	computed scour rate value
$V^*_{H,L,R}$	dimensionless scour propagation velocity
V^*_{Obs}	observed scour rate value
$\overline{V^*_{Est}}$	average value for computed scour rate values
$\overline{V^*_{Obs}}$	average value for observed scour rate values
x_{i}	i-th input variable in MARS model
X	vector of input variables in MARS model
z	specific function in EPR model
α	flow incident angle to the pipeline (or angle of attack)
ϕ	angle of repose of the seabed sediment
Γ	output vector in EPR model
η_i	*i*-th input vector in EPR model
θ_w	Shields' parameter
μ	dynamic viscosity of water
$\prod_{\mathrm{i}}$	i-th non-dimensional parameter
ρ	density of water
ρ_s	density of sediment
ω	set of weighting coefficients in MARS model
ω_{i}	i-th weighting coefficient in MARS model
ω_0	constant coefficient in MARS model
Ω	approximation function in MARS model
ξ_0	bias term in EPR model
ξ_j	*j*-th coefficient in EPR model

Acronyms

ADR	Average Discrepancy Ratio
AI	Artificial Intelligence
ANFIS	Adaptive Neuro-Fuzzy Inference System
ANN	Artificial Neural Network
BP	Back Propagation
CART	Classification and Regression Tree
CBO	Colliding Bodies' Optimization
CW	Clear-Water regime
EPR	Evolutionary Polynomial Regression
GEP	Gene-Expression Programming
GMDH	Group Method of Data Handling
GP	Genetic Programming
IOA	Index Of Agreement
LB	Live-Bed regime
LGP	Linear Genetic Programming
MAE	Mean Absolute Error
MARS	Multivariate Adaptive Regression Splines
MNLR	Multivariate Non-Linear Regression
MSE	Mean Square Error
cMT	Model Tree
PSO	Particle Swarm Optimization
SI	Scatter Index
SSE	Sum of Squared Errors
SVM	Support Vector Machine
WA	Wale Algorithm

Appendix A. Experiments by Cheng et al. (2014): Test Conditions and Main Results

In this Appendix the tests of Cheng et al. [9] are summarized in terms of test conditions and main results. Only the tests considered in this study are given below (Table A1). It should be noted that H is the wave height.

Table A1. Tests of [9]: main characteristics and results under wave-only conditions.

Test# [-]	α [°]	e/D [-]	H [m]	T [s]	U_w [m/s]	KC [-]	θ_w [-]	V_H [mm/s]	V_L [mm/s]	V_R [mm/s]
w15t15e1	0°	0.1	0.13	1.5	0.29	8.7	0.18	3.5	3.6	3.4
w15t15e2	0°	0.2	0.13	1.5	0.29	8.7	0.18	3.8	4.8	2.8
w15t15e3	0°	0.3	0.13	1.5	0.29	8.7	0.18	2.6	2.6	2.6
w15t15e4	0°	0.4	0.13	1.5	0.29	8.7	0.18	1.9	2.0	1.8
w16t18e1	0°	0.1	0.17	1.8	0.44	15.8	0.30	5.7	6.3	5.1
w16t18e2	0°	0.2	0.17	1.8	0.44	15.8	0.30	4.9	5.4	4.4
w16t18e3	0°	0.3	0.17	1.8	0.44	15.8	0.30	4.3	4.6	4.0
w16t18e4	0°	0.4	0.17	1.8	0.44	15.8	0.30	3.2	3.7	2.8
w16t20e1	0°	0.1	0.15	2.0	0.45	18.0	0.29	6.1	6.5	5.7
w16t20e2	0°	0.2	0.15	2.0	0.45	18.0	0.29	5.4	5.3	5.5
w16t20e3	0°	0.3	0.15	2.0	0.45	18.0	0.29	4.7	4.7	4.7
w16t20e4	0°	0.4	0.15	2.0	0.45	18.0	0.29	3.4	3.0	3.7
a15w15t15e3	15°	0.3	0.13	1.5	0.29	8.7	0.18	2.2	1.6	2.7
a15w15t15e4	15°	0.4	0.13	1.5	0.29	8.7	0.18	1.5	1.6	1.5
a15w16t18e1	15°	0.1	0.17	1.8	0.44	15.8	0.30	5.4	5.2	5.6
a15w16t18e2	15°	0.2	0.17	1.8	0.44	15.8	0.30	4.3	4.1	4.5
a15w16t18e3	15°	0.3	0.17	1.8	0.44	15.8	0.30	4.2	4.0	4.4
a15w16t18e4	15°	0.4	0.17	1.8	0.44	15.8	0.30	2.8	2.6	3.1
a15w16t20e2	15°	0.2	0.15	2.0	0.45	18.0	0.29	4.8	4.4	5.2
a15w16t20e3	15°	0.3	0.15	2.0	0.45	18.0	0.29	4.0	3.7	4.3
a15w16t20e4	15°	0.4	0.15	2.0	0.45	18.0	0.29	3.2	3.4	3.0

Table A1. *Cont.*

Test# [-]	α [°]	e/D [-]	H [m]	T [s]	U_w [m/s]	KC [-]	θ_w [-]	V_H [mm/s]	V_L [mm/s]	V_R [mm/s]
a30w15t15e1	30°	0.1	0.13	1.5	0.29	8.7	0.18	2.6	2.2	3.0
a30w15t15e2	30°	0.2	0.13	1.5	0.29	8.7	0.18	1.8	2.2	1.4
a30w16t18e1	30°	0.1	0.17	1.8	0.44	15.8	0.30	3.9	3.9	4.0
a30w16t18e2	30°	0.2	0.17	1.8	0.44	15.8	0.30	3.4	3.2	3.6
a30w16t18e3	30°	0.3	0.17	1.8	0.44	15.8	0.30	2.9	2.7	3.1
a30w16t18e4	30°	0.4	0.17	1.8	0.44	15.8	0.30	2.5	2.3	2.6
a30w16t20e1	30°	0.1	0.15	2.0	0.45	18.0	0.29	3.8	3.4	4.1
a30w16t20e2	30°	0.2	0.15	2.0	0.45	18.0	0.29	3.2	3.1	3.4
a30w16t20e3	30°	0.3	0.15	2.0	0.45	18.0	0.29	2.7	2.3	3.1
a30w16t20e4	30°	0.4	0.15	2.0	0.45	18.0	0.29	2.3	2.3	2.3
a45w15t15e1	45°	0.1	0.13	1.5	0.29	8.7	0.18	1.8	1.9	1.7
a45w15t15e2	45°	0.2	0.13	1.5	0.29	8.7	0.18	1.6	1.8	1.4
a45w16t18e1	45°	0.1	0.17	1.8	0.44	15.8	0.30	2.7	2.8	2.7
a45w16t18e2	45°	0.2	0.17	1.8	0.44	15.8	0.30	2.2	2.5	1.9
a45w16t18e3	45°	0.3	0.17	1.8	0.44	15.8	0.30	1.8	2.0	1.6
a45w16t20e1	45°	0.1	0.15	2.0	0.45	18.0	0.29	2.8	2.4	3.1
a45w16t20e2	45°	0.2	0.15	2.0	0.45	18.0	0.29	2.3	2.3	2.3

Appendix B. Implementation of the Intelligent Computing Methods: Tables and Equations

- GEP-derived equations

The parameters characterizing the GEP method for 3D-scour rate predictions are made explicit in Table A2.

Table A2. Parameters of the GEP model for 3D-scour rate predictions.

Parameters	Description of Parameters	Setting of Parameters
P_1	Function set	$+,-,\cdot,/$, Power (x^2), min(x_1, x_2), ($1-x$), *ln*(x), average(x_1, x_2), *arctan*(x), *tanh*(x)
P_2	Linking function	Addition
P_3	Mutation rate	0.003
P_4	Inversion rate	0.00546
P_5	One-point and two-point recombination rates	0.00277
P_6	Gene recombination rate	0.00277
P_7	Permutation	0.00546
P_8	Maximum tree depth	6
P_9	Number of genes	3
P_{10}	Number of chromosomes	30
P_{11}	Best fitness values	773.67 (V_H^*), 763.76 (V_R^*), 761.35 (V_L^*)

GEP models have been developed via GeneXproTools5 platform. During their development, the values of best-fit function were 1687, 1123, and 1322 for the estimation of V_H^*, V_R^*, and V_L^*, respectively. The best formulations were the following:

$$\begin{gathered} V_H^* = tanh\left(arctan\left(ln\left(\min\left(-3.8810 + KC, \left(1-\tfrac{e}{D}\right)^2\right)\right)\right)\right) + \\ \left[-KC - \tfrac{e}{D} + sin\alpha + \left(2.3133 - \tfrac{e}{D}\right)\cdot\left(1-\tfrac{e}{D}\right)\right] + min\left[tanh\left(\left(\tfrac{\theta_w}{1+sin\alpha}\right)\right)\cdot\left(\tfrac{4.1825+KC}{2}\right), \tfrac{3.9524}{(1+sin\alpha)^2}\right] \end{gathered} \tag{A1}$$

$$V_L^* = \left[-\frac{e}{D} + sin\alpha + \frac{3.5804 - sin\alpha}{2}\right] + \left\{arctan\left[\frac{1-0.5\cdot\left(2+sin\alpha-\frac{e}{D}\right)+min(8.9196-KC,\theta_w-1-sin\alpha)}{2}\right]\right\}^2 + \left[1-\left\{tanh\left(1+\frac{e}{D}+sin\alpha-\left(1-\frac{e}{D}\right)^2\right)\right\}^4\right] \quad (A2)$$

$$V_R^* = \left[1-\frac{e}{D}+\theta_w-(1+sin\alpha)^2\right]\cdot\left(sin\alpha-\frac{e}{D}\right)^{\frac{1}{3}} + \left\{tanh\left[arctan\left(\frac{KC-3.836}{2}\right)\right]+1-\frac{e}{D}\right\}^2 + 1 - \left[tanh\left(1-\frac{e}{D}\right)-2-KC+\frac{e}{D}\right]\cdot\theta_w^2 \quad (A3)$$

- EPR-derived equations

EPR models have been developed by EPR MOGA-XL software. The best formulations for the estimation of V_H^*, V_R^*, and V_L^* were the following:

$$V_H^* = 0.3235\cdot(KC)^{0.5} + 0.02161\cdot\left(1-\frac{e}{D}\right)^{0.5}KC^{0.5}{\theta_w}^2\cdot ln\left[\frac{KC^2\cdot{\theta_w}^2}{\left(1-\frac{e}{D}\right)^2\cdot(1+sin\alpha)^{0.5}}\right] + 0.00278\cdot(1+sin\alpha)^{0.5}\left(1-\frac{e}{D}\right)^{0.5}KC^{0.5}{\theta_w}^{0.5}\cdot ln\left[\frac{\left(1-\frac{e}{D}\right)^{0.5}\cdot{\theta_w}^{0.5}}{(1+sin\alpha)}\right] + 0.01438\cdot(1+sin\alpha)^2KC^{0.5}\theta_w\cdot ln\left[\left(1-\frac{e}{D}\right)^{0.5}\cdot(1+sin\alpha)^2\cdot\theta_w^2KC^{0.5}\right] \quad (A4)$$

$$V_L^* = 0.01157\cdot Ln\left[\frac{\theta_w}{KC^2\cdot(1+sin\alpha)}\right] + 0.001464\cdot ln\left[\left(1-\frac{e}{D}\right)\cdot(1+sin\alpha)\cdot\theta_w^{0.5}\right] + 0.3205\cdot\left(1-\frac{e}{D}\right) + 0.01153\cdot\frac{\left(1-\frac{e}{D}\right)^2\cdot{\theta_w}^2}{KC}\cdot ln\left[\frac{\left(1-\frac{e}{D}\right)^2\cdot{\theta_w}^2\cdot KC^2}{(1+sin\alpha)^{0.5}}\right] + 0.4768\cdot(1+sin\alpha)\cdot ln\left[\frac{1}{(1+sin\alpha)^{1.5}}\right] \quad (A5)$$

$$V_R^* = 0.000671\cdot ln\left[\frac{\theta_w}{KC^2\cdot(1+sin\alpha)^{1.5}}\right] + 0.3438\cdot KC^{0.5}\cdot\theta_w + 0.01371\cdot\frac{\left(1-\frac{e}{D}\right)^2\cdot KC}{\theta_w}\cdot ln\left[\frac{KC^{1.5}}{\left(1-\frac{e}{D}\right)\cdot(1+sin\alpha)^{1.5}\cdot\theta_w^2}\right] + 0.57711\cdot(1+sin\alpha)^{0.5}\left(1-\frac{e}{D}\right)^{0.5}\cdot ln\left[\frac{\left(1-\frac{e}{D}\right)^{0.5}}{(1+sin\alpha)^2}\right] + 0.02773\cdot(1+sin\alpha)^2\left(1-\frac{e}{D}\right)^2\cdot ln\left[\frac{\left(1-\frac{e}{D}\right)^2\cdot(1+sin\alpha)^2}{\theta_w^{0.5}}\right] \quad (A6)$$

- MARS-derived equations

The formulations for the basis functions (BFs) and their weighting coefficients associated with the estimation of 3D- scour rates are listed in Table A3.

MARS models have been developed by MATLAB2015 software. The best formulations for the estimation of V_H^*, V_R^*, and V_L^* were the following:

$$V_H^* = 3.4903 - 13.568\cdot BF_1 + 22.259\cdot BF_2 - 4.7028\cdot BF_3 - 20.026\cdot BF_4 - 3.5962\cdot BF_5 + 4.6579\cdot BF_6 \quad (A7)$$

$$V_L^* = 3.8165 - 4.4937\cdot BF_1 + 1.8162\cdot BF_2 + 3.2772\cdot BF_3 - 4.1566\cdot BF_4 - 0.20284\cdot BF_5 - 0.2084\cdot BF_6 - 7.4663\cdot BF_7 \quad (A8)$$

$$V_R^* = 3.3939 - 0.18259\cdot BF_1 + 4.3799\cdot BF_2 - 8.1073\cdot BF_3 - 4.2723\cdot BF_4 \quad (A9)$$

Table A3. Explored basis functions (BFs) in MARS models for 3D-scour rate predictions.

BFs for V_H^* **Prediction**	
BF_1	$max\{0,\ 0.29 - \theta_w\}$
BF_2	$max\{0,\ 0.29 - \theta_w\} \cdot max\{0, [(1 + sin\alpha) - 1.2587]\}$
BF_3	$max\{0,\ [0.8 - (1 - e/D)]\}$
BF_4	$max\{0, [0.8 - (1 - e/D)]\} \cdot max\{0, [1.2587 - (1 + sin\alpha)]\}$
BF_5	$max\{0, [(1 + sin\alpha) - 1.258]\}$
BF_6	$max\{0, [1.2587 - (1 + sin\alpha)]\}$
BFs for V_L^* **Prediction**	
BF_1	$max\{0,\ [(1 + sin\alpha) - 1.258]\}$
BF_2	$max\{0,\ [1.258 - (1 + sin\alpha)]\}$
BF_3	$max\{0,\ [(1 - e/D) - 0.8]\}$
BF_4	$max\{0,\ [0.8 - (1 - e/D)]\}$
BF_5	$max\{0,\ 15.8 - KC\}$
BF_6	$max\{0,\ 15.8 - KC\} \cdot max\{0, [(1 + sin\alpha) - 1.258]\}$
BF_7	$max\{0, [0.8 - (1 - e/D)]\} \cdot max\{0, [1.2587 - (1 + sin\alpha)]\}$
BFs for V_R^* **Prediction**	
BF_1	$max\{0,\ 15.8 - KC\}$
BF_2	$max\{0,\ [(1 - e/D) - 0.7]\}$
BF_3	$max\{0,\ [0.7 - (1 - e/D)]\}$
BF_4	$max\{0,\ [(1 + sin\alpha) - 1.258]\}$

- M5MT-derived equations

Tables A4 and A5 summarize, respectively, M5 rules and associated linear equation returned by Model Tree for the prediction of V_H^*. 12 rules were provided and, additionally, all the four input variables were used as splitting parameters.

Table A4. List of explored rules returned by M5MT model to predict V_H^*.

```
If 1+ sinα <= 1.379:
|   If 1   e/D <= 0.75:
|   |   If KC <= 12.25: LM#1
|   |   If KC > 12.25:
|   |   |   If 1   e/D <= 0.65: LM#2
|   |   |   If 1   e/D > 0.65: LM#3
|   If 1   e/D > 0.75:
|   |   If 1+ sinα <= 1.129:
|   |   |   If θw <= 0.295: LM#4
|   |   |   If θw > 0.295: LM#5
|   |   If 1+ sinα > 1.129: LM#6
If 1+ sinα > 1.379:
|   If KC <= 12.25:
|   |   If 1+ sinα <= 1.603: LM#7
|   |   If 1+ sinα > 1.603: LM#8
|   If KC > 12.25:
|   |   If 1-e/D <= 0.65: LM#9
|   |   If 1- e/D > 0.65:
|   |   |   If 1+ sinα <= 1.603:
|   |   |   |   If 1   e/D <= 0.75: LM#10
|   |   |   |   If 1- e/D > 0.75: LM#11
|   |   |   If 1+ sinα > 1.603: LM#12
```

Table A5. Linear equations obtained by M5MT model to predict V_H^*.

	Linear Equations by M5MT Model
#1	$V_H^* = -1.5478 - 2.1346 \cdot (1 + sin\alpha) + 5.5576 \cdot \left(1 - \frac{e}{D}\right) + 0.0246 \cdot KC + 9.5079 \cdot \theta_w$
#2	$V_H^* = -1.602 - 2.1346 \cdot (1 + sin\alpha) + 5.811 \cdot \left(1 - \frac{e}{D}\right) + 0.0211 \cdot KC + 9.5079 \cdot \theta_w$
#3	$V_H^* = -1.602 - 2.1346 \cdot (1 + sin\alpha) + 5.811 \cdot \left(1 - \frac{e}{D}\right) + 0.0211 \cdot KC + 9.5079 \cdot \theta_w$
#4	$V_H^* = -0.4133 - 2.793 \cdot (1 + sin\alpha) + 4.6905 \cdot \left(1 - \frac{e}{D}\right) + 12.512 \cdot \theta_w$
#5	$V_H^* = -0.4133 - 2.793 \cdot (1 + sin\alpha) + 4.6905 \cdot \left(1 - \frac{e}{D}\right) + 12.5124 \cdot \theta_w$
#6	$V_H^* = -0.2965 - 2.8258 \cdot (1 + sin\alpha) + 4.6905 \cdot \left(1 - \frac{e}{D}\right) + 12.5124 \cdot \theta_w$
#7	$V_H^* = -1.2191 - 2.4643 \cdot (1 + sin\alpha) + 3.5315 \cdot \left(1 - \frac{e}{D}\right) + 0.0209 \cdot KC + 5.8612 \cdot \theta_w$
#8	$V_H^* = -1.2191 - 2.4643 \cdot (1 + sin\alpha) + 3.5315 \cdot \left(1 - \frac{e}{D}\right) + 0.0209 \cdot KC + 5.8612 \cdot \theta_w$
#9	$V_H^* = 1.3753 - 2.6322 \cdot (1 + sin\alpha) + 3.8196 \cdot \left(1 - \frac{e}{D}\right) + 0.0166 \cdot KC + 5.8612 \cdot \theta_w$
#10	$V_H^* = 1.3642 - 2.6582 \cdot (1 + sin\alpha) + 3.8972 \cdot \left(1 - \frac{e}{D}\right) + 0.0166 \cdot KC + 5.8612 \cdot \theta_w$
#11	$V_H^* = 1.3649 - 2.6582 \cdot (1 + sin\alpha) + 3.8972 \cdot \left(1 - \frac{e}{D}\right) + 0.0166 \cdot KC + 5.8612 \cdot \theta_w$
#12	$V_H^* = 1.4057 - 2.6653 \cdot (1 + sin\alpha) + 3.8509 \cdot \left(1 - \frac{e}{D}\right) + 0.0166 \cdot KC + 5.8612 \cdot \theta_w$

Tables A6 and A7 show that in the case of V_L^* estimation, 12 rules were obtained from M5 analysis with the splitting of the parameters $1 + \sin\alpha$, $1 - e/D$, and KC.

Table A6. List of explored rules returned by M5MT model to predict V_L^*.

Rule#1
IF 1+$sin\alpha$ <= 1.603, 1 e/D <= 0.65, KC > 12.25, 1+ $sin\alpha$ <= 1.379
THEN LM#1
Rule#2
IF 1+ $sin\alpha$ <= 1.603, 1 e/D > 0.65, KC <= 12.25, 1+ $sin\alpha$ <= 1.129
THEN LM#2
Rule#3
IF KC > 12.25, 1+ $sin\alpha$ <= 1.379, 1 e/D <= 0.75
THEN LM#3
Rule#4
IF KC > 12.25, 1+ $sin\alpha$ <= 1.379
THEN LM#4
Rule#5
IF KC > 12.25, 1+ $sin\alpha$ > 1.603, 1 e/D <= 0.85
THEN LM#5
Rule#6
IF KC <= 12.25, 1+ $sin\alpha$ <= 1.603, 1 e/D <= 0.65
THEN LM#6
Rule#7
IF KC <= 12.25, 1+ $sin\alpha$ <= 1.603
THEN LM#7
Rule#8
IF 1+ $sin\alpha$ <= 1.603, 1 e/D <= 0.7
THEN LM#8
Rule#9
IF KC <= 12.25
THEN LM#9
Rule#10
IF 1+ $sin\alpha$ <= 1.603, 1 e/D <= 0.85
THEN LM#10
Rule#11
IF 1+ $sin\alpha$ <= 1.603
THEN LM#11 = 3.7233×(1+ $sin\alpha$)+8.642
Otherwise
Rule#12 LM#12

Table A7. Linear equations obtained by M5MT model to predict V_L^*.

	Linear Equations by M5MT Model
#1	$V_L^* = 0.4207 - 2.2586 \cdot (1 + sin\alpha) + 4.0342 \cdot \left(1 - \frac{e}{D}\right) + 0.1384 \cdot KC$
#2	$V_L^* = 0.71 - 2.5502 \cdot (1 + sin\alpha) + 4.1387 \cdot \left(1 - \frac{e}{D}\right) + 0.1501 \cdot KC$
#3	$V_L^* = 2.2943 - 2.8549 \cdot (1 + sin\alpha) + 4.6573 \cdot \left(1 - \frac{e}{D}\right) + 0.0646 \cdot KC$
#4	$V_L^* = 3.1812 - 3.2497 \cdot (1 + sin\alpha) + 4.1348 \cdot \left(1 - \frac{e}{D}\right) + 0.0691 \cdot KC$
#5	$V_L^* = 2.1734 - 2.5643 \cdot (1 + sin\alpha) + 3.8075 \cdot \left(1 - \frac{e}{D}\right) + 0.0639 \cdot KC$
#6	$V_L^* = 0.6837 - 1.3108 \cdot (1 + sin\alpha) + 2.6787 \cdot \left(1 - \frac{e}{D}\right) + 0.0823 \cdot KC$
#7	$V_L^* = 1.9259 - 2.0976 \cdot (1 + sin\alpha) + 2.7366 \cdot \left(1 - \frac{e}{D}\right) + 0.0829 \cdot KC$
#8	$V_L^* = 3.6165 - 3.1164 \cdot (1 + sin\alpha) + 3.6362 \cdot \left(1 - \frac{e}{D}\right) + 0.046 \cdot KC$
#9	$V_L^* = 4.4871 - 3.7599 \cdot (1 + sin\alpha) + 3.2793 \cdot \left(1 - \frac{e}{D}\right) + 0.0609 \cdot KC$
#10	$V_L^* = 4.2091 - 3.3313 \cdot (1 + sin\alpha) + 4.2603 \cdot \left(1 - \frac{e}{D}\right)$
#11	$V_L^* = 8.462 - 3.7233 \cdot (1 + sin\alpha)$
#12	$V_L^* = 2.1843$

13 rules in terms of linear regression equations were extracted from M5 to predict V_R^*, as indicated in Tables A8 and A9.

Table A8. List of explored rules returned by M5MT model to predict V_R^*.

If KC <= 12.25:
| If 1 e/D <= 0.85:
| | If 1+ $sin\alpha$ <= 1.379: LM#1
| | If 1+ $sin\alpha$ > 1.379: LM#2
| If 1 e/D > 0.85: LM#3
If KC > 12.25:
| If 1+ $sin\alpha$ <= 1.379:
| | If 1 e/D <= 0.75:
| | | If 1 e/D <= 0.65: LM#4
| | | If 1 e/D > 0.65:
| | | | If 1+ $sin\alpha$ <= 1.129: LM#5
| | | | If 1+ $sin\alpha$ > 1.129: LM#6
| | If 1 e/D > 0.75:
| | | If KC <=16.9: LM#7
| | | If KC > 16.9: LM#8
| If 1+ $sin\alpha$ > 1.379:
| | If 1+ $sin\alpha$ <= 1.603:
| | | If 1 e/D <= 0.85:
| | | | If 1 e/D <= 0.75: LM#9
| | | | If 1 e/D > 0.75: LM#10
| | | If 1 e/D > 0.85: LM#11
| | If 1+ $sin\alpha$ > 1.603:
| | | If 1 e/D <= 0.85: LM#12
| | | If 1 e/D > 0.85: LM#13

Table A9. Linear equations obtained by M5MT model to predict V_R^*.

	Linear Equations by M5MT Model
#1	$V_R^* = 0.3424 - 2.154 \cdot (1 + sin\alpha) + 4.0823 \cdot \left(1 - \frac{e}{D}\right) + 0.1151 \cdot KC$
#2	$V_R^* = 0.3424 - 2.1582 \cdot (1 + sin\alpha) + 4.0823 \cdot \left(1 - \frac{e}{D}\right) + 0.1151 \cdot KC$
#3	$V_R^* = 0.2434 - 2.1481 \cdot (1 + sin\alpha) + 4.2686 \cdot \left(1 - \frac{e}{D}\right) + 0.1151 \cdot KC$
#4	$V_R^* = 0.4536 - 2.0057 \cdot (1 + sin\alpha) + 4.9602 \cdot \left(1 - \frac{e}{D}\right) + 0.0951 \cdot KC$
#5	$V_R^* = 0.5106 - 2.0057 \cdot (1 + sin\alpha) + 4.8913 \cdot \left(1 - \frac{e}{D}\right) + 0.0951 \cdot KC$
#6	$V_R^* = 0.5106 - 2.0057 \cdot (1 + sin\alpha) + 4.8913 \cdot \left(1 - \frac{e}{D}\right) + 0.0951 \cdot KC$
#7	$V_R^* = 0.902 - 2.0057 \cdot (1 + sin\alpha) + 4.36 \cdot \left(1 - \frac{e}{D}\right) + 0.0964 \cdot KC$
#8	$V_R^* = 0.9056 - 2.0057 \cdot (1 + sin\alpha) + 4.36 \cdot \left(1 - \frac{e}{D}\right) + 0.0964 \cdot KC$
#9	$V_R^* = 2.4404 - 2.8991 \cdot (1 + sin\alpha) + 4.341 \cdot \left(1 - \frac{e}{D}\right) + 0.0685 \cdot KC$
#10	$V_R^* = 2.4387 - 2.8991 \cdot (1 + sin\alpha) + 4.3454 \cdot \left(1 - \frac{e}{D}\right) + 0.0685 \cdot KC$
#11	$V_R^* = 1.3649 - 2.8991 \cdot (1 + sin\alpha) + 4.3111 \cdot \left(1 - \frac{e}{D}\right) + 0.0685 \cdot KC$
#12	$V_R^* = 2.5596 - 3.0402 \cdot (1 + sin\alpha) + 4.3839 \cdot \left(1 - \frac{e}{D}\right) + 0.0685 \cdot KC$
#13	$V_R^* = 2.5652 - 3.0402 \cdot (1 + sin\alpha) + 4.3839 \cdot \left(1 - \frac{e}{D}\right) + 0.0685 \cdot KC$

References

1. Xie, L.; Zhu, Y. Scour protection of underwater pipelines. *J. Civ. Eng. Constr.* **2018**, *7*, 171–177. [CrossRef]
2. Sumer, B.M.; Fredsøe, J. *The Mechanics of Scour in the Marine Environment*; Advanced Series on Ocean Engineering; World Scientific Publishing: Singapore, 2002; Volume 17.
3. Cevik, E.; Yüksel, Y. Scour under submarine pipelines in waves in shoaling conditions. *J. Waterw. Port Coast. Ocean Eng.* **1999**, *125*, 9–19. [CrossRef]
4. Hansen, E.A. Scour below pipelines and cables: A simple model. In Proceedings of the 11th Offshore Mechanics and Arctic Engineering Conference, ASME, Calgary, AB, Canada, 7–12 June 1992; Volume V-A, pp. 133–138.
5. Kumar, A.V.; Neelamani, S.; Rao, S.N. Wave pressures and uplift forces on and scour around submarine pipeline in clayey soil. *Ocean Eng.* **2003**, *30*, 271–295. [CrossRef]
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. Yang, L.; Shi, B.; Guo, Y.; Wen, X. Calculation and experiment on scour depth for submarine pipeline with a spoiler. *Ocean Eng.* **2012**, *55*, 191–198. [CrossRef]
8. Wu, Y.; Chiew, Y.-M. Three-dimensional scour at submarine pipelines. *J. Hydraul. Eng.* **2012**, *138*, 788–795. [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. Yang, L.; Shi, B.; Guo, Y.; Zhang, L.; Zhang, J.; Han, Y. Scour protection of submarine pipelines using rubber plates underneath the pipes. *Ocean Eng.* **2014**, *84*, 176–182. [CrossRef]
11. Yusoff, M.A.M.; Azamathulla, H.M.; Ghani, A.A. Time variations of scour below submerged skewed pipelines. *IOP Conf. Series: Mater. Sci. Eng.* **2016**, *136*, 012071. [CrossRef]
12. Dogan, M.; Aksoy, A.O.; Arisoy, Y.; Guney, M.S.; Abdi, V. Experimental investigation of the equilibrium scour depth below submerged pipes both in live-bed and clear-water regimes under the wave effect. *Appl. Ocean Res.* **2018**, *80*, 49–56. [CrossRef]
13. Yang, S.; Shi, B.; Guo, Y.; Yang, L. Investigation on scour protection of submarine piggyback pipeline. *Ocean Eng.* **2019**, *182*, 442–450. [CrossRef]
14. Zang, Z.; Tang, G.; Chen, Y.; Cheng, L.; Zhang, J. Predictions of the equilibrium depth and time scale of local scour below a partially buried pipeline under oblique currents and waves. *Coast. Eng.* **2019**, *150*, 94–107. [CrossRef]
15. Azamathulla, H.M.; Zakaria, N.A. Prediction of scour below submerged pipeline crossing a river using ANN. *Water Sci. Technol.* **2011**, *63*, 2225–2230. [CrossRef] [PubMed]
16. Etemad-Shahidi, A.; Yasa, R.; Kazeminezhad, M. Prediction of wave-induced scour depth under submarine pipelines using machine learning approach. *Appl. Ocean Res.* **2011**, *33*, 54–59. [CrossRef]
17. Zanganeh, M.; Yeganeh-Bakhtiary, A.; Bakhtyar, R. Combined particle swarm optimization and fuzzy inference system model for estimation of current-induced scour beneath marine pipelines. *J. Hydroinform.* **2010**, *13*, 558–573. [CrossRef]
18. Parsaie, A.; Haghiabi, A.H.; Moradinejad, A. Prediction of scour depth below river pipeline using support vector machine. *KSCE J. Civ. Eng.* **2019**, *23*, 2503–2513. [CrossRef]
19. Yasa, R.; Shahidi, A.E. Classification and regression trees approach for predicting current-induced scour depth under pipelines. *J. Offshore Mech. Arct. Eng.* **2013**, *136*, 011702. [CrossRef]
20. Azamathulla, H.M.; Yusoff, M.A.M. Soft computing for prediction of river pipeline scour depth. *Neural Comput. Appl.* **2013**, *23*, 2465–2469. [CrossRef]
21. Najafzadeh, M.; Sarkamaryan, S. Extraction of optimal equations for evaluation of pipeline scour depth due to currents. *Proc. Inst. Civ. Eng. Mar. Eng.* **2018**, *171*, 1–10. [CrossRef]

22. Azamathulla, H.M.; Ghani, A.A. Genetic programming to predict river pipeline scour. *J. Pipeline Syst. Eng. Pract.* **2010**, *1*, 127–132. [CrossRef]
23. Najafzadeh, M.; Barani, G.-A.; Azamathulla, H.M. Prediction of pipeline scour depth in clear-water and live-bed conditions using group method of data handling. *Neural Comput. Appl.* **2012**, *24*, 629–635. [CrossRef]
24. Najafzadeh, M.; Barani, G.-A.; Kermani, M.R.H. Estimation of pipeline scour due to waves by GMDH. *J. Pipeline Syst. Eng. Pract.* **2014**, *5*, 06014002. [CrossRef]
25. Haghiabi, A.H. Prediction of river pipeline scour depth using multivariate adaptive regression splines. *J. Pipeline Syst. Eng. Pract.* **2017**, *8*, 04016015. [CrossRef]
26. Najafzadeh, M.; Saberi-Movahed, F. GMDH-GEP to predict free span expansion rates below pipelines under waves. *Mar. Georesources Geotechnol.* **2019**, *37*, 375–392. [CrossRef]
27. 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]
28. Moncada-M, A.T.; Aguirre-Pe, J. Scour below pipeline in river crossings. *J. Hydraul. Eng.* **1999**, *125*, 953–958. [CrossRef]
29. Azamathulla, H.M.; Guven, A.; Demir, Y.K. Linear genetic programming to scour below submerged pipeline. *Ocean Eng.* **2011**, *38*, 995–1000. [CrossRef]
30. Sharafati, A.; Yasa, R.; Azamathulla, H.M. Assessment of stochastic approaches in prediction of wave-induced pipeline scour depth. *J. Pipeline Syst. Eng. Pract.* **2018**, *9*, 04018024. [CrossRef]
31. Wu, Y.; Chiew, Y.-M. Mechanics of three-dimensional pipeline scour in unidirectional steady current. *J. Pipeline Syst. Eng. Pract.* **2013**, *4*, 3–10. [CrossRef]
32. Ferreira, C. Gene expression programming: A new adaptive algorithm for solving problems. *Complex Syst.* **2001**, *13*, 87–129.
33. Giustolisi, O.; Savic, D.A. A symbolic data-driven technique based on evolutionary polynomial regression. *J. Hydroinform.* **2006**, *8*, 207–222. [CrossRef]
34. Berardi, L.; Giustolisi, O.; Kapelan, Z.; Savic, D.A. Development of pipe deterioration models for water distribution systems using EPR. *J. Hydroinform.* **2008**, *10*, 113–126. [CrossRef]
35. Giustolisi, O.; Savic, D.A. Advances in data-driven analyses and modelling using EPR-MOGA. *J. Hydroinform.* **2009**, *11*, 225–236. [CrossRef]
36. Savic, D.A.; Giustolisi, O.; Laucelli, D. Asset deterioration analysis using multi-utility data and multi-objective data mining. *J. Hydroinform.* **2009**, *11*, 211–224. [CrossRef]
37. Friedman, J.H. Multivariate adaptive regression splines. *Ann. Stat.* **1991**, *19*, 1–67. [CrossRef]
38. 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; Adams & Sterling, Ed.; World Scientific: Singapore, 1992; pp. 343–348.