12-2019

A Machine Learning Assessment to Predict the Sediment Transport Rate Under Oscillating Sheet Flow Conditions

Huy Vu
University of New Orleans

A MACHINE LEARNING ASSESSMENT TO PREDICT THE SEDIMENT TRANSPORT RATE UNDER OSCILLATING SHEET FLOW CONDITIONS

An Honors Thesis

Presented to

the Department of Computer Science

of the University of New Orleans

In Partial Fulfillment

of the Requirements for the Degree of

Bachelor of Science, with University High Honors

and Honors in Computer Science

by

Huy Vu

December 2019

Acknowledgments

I am grateful to work on this project and gain insightful experience in machine learning and academic research. The process of researching and writing a thesis is strenuous but rewarding. This thesis will never become a possibility without the encouragement, expertise, and feedback from Dr. Md Tamjidul Hoque and Dr. Mahdi Abdelguerfi, who are my thesis advisors and my dearest professors. I would like to express my sincere thanks of gratitude to them for giving me their invaluable guidance, computing resources, and unfailing support throughout the research-writing process.

I would like to thank Dr. Elias Ioup for introducing me to the research topic. I would like to send many thanks to Dr. Margaret Palmsten and Dr. Sam Bateman for your time computing the datasets and your clear explanations for my questions on the research topic.

I would like to give a huge thanks to my graduate friend Pujan Pokhrel for offering his assistance and expert advice on machine learning and academic writing.

Finally, I would like to express my gratitude towards my family who always supports me financially and emotionally.

Table of Contents

List of Tables

List of Figures

Abstract

The two-phase flow approach has been the conventional method designed to study the sediment transport rate. Due to the complexity of sediment transport, the precisely numerical models computed from that approach require initial assumptions and, as a result, may not yield accurate output for all conditions. This research work proposes that Machine Learning algorithms can be an alternative way to predict the processes of sediment transport in two-dimensional directions under oscillating sheet flow conditions, by utilizing the available dataset of the SedFoam multidimensional two-phase model. The assessment utilized linear regression and gradient boosting algorithm to analyze the lowest average mean squared error in each case and search for the best partition method based on the domain height of the simulation setup.

Keywords: Machine Learning, Sediment Transport, Two-phase Models Analysis, Linear Regression, Gradient Boosting

1. Introduction

1.1. Background

The coastal region is subjected to erosion due to many factors such as the worsen sea-level rise, wave-induced currents, and storms. Erosion is the direct consequence of sediment transport, which is the movement of granular particles caused mainly by the fluid movement and wind flow [1]. The scientific research on sediment transport is key to understand and solve erosion-related problems that occur near the coastal areas. Such issues can be a significant threat to the coastal infrastructure as the coastline recession damages the beachfront properties, to the coastal economy as the tourism value decreases due to the loss of beaches, and to the local coastal community as erosion swallows the land forcing people to move to the interior regions.

Driven by the technological revolution, the study of sediment transport in the fluid is evolving as new tools lead to new approaches and methods to predict sediment transport. Advanced numerical models for sediment movement forecasting are improving with more accurate model formulation utilizing the parameterization of hydrodynamic forces [2]. For mild wave conditions, parameterization of bottom shear stress and parameterization of the total bed shear stress and the total sediment flux were used to establish the model [3-4]. However, such methods were less accurate for extreme wave events, since other factors can become dominant for bed destabilization. In the past, various numerical models [5-11] partitioned the sediment transport into components: bedload, the concentrated region of sediment transport, and suspended load, part of downstream-carried sediment, since bedload transport rates of deposit are much challenging to measure than suspended load rates [12].

In recent years, various researchers have employed the multiphase flow approach to generalize the models for sediment transport without having to divide it into two components. In particular, the two-phase models of sediments and waters follow two schemes: Eulerian [13-18] and Lagrangian [19-22]. They work by defining the parameters of particle motions and the relationships between them. The Lagrangian approach focuses on the individual particle movement, while the Eulerian approach emphasizes the flow at a specific point in space as a function of time [23]. Most two-phase models are based on Reynolds-averaged approach, which is the time-averaging method to reduce the range of scales into one-dimensional-vertical (1DV) formation [24]. 1DV models failed to capture more complex conditions that involve multi-dimensional fluid and particle interactions such as turbulence. More advanced two-phase models [25-28] have been extensively researched and developed to account for the multidimensional aspect of sediment transports.

An alternative approach to the conventional sediment transport model is using well-established Machine Learning (ML) algorithms. Instead of focusing on the genesis of sediment transport, ML algorithms can find the patterns and structures of data to produce a generic mathematical model. Such algorithms are depended solely on the training dataset [29]. The generic model is the basis for predicting the output quickly from input data. Researches in various fields have been utilizing ML, such as gene prediction [30], natural-language processing [31], image recognition [32], and so on. The study, carried by [33], utilizes two supervised ML methods, artificial neural network and model trees, to model the bedload and total load transport showed better accuracy than some bedload models [34-38]. ML models in [39] predicting the bed-load transport rates in gravel-bed streams and rivers in Idaho showed “superior” results comparing to some well-known bedload formulae.

The purpose of this paper is to apply the available dataset in the SedFoam model [28] on assessing the effectiveness of ML algorithms to predict sediment transport and of partitioning the dataset. Finding the most optimal way for partitioning will assist further ML model formulation. Model assessment can prevent overfitting when the model is not generalized enough to capture unseen data, and underfitting when the model cannot capture the underlying pattern of the data. I used linear regression and gradient boosting estimators to construct the models. Linear regression is a standard and fast estimator that is wildly used in physics. Gradient boosting is a nonlinear approach to estimate the outputs. It can fit more complicated data into a model. To further evaluate the dataset, I performed several splitting tests. I first split the dataset into even parts where each part was group by its height. I also split the dataset into three sections with various sizes. In each scenario, I recorded the average mean squared error of the models for the accuracy analysis. Partitioning the dataset by the domain height is essential since different height regions will have different initial concentration field.

1.2. Related Work

The SedFoam model derives the idea from the two-phase model. The model was able to resolve processes of multi-dimensional sediment transport and validated for 2DV Reynolds-averaged condition. Under various simulations for oscillatory sheet flow conditions, the model shows multiple agreements between the computed experimental sediment concentrations. However, the model cannot capture the burst events during the flow reversal and outputs a higher flow velocity on the positive flow peak of the free-stream velocity. The inaccuracy is mainly caused by the limitation in describing the model with a precise mathematical formulation. Hence, the alternative approach using ML models should be considered and studied.

2. Methods and data

2.1 Computational Domain Configuration

Two-dimensional (2D) simulation for oscillatory sheet conditions, that is set up by [28], is similar to Figure 1 but omits the spanwise direction by setting it the 'empty' boundary condition. The total domain height L_z is 0.5 m. The initial sediment bed depth is set to be h_b = 0.1 m. The medium sand's diameter has a value of d = 0.28 mm. The training dataset utilized 360 grid points along the vertical direction and 100 grid points for the horizontal.

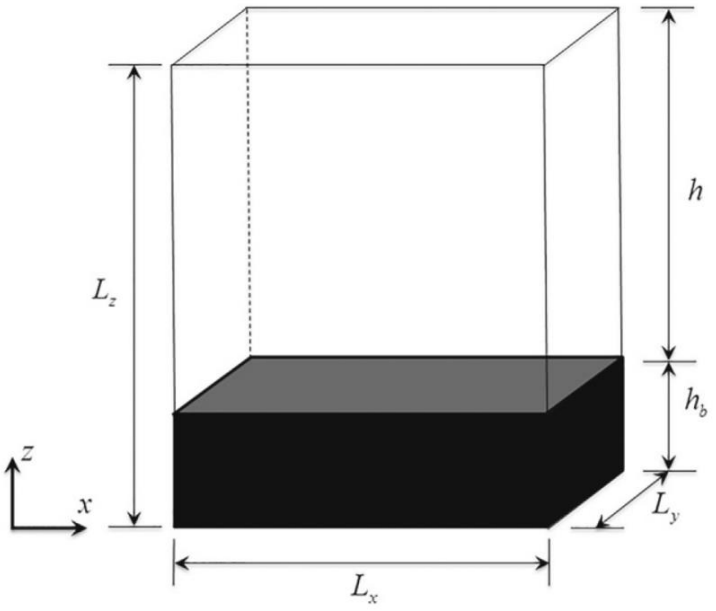

Fig. 1. Sketch of simulation setup. Reprinted from the 3D simulation in [28]. x: streamwise direction. y: spanwise direction. z: vertical direction. L_x: length of the grid in the streamwise direction. L_y: length of the grid in the spanwise direction. L_z: length of the grid in the vertical direction. h_b: initial sediment bed depth. h: initial water depth. The shaded grid represents the sediment. The unshaded grid represents the water.

2.2 Oscillatory Sheet Flow Setup

In simulation [28], the wave period is $T = 5$ s. The oscillatory flow is created by a mean streamwise pressure gradient. The asymmetric flow is based on the second-order Stokes wave motion ([28]). The maximum and minimum free-stream velocities are calculated as U_{max} and U_{min} with $U_{max} = 1.5\ m/s$ and $U_{min} = 1.5\ m/s$. The flow asymmetry is defined as $a = U_{max}/(U_{max} - U_{\min}) = 0.63$.

2.3 Dataset Format

The dataset is from the 2D SedFoam model that runs with an initially flatbed. The dataset showed multiple agreements between the computed dataset with the measured one in [40]. According to [28], the oscillatory flow dataset contains the following parameters:

$$\mathrm{t, z, x, \phi}, u_x^f, u_z^f, u_x^s, u_z^s \tag{1}$$

where t is the time at which the output is computed; z is the position of the grid point on the vertical axis; x is the position of the grid point on the horizontal axis; ϕ is the sediment concentration; u_x^f is the fluid free-stream velocity in the horizontal direction; u_z^f is the fluid free-stream velocity in the vertical direction; u_x^s is the sediment free-stream velocity in the horizontal direction; and u_z^s is the sediment free-stream velocity in the vertical direction. The input variables are t, z, and x. The output variables are ϕ, u_x^f, u_z^f, u_x^s, and u_z^s. Five output variables are grouped individually with the same set of input variables to create different scenarios. A derived model from each scenario can only predict one output variable. There are

36,000,000 data samples in every scenario. The descriptive statistics of the dataset is described in Table 1 below.

Table 1. Descriptive statistics of the variables.

	t (s)	z (m)	x (m)	ϕ	u_x^f (m/s)	u_z^f(m/s)	u_x^s (m/s)	u_z^s (m/s)
Median	25.03	0.09	0.2	0.067	0.00	0.00	0.00	0.00
Mean	25.03	0.13	0.20	0.25	0.00	0.00	0.00	-0.01
Std.	14.43	0.13	0.12	0.28	0.80	0.04	0.80	0.05
min	0.05	0.00	0.00	0.00	-5.18	-4.26	-5.12	-2.56
max	50.00	0.50	0.40	0.62	3.78	2.22	3.75	12.39

2.4 Data Transformation

Data preparation is crucial in Machine Learning. It ensures the dataset has a desired variance and bias. The implementation for this analysis is summarized as follows: In the beginning, datasets are loaded and transformed into a data frame using Pandas library. A data frame is a two-dimensional data structure with its columns representing the input and output parameters. For every output parameter, a new data frame is constructed so that it contains all input parameters and only one output. The domain height is partitioned in two main ways: the even-sized *k*-partitions (see Section 3.1) and various-sized tripartition of the height domain (see Section 3.2). The new data frame is produced based on the domain height group which is explained in the later section. A loop runs through each row in the new data frame to check which height group the dataset (for that row) belongs to. The resulted outcome contains an array

of data frames of different height categories. Each dataset within that array is divided into training and testing sets. The training set is used to train the model so its result model can then be validated against the testing set. I archive this by using *k*-fold cross-validation method, which divides the dataset into k subsets and repeats the holdout method for k times. In each time, pick one subset out of k subsets as the testing set. The remaining subsets ($k-1$ subsets) become training set. I can then average the prediction scores of all subsets. As a conventional rule, k is set to 5 since it empirically outputs the best error estimates.

2.5 Linear Regression

Multiple linear regression was chosen in this thesis to estimate how a well polynomial model can fit the data. It considers most process to be linear in physics. It can quickly produce a model and has been employed to model sediment transport in recent studies. [42] developed a total bed material equation for Malaysia rivers using linear regression that was resulted in outperforming the commonly used models: Graf, Yang, and Acker-White. Regression is the most well-known ML algorithm for modeling in supervised learning. It is an established statistical technique that produces a model that simulates the relationship of the independent variable, called feature, and the dependent variable, called response variable. Simple linear regression is the set of processes for predicting the polynomial relationship between one feature and a response variable. Multiple linear regression employs two or more features and a response variable to model their linear relationship. According to [41], a multiple linear regression model for k observations is defined as:

$$\mathrm{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \epsilon \qquad (2)$$

where $\hat{y}$ is the expected value, ϵ is the model's error term, and $\beta_0, \beta_1, \beta_2, \ldots, \beta_k$ are the unknown coefficients. In linear regression, the coefficients are estimated and denoted by $\hat{\beta}_i, i = 0, \ldots, k$. They are used to compute the predicted response value:

$$\widehat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \cdots + \hat{\beta}_k x_{ik} \tag{3}$$

The coefficient of determination, denoted as R^2, measures the rate of variation in the response variable that is predictable from the independent one that is computed as:

$$R^2 = \frac{\sum_{j=1}^{n} \left(\widehat{y}_i - \bar{y}\right)^2}{\sum_{j=1}^{n} (y_i - y)^2} \tag{4}$$

where n is the number of observations. It helps to evaluate the model's performance. The closer it gets to 1, the better the predictions fit the data.

2.6 Gradient Boosting

This thesis uses XGBoost (eXtreme Gradient Boost) library as an estimator for its being robust to noise, nonlinear, and a tree-based method, which takes less time to run. It is an improved method based on the gradient boosting algorithm, which was proposed by [43]. Gradient boosting is the algorithm that converts the ensembles of weak learners into strong ones [44]. It archived such conversion by iteratively generating new models. Weak learners are decision trees in gradient boosting. In each stage, a procedure is similar to gradient descent is performed. A new model is constructed by adding trees to the tree ensemble model so that it minimizes the loss of the model.

In XGBoost, the established learning objective that measures the performance was established in [45] for a collection F of k trees as follows:

$$\mathcal{L}(\phi) = \sum_{i} l\left(\hat{y}_i, y_i\right) + \sum_{k} \Omega\left(f_k\right)) \tag{5}$$

where l is the loss function that measures the difference between the target y_i and the prediction $\hat{y}_i$. The predicted value $\hat{y}_i$ is computed as below:

$$\hat{y}_i = \sum_{i=1}^{k} f_k(x_i)\,, f_k \in F \tag{6}$$

The second term Ω in (5) helps regularize the complexity of the model to avoid over-fitting. It is defined as below:

$$\Omega(f_k) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2 \tag{7}$$

where T represents the number of leaves in the tree. f_k stands for the independent tree structure and leaf weight w. γ is the complexity in each leaf. λ is the parameter to scale the penalty.

2.7 Statistical Analysis

The *Mean Squared Error* (MSE) was employed to determine the accuracy of the model. MSE measures the average squared difference between the outcome value from the model ($\hat{\mathrm{Y}}_i$) and the actual value (Y_i). For n predictions, the MSE is defined as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(\mathrm{Y}_i - \hat{\mathrm{Y}}_i\right)^2 \tag{8}$$

Note that the result of MSE is always non-negative and the model is more accurate as MSE values get closer to zero.

3. Results and Discussion

3.1 Even-sized *k*-partitions of domain heights

In this partitioning method, the domain height is split into even k parts (k ranges from 1 to 10), or k categories, according to the partition size. The shifts in the MSE values as the partition size increases of the predicted concentration profile, horizontal fluid velocity, vertical fluid velocity, horizontal sediment velocity, and vertical sediment velocity are shown in Figures 2-5 for linear regression and Figures 7-11 for gradient boosting respectively.

3.1.1 Linear Regression

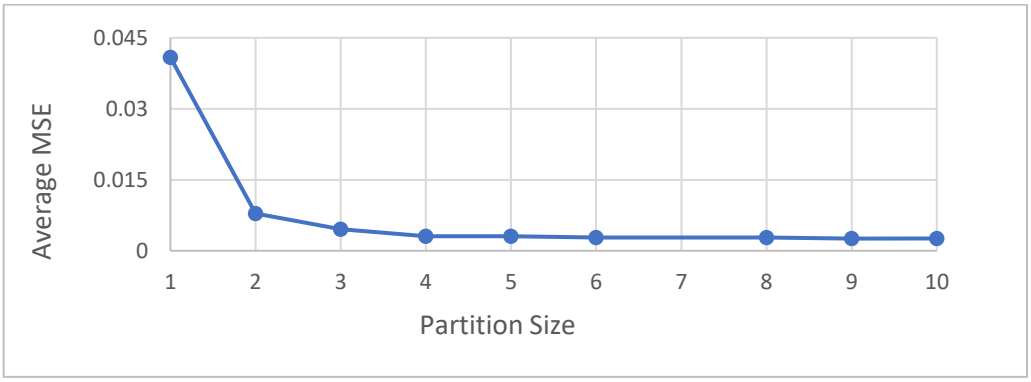

Fig. 2. Variation of average prediction errors of linear-regression-based models in relation to even-sized *k*-partition for the concentration profile.

From Figure 2, the average MSE values decrease rapidly when the partition size goes from 1 to 3 then stays relatively constant afterward.

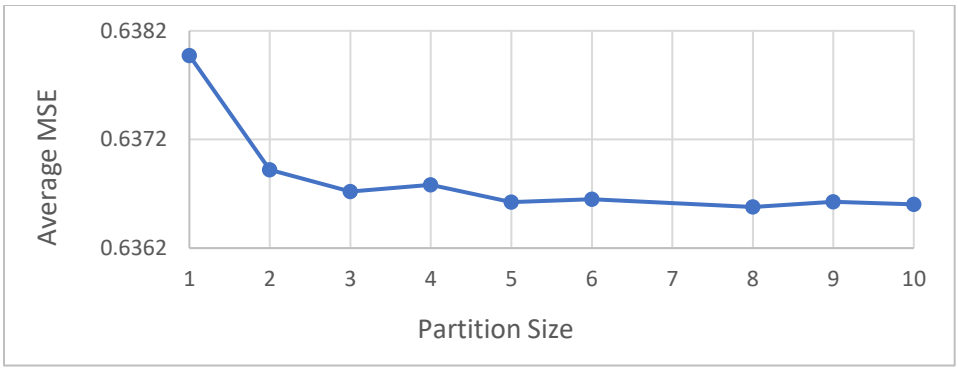

Fig. 3. Variation of average prediction errors of linear-regression-based models in relation to even-sized *k*-partition for the horizontal fluid velocity.

From Figure 3, the average MSE values dropdown steadily till 3 partitions and then show some fluctuations but remain relatively constant afterward.

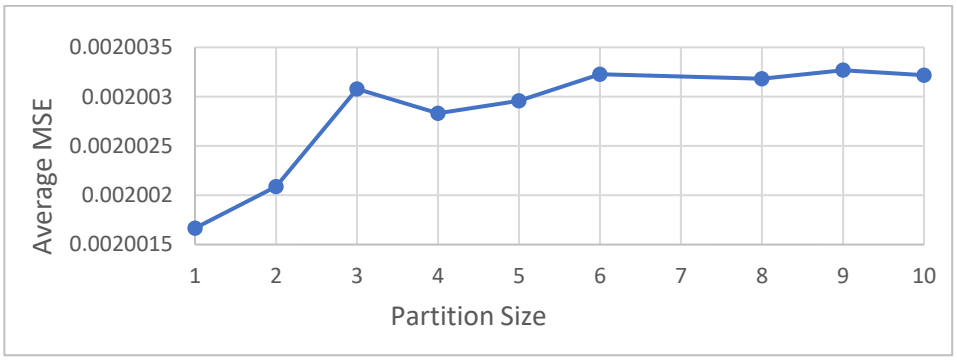

Fig. 4. Variation of average prediction errors of linear-regression-based models in relation to even-sized *k*-partition for the vertical fluid velocity.

Figure 4 shows that the average MSE values fluctuate slightly when the partition size increases. The trend line of this figure suggests that partitioning the domain height with a linear regression estimator may not improve the accuracy of the models for vertical fluid velocity.

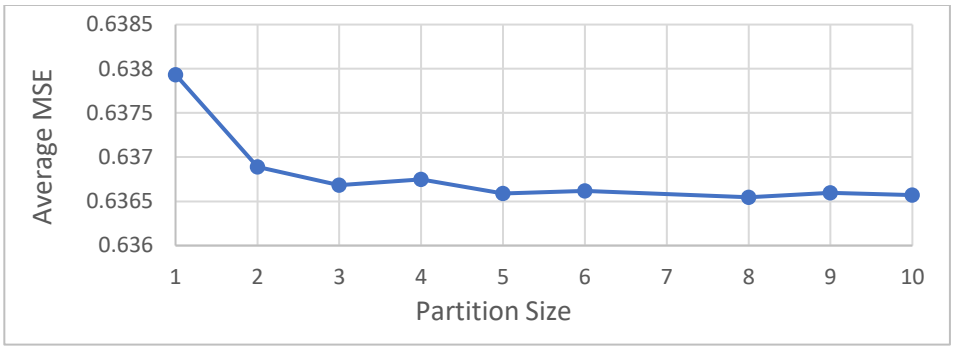

Fig. 5. Variation of average prediction errors of linear-regression-based models in relation to even-sized *k*-partition for the horizontal sediment velocity.

Figure 5 shows that the average MSE values decrease steadily as partition size goes from 1 to 3 and then slightly fluctuate afterward.

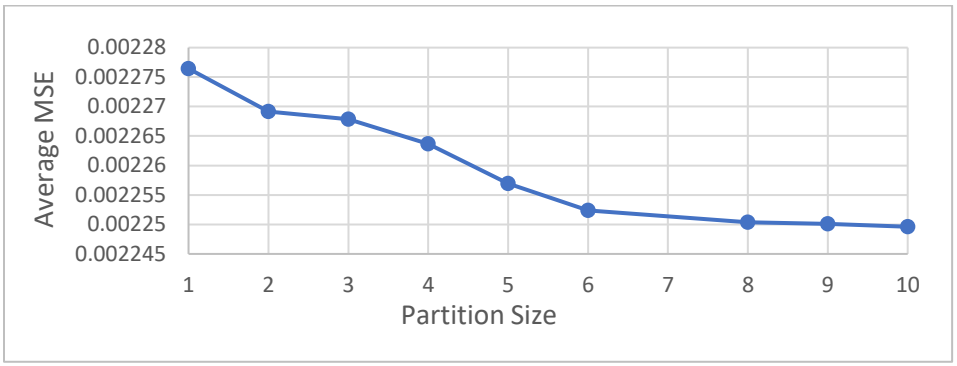

Fig. 6. Variation of average prediction errors of linear-regression-based models in relation to even-sized *k*-partition for the vertical sediment velocity.

Figure 6 shows that the average MSE values decrease steadily until the partition size is 6 partitions and then remain relatively constant afterward.

3.1.2 Gradient Boosting

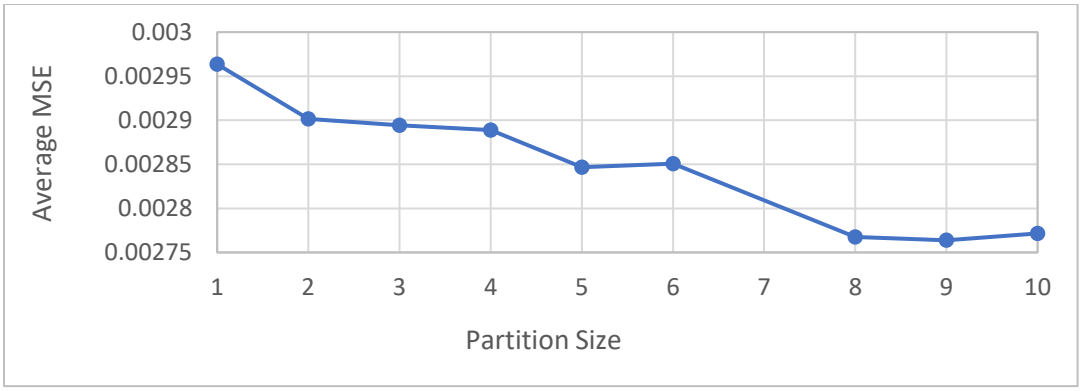

Fig. 7. Variation of average prediction errors of gradient-boosting-based models in relation to even-sized *k*-partition for the concentration profile.

From Figure 7, the average MSE values decrease as the partition size goes up to 8, and then they remain steady.

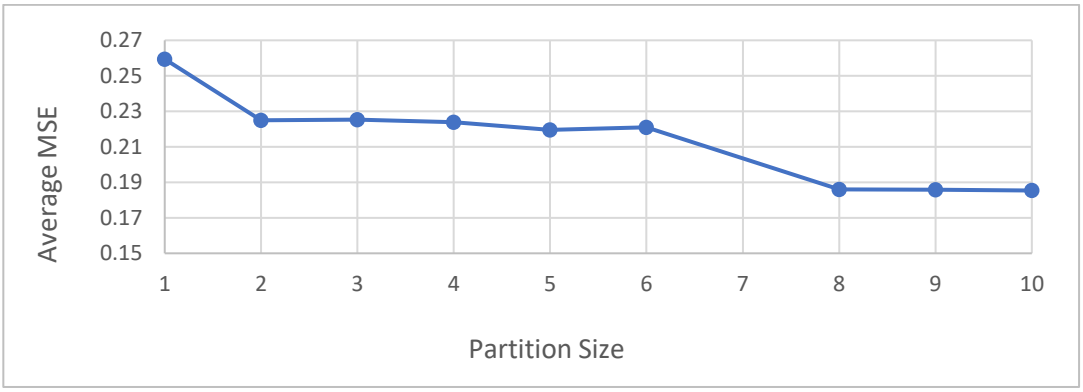

Fig. 8. Variation of average prediction errors of gradient-boosting-based models in relation to even-sized *k*-partition for the horizontal fluid velocity.

From Figure 8, the average MSE values decrease steadily from partition 1 to 2 but remain relatively constant afterward and decrease as the partition size goes from 6 to 8 and again, remain relatively constant.

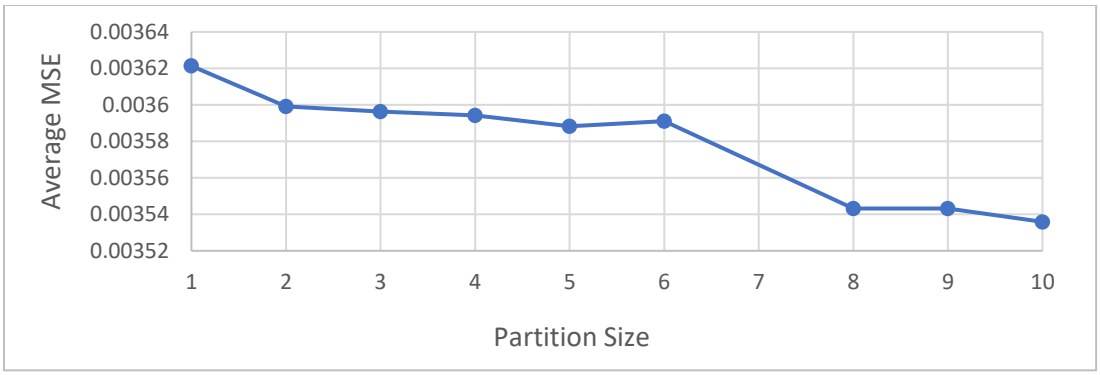

Fig. 9. Variation of average prediction errors of gradient-boosting-based models in relation to even-sized *k*-partition for vertical fluid velocity.

From Figure 9, the average MSE values decrease steadily as the partition size goes from 1 to 2 and then remain relatively steady but again decrease afterward as the partition size goes from 6 to 8.

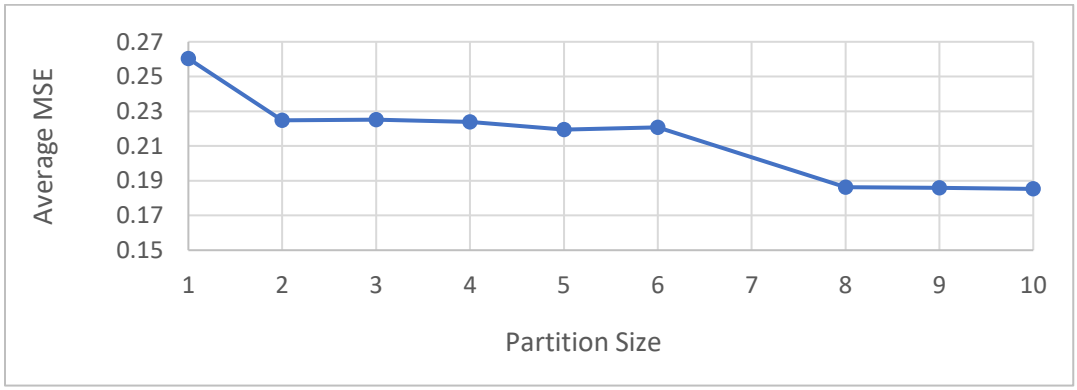

Fig. 10. Variation of average prediction errors of gradient-boosting-based models in relation to even-sized *k*-partition for the horizontal sediment velocity.

From Figure 10, the average MSE values decrease rapidly as the partition size goes from 1 to 2 but then remain relatively constant afterward.

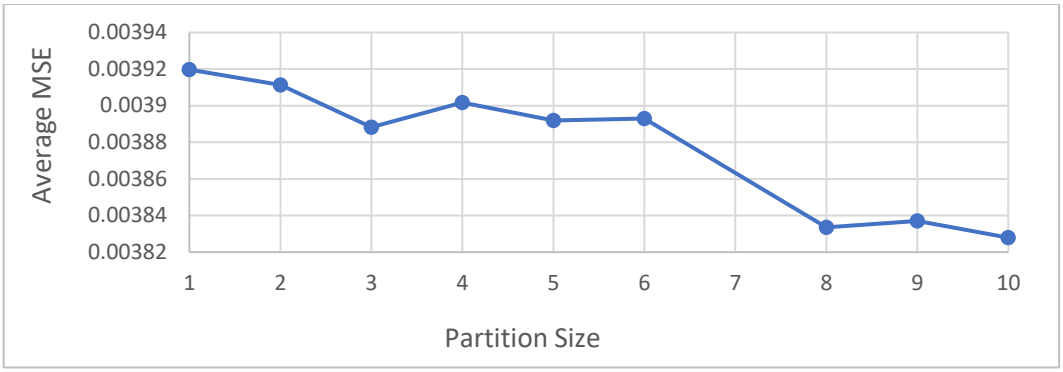

Fig. 11. Variation of average prediction errors of gradient-boosting-based models in relation to even-sized *k*-partition for the vertical sediment velocity.

From Figure 11, the average MSE values slightly decrease as the partition size goes from 1 to 10.

For linear regression and gradient boosting estimators, the average MSE values decrease as the partition size increases. Comparing to linear regression, the gradient boosting method shows a lower average MSE value. Therefore, nonlinear models appear to work better for the sediment transport under oscillatory conditions. The reason can be that wave movement is not linear but more complex than that.

3.2 Tripartition of domain heights

For tripartition, the domain height is divided into three various-sized sections: top, middle, and bottom sections. The middle component wraps around the bed surface ($h_b = 0.1\ m$). The boundaries of the middle components are shown in Table 2. The bottom component's boundary starts at $h = 0\ m$ and reaches the point before the start of the middle component's boundary. The top component's boundary starts at the point after the end of the middle component's boundary all the way to the highest point in the domain height ($h = 0.5m$). There are three cases of expansion to consider: downward expansion, bidirectional expansion, and upward expansion.

The downward expansion keeps the same ending point while decreasing the starting point. On the other hand, the upward expansion keeps the same starting point while increasing the ending point. The bidirectional expansion increases the ending point while decreasing the starting point.

The graphs in Figures 12-21 show that the downward and bidirectional expansion around the bed surface can reduce the MSE values (in most cases). The upward expansion slightly alters the MSE values and even decreases the MSE values in some cases. This suggests the upward expansion towards the water region above the bed surface shows no improvement for sediment transport models. Overall, Tripartition around the bed surface produces a much lower MSE value compared to the result of even-sized partition of the height domain.

Table 2. List of expansion cases based on the middle height ranges of the dataset.

Category	Downward Expansion		Bidirectional Expansion		Upward Expansion	
	Start (m)	End (m)	Start (m)	End (m)	Start (m)	End (m)
1	0.045	0.15	0.135	0.055	0.05	0.155
2	0.04	0.15	0.06	0.14	0.05	0.16
3	0.035	0.15	0.055	0.145	0.05	0.165
4	0.03	0.15	0.05	0.15	0.05	0.17
5	0.025	0.15	0.045	0.155	0.05	0.175
6	0.02	0.15	0.04	0.16	0.05	0.18

3.2.1 Linear Regression

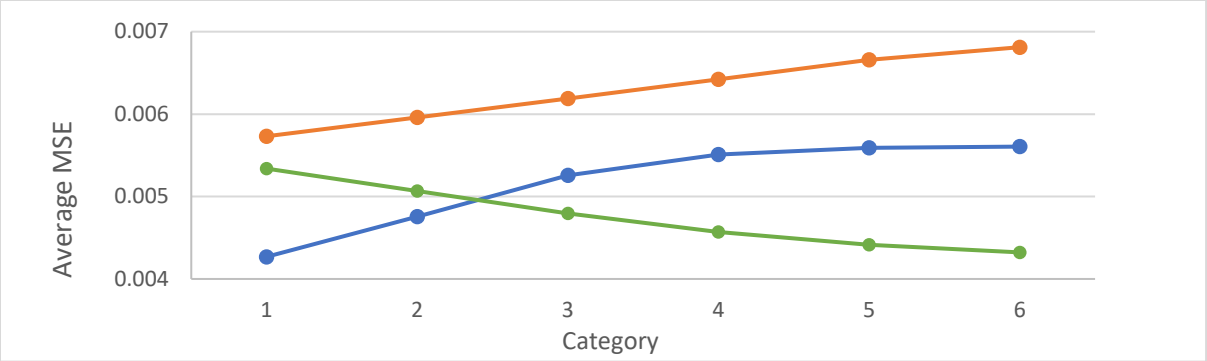

Fig. 12. Variation of average prediction errors of linear-regression-based models in relation to different tripartition categories for the concentration profile. The orange line represents the upward expansion case. The blue line represents the bidirectional expansion case. The green line represents the downward expansion case.

Figure 12 shows that for upward expansion, the average MSE values constantly increase. For bidirectional expansion, they quickly go up from category 1 to 3 and remain constant for the last three categories. For downward expansion, they constantly decrease.

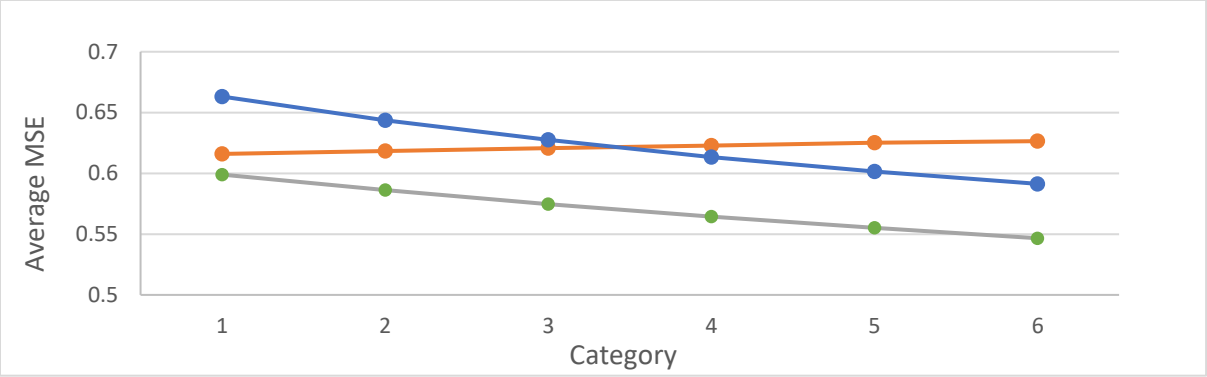

Fig. 13. Variation of average prediction errors of linear-regression-based models in relation to different tripartition categories for the horizontal fluid velocity. The orange line represents the upward expansion case. The blue line represents the bidirectional expansion case. The green line represents the downward expansion case.

From Figure 13, the average MSE values constantly decrease for bidirectional and downward expansion, but they slowly go up for upward expansion.

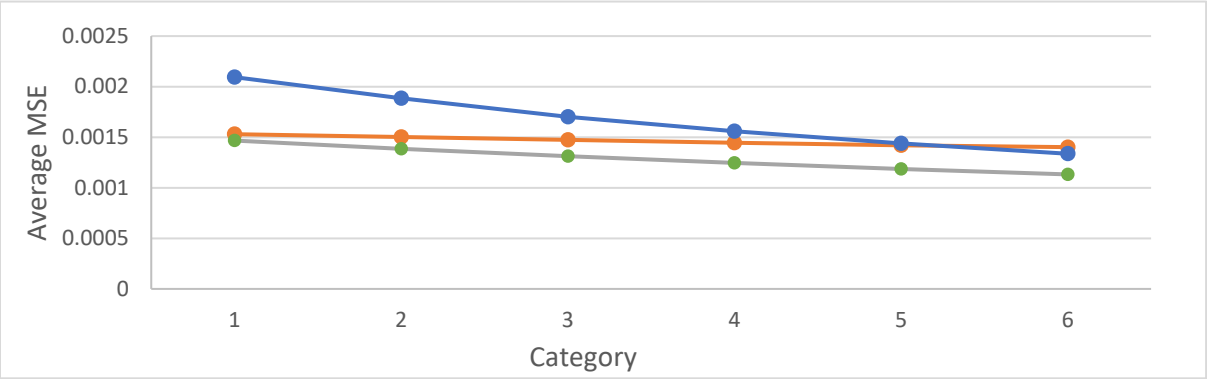

Fig. 14. Variation of average prediction errors of linear-regression-based models in relation to different tripartition categories for the vertical fluid velocity. The orange line represents the upward expansion case. The blue line represents the bidirectional expansion case. The green line represents the downward expansion case.

From Figure 14, the average MSE values slightly decrease for upward, bidirectional, and downward expansion.

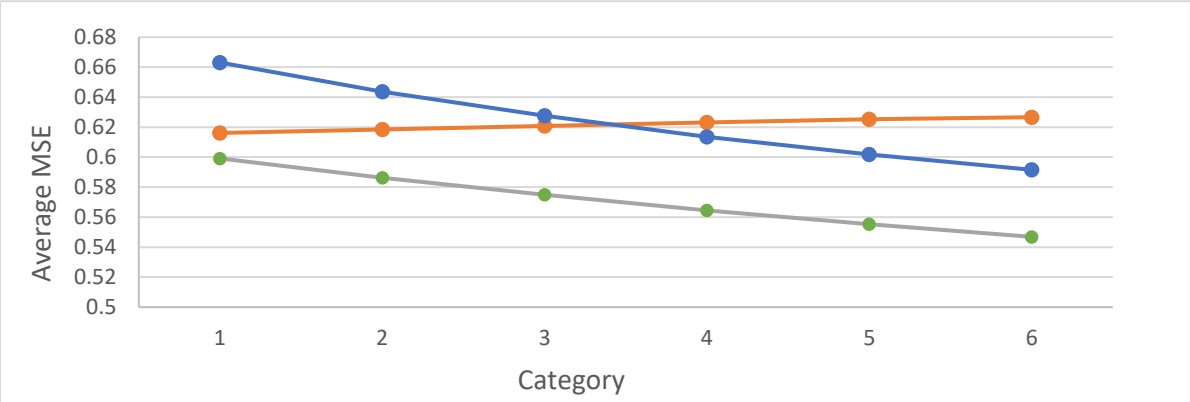

Fig. 15. Variation of average prediction errors of linear-regression-based models in relation to different tripartition categories for the horizontal sediment velocity. The orange line represents

the upward expansion case. The blue line represents the bidirectional expansion case. The green line represents the downward expansion case.

From Figure 15, the average MSE values constantly decrease for downward and bidirectional expansion, but they slowly increase for upward expansion.

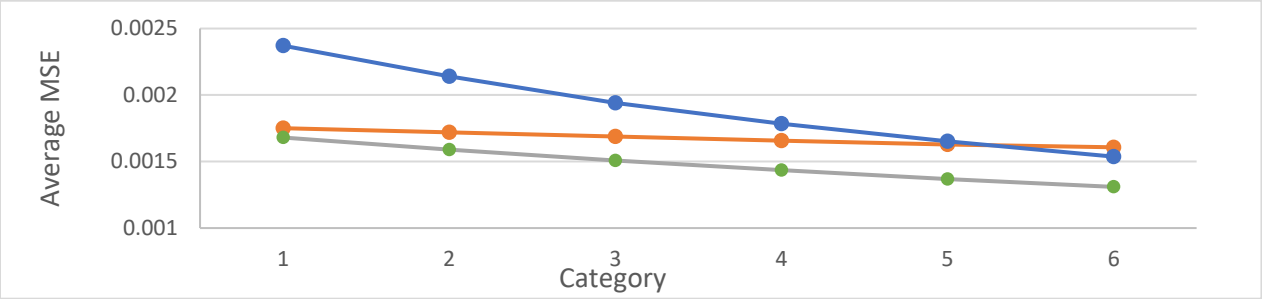

Fig. 16. Variation of average prediction errors of linear-regression-based models in relation to different tripartition categories for the vertical sediment velocity. The orange line represents the upward expansion case. The blue line represents the bidirectional expansion case. The green line represents the downward expansion case.

From Figure 16, the average MSE values constantly decrease for bidirectional and downward expansion, but they show small changes for upward expansion.

3.2.2 Gradient Boosting

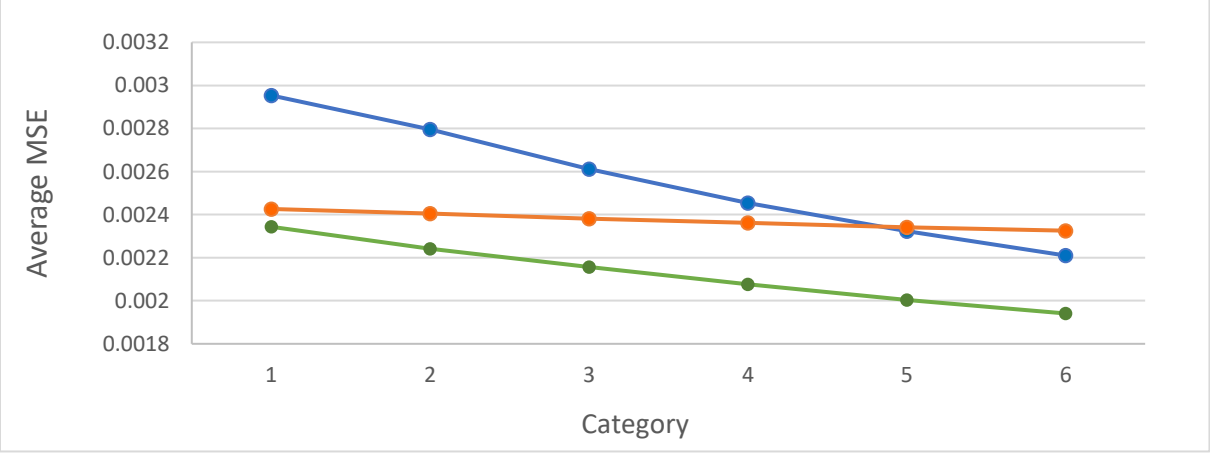

Fig. 17. Variation of average prediction errors of gradient-boosting-based models in relation to different tripartition categories for the concentration profile. The orange line represents the upward expansion case. The blue line represents the bidirectional expansion case. The green line represents the downward expansion case.

From Figure 17, the average MSE values constantly decrease for downward and bidirectional expansion and slowly decrease for upward expansion.

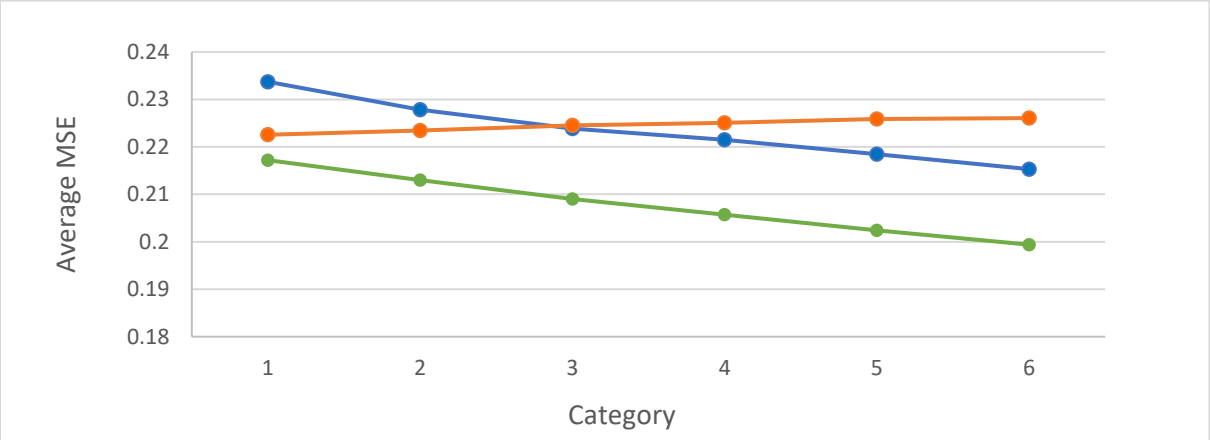

Fig. 18. Variation of average prediction errors of gradient-boosting-based models in relation to different tripartition categories for the horizontal fluid velocity. The orange line represents the upward expansion case. The blue line represents the bidirectional expansion case. The green line represents the downward expansion case.

From Figure 18, the average MSE values constantly decrease for downward and bidirectional expansion, but they slowly increase for upward expansion.

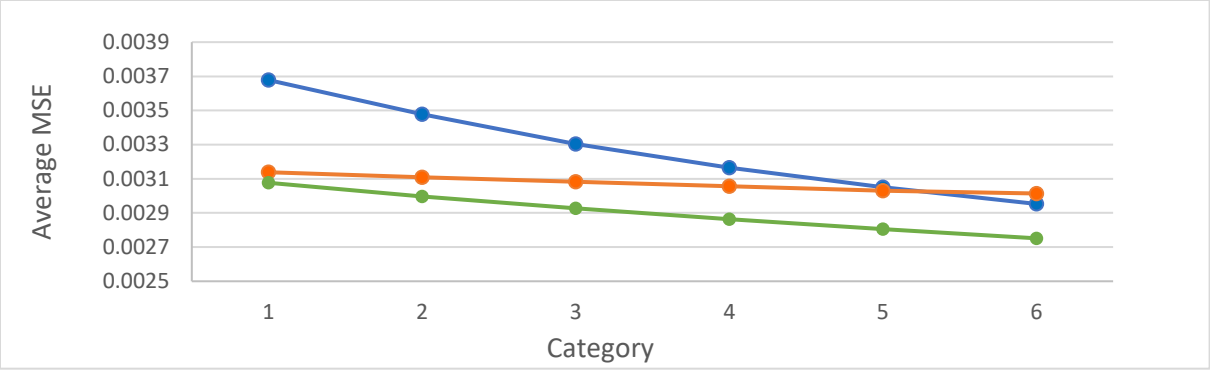

Fig. 19. Variation of average prediction errors of gradient-boosting-based models in relation to different tripartition categories for the vertical fluid velocity. The orange line represents the upward expansion case. The blue line represents the bidirectional expansion case. The green line represents the downward expansion case.

From Figure 19, the average MSE values constantly decrease for downward and bidirectional expansion and slowly decrease for upward expansion.

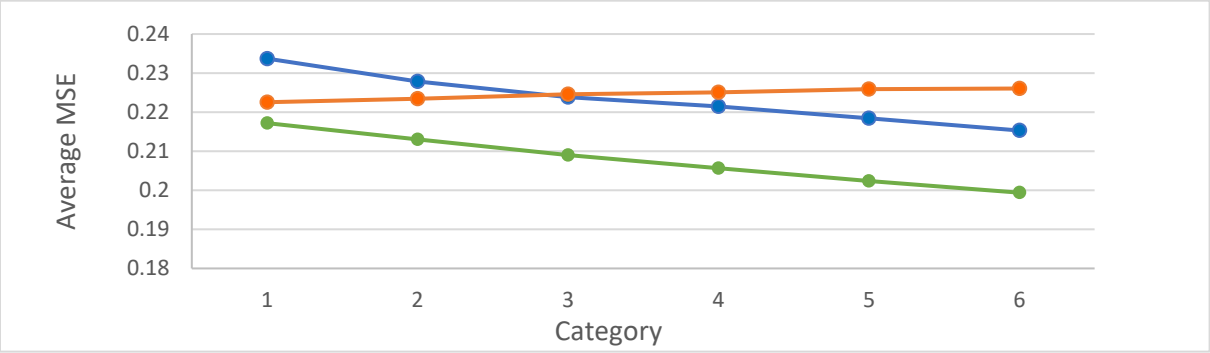

Fig. 20. Variation of average prediction errors of gradient-boosting-based models in relation to different tripartition categories for the horizontal sediment velocity. The orange line represents the upward expansion case. The blue line represents the bidirectional expansion case. The green line represents the downward expansion case.

From Figure 20, the average MSE values constantly decrease for downward and bidirectional expansion, but they slowly increase for upward expansion.

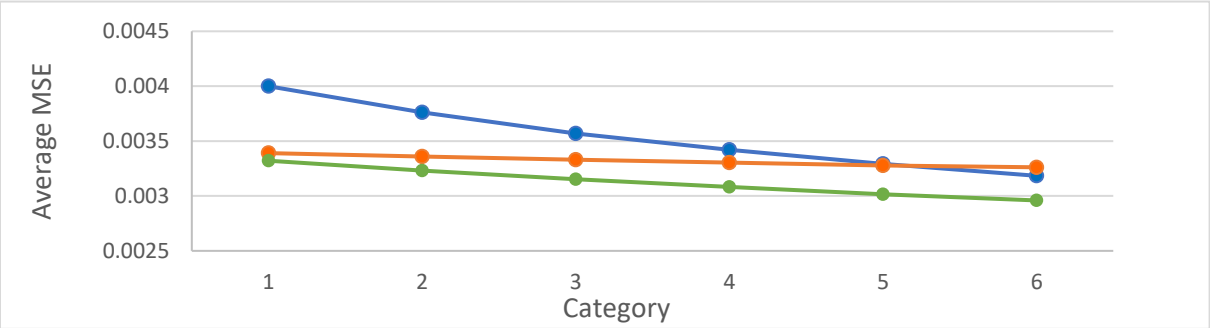

Fig. 21. Variation of average prediction errors of gradient-boosting-based models in relation to different tripartition categories for the vertical sediment velocity. The orange line represents the upward expansion case. The blue line represents the bidirectional expansion case. The green line represents the downward expansion case.

From Figure 21, it can be seen that the average MSE values constantly decrease for downward and bidirectional expansion, but they show no change for upward expansion.

4. Conclusions

In this paper, we have analyzed various statistical machine learning methods that can be used in conjunction to solve bigger problems. Rather than using a single model to predict all the aspects of a simulation, we use various models built on various partitions of the dataset based on the vertical height which improves the performance of our model. A cross-validation approach, namely the *k*-fold cross-validation method, is utilized to assess several sediment transport parameters: the concentration, horizontal fluid velocity, vertical fluid velocity, horizontal sediment velocity, and vertical sediment velocity. The computed models produce promisingly

low MSE values. The data partition is performed in two ways: even-sized partitions and various-sized tripartitions of the domain height. Two estimators (linear regression and gradient boosting) are employed to construct models. We find that gradient boosting, a nonlinear model, greatly outperforms the linear regression model. The results of our experiments show our model to be consistent with the physical laws on how sediment transport occurs. The results show that using separate models for the points with different rates of sediment transport improves the model.

For tripartition, the MSE values decrease as the number of partitions increases and bidirectional and downward expansion around the sediment surface can improve the accuracy while upward expansion does not. It can be explained as various points have similar values for sediment transport at a certain domain height. Thus, making the partitions improved the prediction performance

Similar methods can be adapted in physics, either to speed-up the computation time and make large-scale computation possible or to use various statistical machine learning algorithms for mathematical scaffolding in finding new laws of physics.

5. References

[1] Costa P.J.M. (2016) Sediment Transport. In: Kennish M.J. (eds) Encyclopedia of Estuaries. Encyclopedia of Earth Sciences Series. Springer, Dordrecht.

[2] Moriarty, Julia, et al. "A Hydrodynamic and Sediment Transport Model for the Waipaoa Shelf, New Zealand: Sensitivity of Fluxes to Spatially-Varying Erodibility and Model Nesting." *Journal of Marine Science and Engineering*, vol. 2, no. 2, 2014, pp. 336–369., doi:10.3390/jmse2020336. H. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer-Verlag, 1985, ch. 4.

[3] J.S. Ribberink, Bed-load transport for steady flows and unsteady oscillatory flows, Coast. Eng. 34 (1-2) (1998) 59–82, Jul.

[4] D. Roelvink, A. Reniers, A. van Dongeren, Van Thiel de Vries, J., R. McCall, J. Lescinski, Modelling storm impacts on beaches, dunes and barrier islands, Coast. Eng. 56 (11-12) (2009) 1133–1152, Nov.

[5] N.G. Jacobsen, D.R. Fuhrman, J. Fredsoe, A wave generation toolbox for the open-source CFD library: Openfoam, Int. J. Numer. Methods Fluids 70 (9) (2012) 1073–1088, Nov.

[6] J.C. Warner, C.R. Sherwood, R.P. Signell, C.K. Harris, H.G. Arango, Development of a three-dimensional, regional, coupled wave, current, and sedimenttransport model, Comput. Geosci. 34 (10) (2008) 1284–1306.

[7] G. Lesser, J. Roelvink, J. van Kester, G. Stelling, Development and validation of a three-dimensional morphological model, Coast. Eng. 51 (8-9) (2004) 883–915, Oct.

[8] G. Hoffmans, K. Pilarczyk, Local scour downstream of hydraulic structures, J. Hydraul. Eng. 121 (4) (1995) 326–340, Apr.

[9] X. Liu, M. Garcia, Three-dimensional numerical model with free water surface and mesh deformation for local sediment scour, J. Waterw. Port Coast. Ocean Eng. 134 (4) (2008) 203–217, Jul.

[10] V. Marieu, P. Bonneton, D.L. Foster, F. Ardhuin, Modeling of vortex ripple morphodynamics, J. Geophys. Res. 113 (C9) (2008) C09007, Sep.

[11] Y.J. Chou, O.B. Fringer, A model for the simulation of coupled flow-bed form evolution in turbulent flows, J. Geophys. Res. 115 (C10) (2010) C10041, Oct.

[12] Turowski, Jens & Rickenmann, D. & Dadson, Simon. (2009). The partitioning of the total sediment load of a river into suspended load and bedload. 11. 3973.

[13] P. Dong, K. Zhang, Two-phase flow modelling of sediment motions in oscillatory sheet flow, Coast. Eng. 36 (2) (1999) 87–109.

[14] T.G. Drake, J. Calantoni, Discrete particle model for sheet flow sediment transport in the nearshore, J. Geophys. Res. Oceans 106 (C9) (2001) 19859–19868.

[15] T.J. Hsu, J.T. Jenkins, P.L.F. Liu, On two-phase sediment transport: sheet flow of massive particles, Proc. R. Soc. London Ser. A Math. Phys. Eng. Sci. 460 (2048) (2004) 2223–2250.

[16] M. Li, S. Pan, B.A. O'Connor, A two-phase numerical model for sediment transport prediction under oscillatory sheet flows, Coast. Eng. 55 (12) (2008) 1159–1173, Dec.

[17] L. Amoudry, P.F. Liu, Two-dimensional, two-phase granular sediment transport model with applications to scouring downstream of an apron, Coast. Eng. 56 (7) (2009) 693–702, Jul.

[18] R. Bakhtyar, D.A. Barry, A. Yeganeh-Bakhtiary, L. Li, J.Y. Parlange, G. Sander, Numerical simulation of two-phase flow for sediment transport in the innersurf and swash zones, Adv. Water Resour. 33 (3) (2010) 277–290.

[19] T.G. Drake, J. Calantoni, Discrete particle model for sheet flow sediment transport in the nearshore, J. Geophys. Res. Oceans 106 (C9) (2001) 19859–19868.

[20] J. Calantoni, J.A. Puleo, Role of pressure gradients in sheet flow of coarse sediments under sawtooth waves, J. Geophys. Res. 111 (C1) (2006) C01010, Jan.

[21] J. Calantoni, C.S. Thaxton, Simple power law for transport ratio with bimodal distributions of coarse sediments under waves, J. Geophys. Res. 113 (C3) (2008), Mar.

[22] J. Heald, I. McEwan, S. Tait, Sediment transport over a flat bed in a unidirectional flow: simulations and validation, Phil. Trans. R. Soc. London A 362 (2004) 1973–1986.

[23] Ballio, Francesco, et al. "Lagrangian and Eulerian Description of Bed Load Transport." Journal of Geophysical Research: Earth Surface, vol. 123, no. 2, 13 Feb. 2018, pp. 384–408., doi:10.1002/2016jf004087.

[24] Leschziner, Michael A. "Reynolds-Averaged Navier-Stokes Methods." Encyclopedia of Aerospace Engineering, 2010, doi:10.1002/9780470686652.eae054

[25] L. Amoudry, P.F. Liu, Two-dimensional, two-phase granular sediment transport model with applications to scouring downstream of an apron, Coast. Eng. 56 (7) (2009) 693–702, Jul.

[26] R. Bakhtyar, D.A. Barry, A. Yeganeh-Bakhtiary, L. Li, J.Y. Parlange, G. Sander, Numerical simulation of two-phase flow for sediment transport in the innersurf and swash zones, Adv. Water Resour. 33 (3) (2010) 277–290.

[27] Wachem, Berend & Yu, Xiao & Hsu, Tian-Jian. (2010). A 3D Eulerian-Lagrangian Numerical Model for Sediment Transport.

[28] Cheng, Zhen, et al. “SedFoam: A Multi-Dimensional Eulerian Two-Phase Model for Sediment Transport and Its Application to Momentary Bed Failure.” Coastal Engineering, vol. 119, 2017, pp. 32–50., doi:10.1016/j.coastaleng.2016.08.007.

[29] Mitchell, T. M. (1997). Machine learning, McGraw-Hill, New York.

[30] Hoff, K.J., Tech, M., Lingner, T. et al. Gene prediction in metagenomic fragments: A large scale machine learning approach. BMC Bioinformatics 9, 217 (2008) doi:10.1186/1471-2105-9-217

[31] LAPPIN, S., & SHIEBER, S. (2007). Machine learning theory and practice as a source of insightinto universal grammar. Journal of Linguistics, 43(2), 393-427. doi:10.1017/S0022226707004628

[32] Meiyin Wu and Li Chen, "Image recognition based on deep learning," 2015 Chinese Automation Congress (CAC), Wuhan, 2015, pp. 542-546.

[33] Bhattacharya, B & Price, R. & Solomatine, Dimitri. (2007). A Machine Learning Approach to Modeling Sediment Transport. Journal of Hydraulic Engineering-asce - J HYDRAUL ENG-ASCE. 133. 10.1061/(ASCE)0733-9429(2007)133:4(440).

[34] Bagnold, R. A. (1988b). “An empirical correlation of bedload transportrates in flumes and natural rivers.” The physics of sediment transportby wind and water, ASCE, Reston, Va., 323–345.

[35] Einstein, H. A. (1950). “The bed-load function for sediment transporta-tion in open channel flows.” Technical Bulletin No. 1026, Soil Con-servation Series, U.S. Dept. of Agriculture, Washington, D.C.

[36] van Rijn, L. C. (1984a). “Sediment transport. Part I: Bed-load transport.”J. Hydraul. Eng.,110(10), 1431–1456.

[37] van Rijn, L. C. (1984b). “Sediment transport. Part II: Suspended-loadtransport.” J. Hydraul. Eng.,110(11), 1613–1641.

[38] van Rijn, L. C. (1993).Principles of sediment transport in rivers, estuaries, and coastal areas, Aqua, Amsterdam, The Netherlands.

[39] Kitsikoudis, V., Sidiropoulos, E. & Hrissanthou, V. Water Resour Manage (2014) 28: 3727. https://doi.org/10.1007/s11269-014-0706-z.

[40] T. O’Donoghue, S. Wright, Concentrations in oscillatory sheet flow for well sorted and graded sands, Coast. Eng. 50 (3) (2004) 117–138.

[41] Xin Yan and Xiao Gang Su. 2009. Linear Regression Analysis: Theory and Computing. World Scientific Publishing Co., Inc., River Edge, NJ, USA.

[42] Sinnakaudan, S. K., et al. “Multiple Linear Regression Model for Total Bed Material Load Prediction.” *Journal of Hydraulic Engineering*, vol. 132, no. 5, 2006, pp. 521–528., doi:10.1061/(asce)0733-9429(2006)132:5(521).

[43] Friedman, J.H. Greedy function approximation: A gradient boosting machine. Ann. Stat. 2001, 29, 1189–1232.

[44] Zhou Zhi-Hua (2012). Ensemble Methods: Foundations and Algorithms. Chapman and Hall/CRC. p. 23. ISBN 978-1439830031.

[45] Chen, Tianqi, and Carlos Guestrin. “XGBoost.” *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16*, 2016, doi:10.1145/2939672.2939785.