An optimal statistical regression model for predicting wave-induced equilibrium scour depth in sandy and silty seabeds beneath pipelines

Zhang, Yaqi, Wu, Jinran, Zhang, Shaotong, Li, Guangxue, Jeng, Dong-Sheng, Xu, Jishang, Tian, Zhuangcai and Xu, Xingyu

Zhang, Y., Wu, J., Zhang, S., Li, G., Jeng, D.-S., Xu, J., Tian, Z. and Xu, X. (2022). An optimal statistical regression model for predicting wave-induced equilibrium scour depth in sandy and silty seabeds beneath pipelines. *Ocean Engineering, 258*, Article 111709. https://doi.org/10.1016/j.oceaneng.2022.111709

An optimal statistical regression model for predicting wave-induced equilibrium scour depth in sandy and silty seabeds beneath pipelines

Yaqi Zhang [1,3], Jinran Wu [2], Shaotong Zhang [1,3,*], Guangxue Li [1], Dong-Sheng Jeng [3], Jishang Xu[1], Zhuangcai Tian [4], Xingyu Xu [5]

[1] Key Laboratory of Submarine Geosciences and Prospecting Techniques (Ministry of Education), College of Marine Geosciences, Ocean University of China, Qingdao, 266100, China

[2] School of Mathematical Sciences, Queensland University of Technology, Brisbane, 4001, Australia

[3] School of Engineering and Built Environment, Griffith University Gold Coast Campus, Gold Coast, 4222, Australia

[4] State Key Laboratory for Geomechanics and Deep Underground Engineering, China University of Mining and Technology, Xuzhou, 221116, China

[5] Shengli Oilfield Technology Inspection Centre, SINOPEC, Dongying, 257000, China

* Corresponding author: shaotong.zhang@ouc.edu.cn

Abstract

Equilibrium scour depth (*S*) of seabed is critical to the safety of offshore pipelines which is one of the most important topics in ocean engineering. Compared to sands, few experiments have been done for silty seabed. In the present work, scour experiments under wave-only action were performed for both sandy and silty seabeds. Together with the data from literature, the most abundant dataset at the present stage is established. Based on this, two practical formulas for *S* were obtained with adaptive robust regression (ARR) from a data-driven perspective. One is for sands only that is related to the Keulegan–Carpenter (*KC*) number, pipeline-seabed gap and grain size of sands. The other is a more generalized model for both sands and silts, which is related to the *KC* number and sediment type that is distinguished by introducing a dummy variable. The formulas outperform the commonly-used process-based and data-driven models while also showing good interpretations in physical meaning. For silts from the Yellow River Delta, the *S* in silts is generally 1.2 times of that in sands. The better performance is attributed to (1) the outliers in the dataset are effectively handled with ARR; (2) the most abundant dataset.

Keywords

cohesive sediments; wave-flume experiments; outliers; adaptive robust regression; dummy variable; Yellow River Delta

1. Introduction

Submarine pipelines are important marine installations for the transportation of oils and gases in the ocean, of which the stability is of great significance to ocean engineering. Seabed scour beneath pipelines is the most common threat to pipeline security (Sumer and Fredsøe, 2002; Zhao et al., 2021). Without a precise prediction of the scour depth, the protection design of submarine pipelines will be inappropriate and possibly cause damage to the pipelines (Dong et al., 2017; Fan et al., 2022).

Significant efforts have been devoted to establishing process-based (i.e., physical) models for predicting the equilibrium scour depth (*S*) beneath pipelines by clarifying the physical processes that determine sediment motions. The most commonly-employed empirical formula was fitted by sand data with d_{50} (median particle size of sediment) values of 0.58, 0.36, and 0.18 mm and for a pipe in contact with the bed (*e*/*D*=0; a negative value refers to burial in the present paper) by Sumer and Fredsøe (1990) (SF1990 hereinafter), in which the ratio of the *S* to the pipeline diameter (*D*) is related to the *KC* number for live-bed situations ($\theta > \theta_{\mathrm{cr}}$):

$$S/D = 0.1\sqrt{KC}, \qquad \theta > \theta_{\mathrm{cr}}, \tag{1}$$

where *KC* is the Keulegan–Carpenter number, which is computed from Eq. (13) hereinafter, θ is the Shields parameter and θ_{cr} is the critical Shields parameter for sediment entrainment, which is computed from Eq. (31) hereinafter.

Cevik & Yuksel (1999) modified Eq. (1) to estimate the wave-induced scour around pipelines on a horizontal bed based on their experiments and those of SF90 and Lucassen (1984) as:

$$\frac{S}{D} = 0.11KC^{0.45}, \tag{2}$$

Pu et al. (2001) studied the effect of various soil materials (d_{50}=0.68, 0.47, 0.20, and 0.0047 mm) on *S*/*D* for both live-bed and clear-water ($\theta < \theta_{\mathrm{cr}}$) conditions and proposed the following relationship:

$$\frac{S}{D} = B * KC^{\mathrm{m}}, \tag{3}$$

where m is a constant related to bed materials (e.g., m=3.18 for a sandy bed) and *B* is a function of *e*/*D* (the initial pipe position with respect to the bed), which appears to be an important parameter in determining *S* (Sumer and Fredsøe, 1990). The shape of this equation is similar to that of Eq. (1), but the exponent m=3.18 for sandy beds is quite different from m=0.5 in Eq. (1). In addition, Pu et al. (2001) pointed out that for sandy beds, *B* increases with *e*/*D*. This is contradictory to the case in Eq. (4) proposed by Sumer and Fredsøe (2002) for various *e*/*D* values in live-bed conditions, according to their experiments and the data of Lucassen (1984):

$$\frac{S}{D} = 0.1KC^{0.5}\exp\left(-0.6\frac{e}{D}\right),\ -0.25\ D \leq e \leq 1.2\ D, \tag{4}$$

this model suits the scenario in which the pipe is installed at a depth of 0.25 *D* at the erodible bed up to 1.2 *D* higher than the bed surface (Mousavi et al., 2009). From the perspective of Pu et al. (2001), here, $B = 0.1 * \exp\left(-0.6\frac{e}{D}\right)$, i.e., *B* decreases with *e*/*D*; thus, *S*/*D* decreases with *e*/*D*, contradicting the opinion of Pu et al. (2001)(i.e., for sandy beds, *B* increases with *e*/*D*).

Mousavi et al. (2009) found that when the primary installation depth (initial gap) of the pipe, *e*, exceeds a specified depth, no scouring occurs underneath the pipe in cases of small *KC* numbers (i.e., the effect of wave-seabed interactions is rather low), which is the case in the offshore area where the waves are in the transition zone or deep water:

$$\frac{S+|e|}{D} = 0.1KC^{0.5},\ \text{for}\ KC < 6, \tag{5}$$

As only buried ($e/D < 0$) or no gap ($e/D = 0$) data were considered in Mousavi et al. (2009), when the pipe was buried in the seabed, $|e| > 0$; thus, *S* was reduced for the same *KC* number according to Eq. (5). This result indicates that burial depth (*e*) reduces scour, thereby supporting the conclusion of Sumer and Fredsøe (2002).

The abovementioned works are all from a process-based perspective; however, with more sufficient data on pipeline scour from the worldwide community, data-driven models have emerged as an alternative to process-based models. Machine learning approaches, such as artificial neural networks (ANNs), have been used to increase the accuracy of scour depth prediction (Kazeminezhad et al., 2010). Kızılöz et al. (2015) developed models using the feed forward back propagation (FFBP) ANN technique for both regular and irregular wave conditions. However, two shortcomings of the ANN methods are that they are generally not as transparent as physical models, and more importantly, it is difficult to give clear mathematical formulas that are practical for scour predictions. One step forward was achieved by the statistical learning works of Etemad-Shahidi et al. (2011), who proposed an M5’ model tree that can provide understandable formulas,

$$\frac{S}{D} = 3.344KC^{0.512}\theta^{1.296}\exp\left(-2.32\frac{e}{D}\right), \quad \text{for } \theta \leq 0.064, \tag{6}$$

$$\frac{S}{D} = \begin{cases} 0.149KC^{0.477}\theta^{0.121}\exp\left(-0.472\frac{e}{D}\right) \text{ for } \theta > 0.064, \text{and} \frac{e}{D} \leq 0.145 \\ 0.048KC^{0.782}\theta^{0.121}\exp\left(-0.942\frac{e}{D}\right) \text{ for } \theta > 0.064, \text{and} \frac{e}{D} > 0.145 \end{cases}, \tag{7}$$

and Sharafafi et al. (2018), who proposed the following formulas for clear-water and live-bed scour regimes:

$$\frac{S}{D} = 4.17KC^{0.72}\theta^{1.55}\exp\left(-3.9\frac{e}{D}\right), \quad \text{for } \theta \leq 0.064, \tag{8}$$

$$\frac{S}{D} = \begin{cases} 0.149KC^{0.42}\theta^{0.08}\exp\left(-0.472\frac{e}{D}\right) \text{ for } \theta > 0.064, \text{and} \frac{e}{D} \leq 0.145 \\ 0.073KC^{0.45}\theta^{0.17}\exp\left(-0.094\frac{e}{D}\right) \text{ for } \theta > 0.064, \text{and} \frac{e}{D} > 0.145 \end{cases}. \tag{9}$$

However, most of the aforementioned studies focused on sandy seabeds. Zhang et al. (2019) found that the wave-induced seepage effect in a silty seabed has some promoting effect on the initial scour process. To date, few quantitative studies have been performed on scour in silts. Therefore, studies dedicated to the field assessment of scouring processes in cohesive seabeds still have to resort to formulas established for noncohesive sediments (Xu et al., 2012). Postacchini and Brocchini (2015) explored the understanding and modelling of scouring processes in cohesive seabeds based on dimensional analysis and derived a formula for *S*/*D*. To the best of our knowledge, no existing work has explored the difference in pipeline scour between cohesive and noncohesive sediments from a data-driven perspective.

To this end, a series of laboratory experiments for sandy and silty seabed scouring around a pipeline under waves were conducted in the present study. Combined with the data collected from published literature, the most complete dataset for scour under pipelines in regular waves was established by the present paper. Then, two statistical models were formed based on the experimental data (22 sets) and data collected from the literature (182 sets). Adaptive robust regression, which can handle the outliers in the original dataset, was introduced for modelling. The first model was trained for sands and tested as the optimal model for sands comparing to the existing popular sand scour models. The second model was trained for both sands and silts together; therefore, a generalized model for scour under pipelines in both sands and silts was developed for the first time. The model was proven to outperform popular process-based or data-driven models. The influence of sediment types on scour depth was detected from a data-driven perspective for the first time.

The following parts of the paper are organized as follows: Section 2 describes our laboratory experiments; Section 3 gives the methodology of the statistical learning; Section 4 presents the dataset for statistical learning, which consists of the present experimental data and data from the literature; Section 5 gives the data modelling processes in which a formula for sand scour and a more generalized model for both sands and silts are derived; Section 6 gives the validation of the proposed models with the test sets; this paper ends with a few conclusions in Section 7. The general flow of the present paper is also illustrated in **Figure A1** in the Appendix.

2. Laboratory physical experiments

In this section, the experimental setup and procedure are given, and the key parameters are calculated.

2.1 Experimental setup

Flume: The experiments were carried out in a large wave flume at the Drilling Technology Research Institute of the Sheng-li Petroleum Oilfield Administration Bureau, Dongying, Shandong, China. The wave flume is 62 metres long, 1.5 metres wide and 1.1 metres high, with an maximum working water depth of 0.70 m (**Figure 1**). A wave-maker is installed at one end of the flume, which is composed of a push plate and a control system. At the other end of the flume, a wave dissipation system is installed. It is composed of an artificial gravel beach with a wire mesh covering it. The gravel is used to absorb the incident wave energy for wave elimination, thereby minimizing the wave reflections to avoid affecting the designed wave parameters. The wire mesh is used to prevent the gravel from excessive displacement in case it influences the wave-eliminating effect.

A soil tank 2.4 m long, 1.5 m wide, and 0.5 m deep is located in the middle of the flume. The position of the soil tank is 45 m away from the wave generator (Xu et al., 2010).

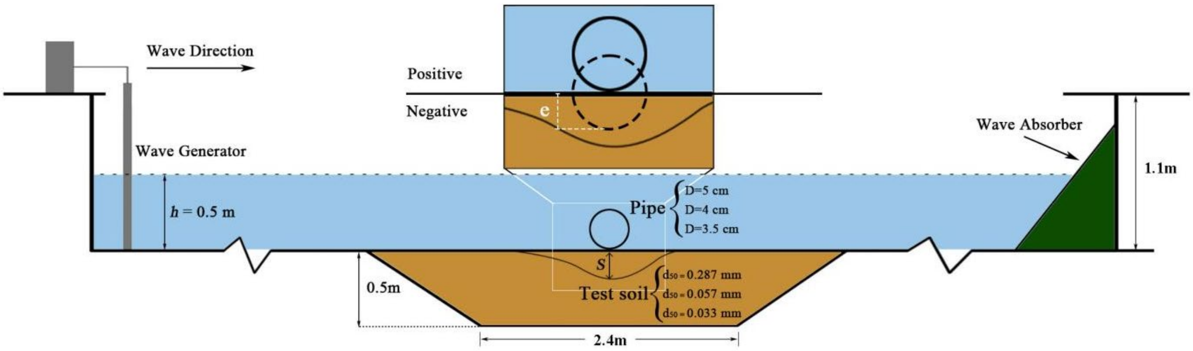

Figure 1. Experimental setup.

Pipelines: Circular pipeline models that are 1.49 m in length and have three different diameters (3.5, 4 and 5 cm) are artificially placed on the soil bed. The pipeline is made of glass fibre reinforced plastic and placed in the centre of the soil tank perpendicular to the wave direction. Both ends of the pipeline are fixed by two rigid rods attached to the flume bottom to prevent both the horizontal and vertical displacement of the pipeline when the bed is scoured, therefore avoiding errors in recording the scour depth. Two initial burial conditions of the pipeline, namely, unburied (e/D=0) and half-buried (e/D=-0.5), are considered in the experiments.

Waves: A push plate is mechanically driven to generate regular waves with wave periods (T_w) of 0.6-3.5 s by setting the motion frequency on the computer (Zhou et al., 2011). The designed wave parameters are summarized in **Table 1**.

Table 1. Designed wave parameters.

Water depth (cm)	H_w (cm)	T_w (s)	e/D	
			Unburied	Half-buried
50	8	1.3	0	- 0.5
50	17	1.9	0	- 0.5

Sediments and Bedforms: The sediments used in the experiment are commercial soil substitutes. The particle size distribution curves of the sediments are measured with a Mastersizer 2000 laser particle size analyser. The median diameters of the three types of sediments (fine sand, very fine sand, and silt) are d_{50} = 0.287 mm, 0.057 mm, and 0.033 mm, respectively. The particle size distribution curves of the sediment are shown in **Figure 2**.

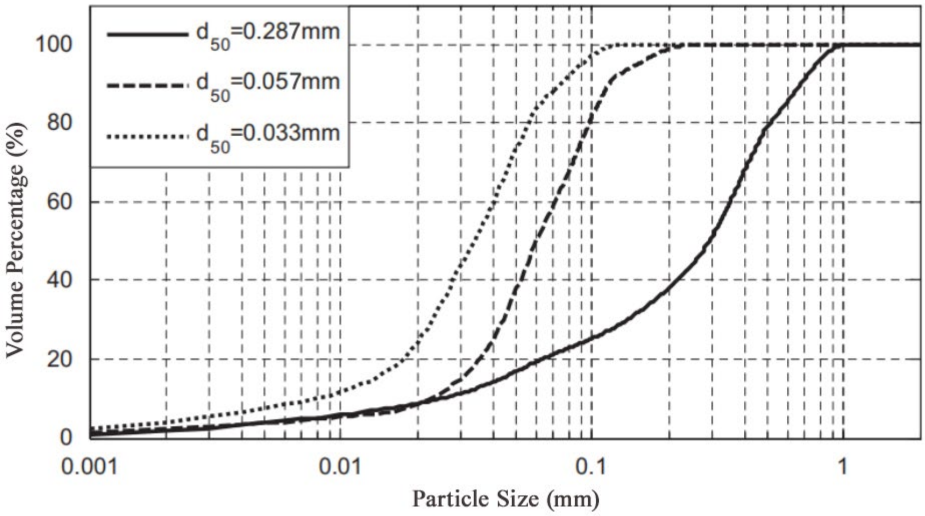

Figure 2. Particle size distribution curves of the three types of sediments (Xu et al., 2010).

Similarity Scale: Considering that the field waves, water depth and pipeline geometries in the Chengdao sea region (Yellow River Delta) where many key pipelines (of the Shengli oilfield) distributed were generally 20 times of the geometries that could be simulated in our wave flume, the geometrical scale was taken as 1:20 (Zhang et al., 2019).

Table 2 Similarity scale of the present experiments

Geometrical scale is 1:20	H_w		h	D			Max.(S/D)
Experimental geometry (m)	0.08	0.17	0.50	0.035	0.04	0.05	0.75
Field geometry (m)	1.60	3.40	10	0.70	0.80	1	0.80

Note: The geometrical scale of 1:20 successfully reproduced experimental results of *S/D* basically on a scale of 1:1 with field measurement (Xu et al., 2012), this, to some extent, proves the rationality of the selected experimental similarity scale. Here h is water depth, H_w is wave height, D is pipeline diameter.

The wave heights (H_w) of 1.6 m and 3.4 m in Table 2 are well within the normal range of H_w in the Chengdao sea; the water depth (h) in this region also fluctuates around 10 m (Zhang et al., 2021a, b). The pipeline diameters (D) are normally 0.2-1 m (Zhang, 2019). The *S/D* results from our experiments are within 0-0.75 which are well consistent with the field survey results (maximum *S/D* was 0.80, mostly < 0.60) of Xu et al. (2012) (see Table 2). This proves that our design of similarity scale is reasonable.

2.2 Experimental procedure

The experimental procedure is outlined as follows:

(1) The test soils were mixed with water to form a slurry, which was then backfilled into the soil tank to form a freshly deposited (remodelled) seabed. The seabed was properly levelled to form an initially flat surface.

(2) The flume was filled slowly with water to a depth of 50 cm.

(3) Small waves were initiated to accelerate the consolidation of the freshly deposited seabed.

(4) Waves were stopped for the static consolidation of the seabed until the pore pressure stabilized.

(5) The pipeline was placed horizontally in the centre of the soil tank, both ends were fixed to two rigid rods emerging from the bed surface to ensure that the pipeline elevation did not change to prevent error when recording the scour depth.

(6) The waves were switched on and off according to the schedule in **Table 3**, and the wave parameters were recorded with a wave gauge. Because the waves deformed after a period of time, the waves were generated intermittently according to the schedule in **Table 3**, i.e., waves were stopped after a period of time and continued after the water surface recovered to flat.

(7) Waves were finally stopped until no significant variation in the scour depth beneath the pipeline was observed.

(8) The equilibrium scour depth under the pipeline was measured from the side of the flume. Although the equilibrium scour depths were in 3D patterns, we have visually confirmed that the measured value from the side wall of the flume is consistent with the equilibrium scouring depth in the three-dimensional structure. The major difference between 2D and 3D is the local structure of the scour holes, not the equilibrium scour depths. Nevertheless, our future works will try to use 3D imaging technique of the scour hole to determine equilibrium scour depth.

(9) The pipeline was uninstalled, the bed surface was restored to level, and the next test was continued.

Table 3. Experimental procedure and corresponding wave parameters.

Exp. No.	H_w (cm)	T_w (s)	Duration* (min)
G-1-1-03	17	1.9	20+(15)+20
G-1-1-06	17	1.9	12+(18)+15+(10)+10
G-1-2-01	8	1.3	15+(81)+15+(73)+15+(55)+17+(5)+17
G-1-2-02	17	1.3	15+(82)+15+(53)+15+(52)+20
G-1-2-03	17	1.9	15+(52)+15+(62)+15+(55)+20
G-1-2-04	8	1.3	55+(41)+55
G-1-2-05	17	1.3	15+(55)+15+(4)+15+(44)+15+(5)+15
G-1-2-06	17	1.9	15+(56)+15+(53)+15+(56)+15+(6)+15
G-1-2-07	17	1.9	15+(12)+20
G-2-1-03	17	1.9	15+(11)+15+(9)+20
G-2-1-06	17	1.9	15+(14)+15+(6)+15
G-2-2-03	17	1.9	10+(43)+10+(101)+10+(?)+10
G-2-2-06	17	1.9	15+(35)+15+(32)+20
G-2-2-07	17	1.9	15+(14)+15+(9)+15
G-3-1-03	17	1.9	10+(3)+15+(5)+15+(5)+15
G-3-1-06	17	1.9	10+(7)+10+(5)+10+(5)+10+(4)+10+(12)+15
G-3-2-01	8	1.3	15+(6)+15+(8)+17
G-3-2-02	17	1.3	15+(6)+20+(8)+20
G-3-2-03	17	1.9	15+(8)+15+(5)+15+(5)+15
G-3-2-04	8	1.3	15+(6)+15+(4)+15+(4)+15+(5)+20
G-3-2-06	17	1.9	15+(5)+15+(6)+15+(5)+15
G-3-2-07	17	1.9	10+(5)+10+(6)+10+(6)+10+(5)+10

* Duration of 20+(15)+20 means that the waves were active for an initial 20 minutes, followed by a 15-minute rest and then another 20-minute wave action.

2.3 Determination of key parameters

Several key parameters, which are potentially important for pipeline scour (e.g., the

Keulegan-Carpenter (*KC*) number, Shields parameter, Ursell parameter etc), were determined as follows:

(1) Wavelength L

For a limited water depth (1/20＜h/L＜1/2), the wavelength (L) was calculated by the linear wave dispersion equation (Dean and Dalrymple, 1991) as:

$$L = \frac{\mathrm{g}}{2\pi} {T_{\mathrm{w}}}^{2} \tanh(\frac{2\pi}{L} h), \tag{10}$$

where g is the acceleration due to gravity and h is the still water depth.

(2) Maximum water particle velocity on the bed in the absence of the pipe, U_{m}

U_{m} was calculated based on the second-order Stokes wave theory:

$$U_{\mathrm{m}} = \frac{\pi H_{\mathrm{w}}}{T_{\mathrm{w}} \sinh(kh)} + \frac{3}{4} \frac{\pi^2 {H_{\mathrm{w}}}^2}{T_{\mathrm{w}} L \sinh^4(kh)}, \tag{11}$$

where $k = 2\pi/L$ is the wave number.

(3) Ursell parameter, U_{r}, is a parameter that is commonly used to evaluate the wave nonlinearity. As wave nonlinearities influence the vortex formation and development around an obstacle, hence, it also influences the scour depth (Corvaro et al., 2018). The Ursell number was calculated by

$$U_{\mathrm{r}}=H_{\mathrm{w}}L^2/h^3 \tag{12}$$

where H_{w} is the incident wave height, h is the water depth, L is wavelength. For long waves ($L \gg h$) of small height H_{w}, i.e. $U_{\mathrm{r}} \ll 32\pi^2/3 \approx 100$, the second-order Stokes theory is applicable. Here, U_{r} was found to range from 3.41-19.75, therefore, the accuracy of using the second order Stokes wave theory was further evaluated. We find that using second order Stokes wave theory improves ca. 11 % from linear theory, which means further improvement will be less than 11 % even using higher-order Stokes theory. Therefore, second order Stokes wave theory was employed in the present paper, to keep consistent with the previous paper of Zhou et al. (2011) (Eq. 2 therein) which was based on the same experimental data.

(4) Keulegan-Carpenter (*KC*) number (Sumer and Fredsøe, 1990)

$$KC = \frac{U_{\mathrm{m}} T_{\mathrm{w}}}{D}, \tag{13}$$

(4) Shields parameter θ (Nielsen et al., 2001)

$$\theta = \frac{U_*^2}{\mathrm{g}(\frac{\rho_s}{\rho} - 1) d_{50}}, \tag{14}$$

where ρ_s is the density of sediment particles and ρ is the density of seawater.

(5) Friction velocity U_*

$$U_* = \sqrt{\tau_\mathrm{w}/\rho}, \tag{15}$$

(6) Wave shear stress τ_w (Nielsen, 2009, p 213)

$$\tau_\mathrm{w} = 0.5\rho f_\mathrm{w} U_\mathrm{m}^2, \tag{16}$$

(7) Wave friction factor f_w (Nielsen, 2009, p 214)

$$f_\mathrm{w} = \exp[5.5(\frac{2.5d_{50}}{A})^{0.2} - 6.3], \tag{17}$$

(8) Semi-excursion of wave particles at the bottom, A (Nielsen, 2009, p 216)

$$A = \frac{\sqrt{2}U_\mathrm{m}}{2\pi/T_\mathrm{w}}, \tag{18}$$

3. Statistical learning method

The statistical learning method and the evaluation criteria of the optimal model are elaborated in this section. Adaptive robust regression (ARR) is employed for data modelling, as it is good at handling outliers. The Akaike information criterion (AIC) is used as the optimal model evaluation criterion, which is detailed in subsection 3.2.

3.1 Adaptive robust regression

In statistical modelling, given a dataset (x_i, y_i), $x_i \in \mathrm{R}^d$, R is the real number field, d is the dimension of x_i, $y_i \in \mathrm{R}$, i=1, 2..., n (n is the amount of the sample), an ordinary linear regression model can be given as

$$y_i = x_i^T \beta + \sigma\mu_i, \tag{19}$$

where the dependent variable is y_i; the independent variable is $x_i = (1, x_{i1}..., x_{id})'$, the upper corner mark T means the transpose of matrix; the scaled noise is μ_i, σ (>0) is the scale coefficient of the noise [In generalized linear model, the noises are assumed to comply with a standardized distribution with a scale. Corresponding to the unit of y_i, the value of σ would change to scale noises to follow a standardized distribution]; and the coefficient of regressor is $\beta = (\beta_0, \beta_1..., \beta_d)'$. A special case of estimation for β is the least squares (LS) estimation with the scale parameter σ =1:

$$\hat{\beta} = (\sum_{i=1}^n x_i\, x_i^T)^{-1}(\sum_{i=1}^n x_i\, y_i^T), \tag{20}$$

where $\hat{\beta}$ means the estimation of β.

However, the LS approach can not effectively handle the presence of outliers for data modelling (Zhang et al., 2021c; 2022). Outliers generally introduce large errors, which significantly dominate the parameter estimation with the LS approach (Huang et al., 2015). For example, assuming 100 residuals for normal samples range from −0.5 and 0.5 while the only residual from the outlier is 100. Using the general LS approach to calculate the loss, the contribution from one outlier is much larger than the total contribution from the normal samples; correspondingly, the estimation is unreliable. Therefore, an *M*-estimation was proposed by Huber et al. (1973) to handle outliers to obtain a robust estimation. The *M*-estimation based on Huber's loss function minimizes the following formulation (Maronna, 1976):

$$\hat{\beta} = \underset{\beta}{\mathrm{Argmin}} \sum_{i=1}^{\mathrm{n}} \rho\left(\frac{y_i - x_i^T \beta}{\hat{\sigma}}\right), \tag{21}$$

where Argmin means to minimize the following formula, with a given estimate of the scale parameter $\hat{\sigma}$, the dispersion function $\rho(\cdot)$ for Huber's loss function is given as

$$\rho(\epsilon) = \begin{cases} \frac{1}{2}\epsilon^2, & |\epsilon| \le \tau \\ |\epsilon|\tau - \frac{\tau^2}{2}, & |\epsilon| > \tau \end{cases}, \tag{22}$$

where a hyper-parameter τ is used to control the loss calculation, when the absolute value of the residual ϵ is larger than τ, the l_1-norm loss function which is robust to outliers is used; otherwise, the traditional l_2-norm loss function is employed. The gradient of $\rho(\epsilon)$, i.e., $\Psi(\epsilon)$ can be obtained as

$$\Psi(\epsilon) = \begin{cases} \epsilon, & |\epsilon| \le \tau \\ \mathrm{Sign}(\epsilon) \times \tau, & |\epsilon| > \tau \end{cases}. \tag{23}$$

where Sign(.) is the sign function, which equals to 1 when $\epsilon > \tau$ and equals to -1 when $\epsilon < -\tau$.

Therefore, the robust estimator of β can be achieved by solving the equation

$$\sum_{i=1}^{n} x_i \Psi\left(\frac{y_i - x_i^T \beta}{\hat{\sigma}}\right) = 0_{n\times1}, \tag{24}$$

It can be found that two hyper-parameters, σ and τ, would determine the performance of these robust estimators. Following the recommendation of Fu et al. (2020), the robust estimator of $\hat{\sigma}$ is calculated by the median absolute deviation estimator as

$$\hat{\sigma} = \frac{\mathrm{Median}(|y_i - x_i^T \beta|)}{0.6745}, \tag{25}$$

where Median(.) is the median function which automatically ranks the order of (.) and selects median. For the second parameter τ, according to the work of Wang et al. (2007), a data-dependent tuning optimal estimated $\hat{\tau}^*$ is used in our model as

$$\hat{\tau}^* = \underset{\tau}{\text{Argmax}} \frac{\left\{\sum_{i=1}^{\mathrm{n}} I(|\hat{e}_i| \le \tau)\right\}^2}{\mathrm{n} \sum_{i=1}^{\mathrm{n}} \{I(|\hat{e}_i| \le \tau)\Psi^2(\hat{e}_i) + \tau^2 I(|\hat{e}_i| > \tau)\}}, \tag{26}$$

where Argmax means to maximize the following formula; $\hat{e}_i = (y_{\mathrm{i}} - x_i^T \beta)/\hat{\sigma}$; I(.) is an indicator function which equals to 1 if the condition (.) is satisfied, otherwise, I(.)=0. This tuning method is also popular in environmental modelling applications, in which the recommended sequence for τ ranges from 0 to 3 with an interval of 0.1 (Wang et al., 2018; Callens et al., 2020).

Finally, the training procedure for our robust regression with a tuning hyper-parameter τ can be implemented as follows:

Step 1. Obtain initial estimates of β with a median (i.e., l_1-norm) regression without the scale parameter as

$$\hat{\beta} = \underset{\beta}{\text{Argmin}} \sum_{i=1}^{\mathrm{n}} \left| y_i - x_i^T \beta \right|, \tag{27}$$

to obtain the initial residual for the following optimization steps.

Step 2. Obtain the estimate of scale parameter $\hat{\sigma}$ by Eq. (25).

Step 3. Obtain the tuning hyper-parameter $\hat{\tau}^*$ by solving Eq. (26) with the recommended τ sequence.

Step 4. Obtain the new β estimate with estimated $\hat{\sigma}$ and tuning $\hat{\tau}^*$ by solving Eq. (24).

3.2 Model selection

This subsection illustrates the procedure of model selection for our adaptive robust regression with the Akaike information criterion (AIC), which can evaluate the performance of each model based on the provided data. Akaike Information Criterion (AIC) from information theory is an indicator of the prediction error to measure the quality of the statistical models for the investigated data (Akaike, 1974). The AIC can effectively balance the performance of model fitting and the simplicity (i.e., it can handle the risks of overfitting and underfitting; overfitting means the model is too much dependent on the training dataset, which limits its application to other scenarios; underfitting means the model have not included all key parameters which limits the model performance). Considering the advantage of balancing the over- and under-fitting, AIC method was employed for model selection. The definition of AIC is given by Akaike (1974) as:

$$\mathrm{AIC} = -2\log(-Q) + 2K, \tag{28}$$

where K=d+3 is the number of independently adjusted parameters in the model with d the dimension of x_i; Q is the maximum likelihood function value for the adaptive robust regression:

$$Q = \prod_{i=1}^{\mathrm{n}} f(e_i) = \frac{1}{C^n(\hat{\tau}^*,\hat{\sigma})} \exp\left(-\sum_{i=1}^{\mathrm{n}} \rho_{\hat{\tau}^*,\hat{\sigma}}(e_i)\right), \tag{29}$$

where $f(.)$ is the probability density function,

$$C(\hat{\tau}^*,\hat{\sigma}) = \hat{\sigma}\sqrt{2\pi}[2\Phi(\hat{\tau}^*) - 1] + \frac{2\hat{\sigma}}{\hat{\tau}^*}\exp\left(-\frac{(\hat{\tau}^*)^2}{2}\right), \tag{30}$$

where $\Phi(\cdot)$ is the cumulative probability function of the standardized normal distribution. For model selection with the AIC, the model with the smallest AIC is recommended as the preferable model (Hastie et al., 2009). Therefore, there is a counterbalance between Q and K in Eq. (28).

4. Dataset for statistical regression modelling

The dataset used in this study consists of two parts: our experimental data for sands and silts (**Table 4**) and the data for sands from the literatures (**Table 5**). Considering that silt data are rare, it is expected that some comparisons can be made with abundant sand data. The entire training dataset is given in the Appendix.

4.1 Experimental results

Our laboratory experiments yield 13 data points for silts and 9 for sands.

Table 4. Results of our experiments.

Parameters		h	H_w	T_w	L	U_m	D	d_{50}	$\boldsymbol{\theta}$	KC	S/D	S
No.	Exp. No.	cm	cm	s	m	m/s	m	mm	Dimensionless			cm
1	G-1-1-03	50	17	1.9	3.81	0.346	0.050	0.287	0.158	13.2	0.42	2.1
2	G-1-1-06	50	17	1.9	3.81	0.346	0.050	0.287	0.158	13.2	0.28	1.4
3	G-1-2-01	50	8	1.3	2.31	0.108	0.040	0.287	0.030	3.5	0.23	0.9
4	G-1-2-02	50	17	1.3	2.31	0.232	0.040	0.287	0.098	7.6	0.25	1.0

5	G-1-2-03	50	17	1.9	3.81	0.346	0.040	0.287	0.158	16.4	0.22	0.88
6	G-1-2-04	50	8	1.3	2.31	0.108	0.040	0.287	0.030	3.5	0.00	0
7	G-1-2-05	50	17	1.3	2.31	0.232	0.040	0.287	0.098	7.6	0.00	0
8	G-1-2-06	50	17	1.9	3.81	0.346	0.040	0.287	0.158	16.4	0.00	0
9	G-1-2-07	50	17	1.9	3.81	0.346	0.035	0.287	0.158	18.8	0.29	1
10	G-2-1-03	50	17	1.9	3.81	0.346	0.050	0.057	0.454	13.2	0.54	2.7
11	G-2-1-06	50	17	1.9	3.81	0.346	0.050	0.057	0.454	13.2	0.40	2.0
12	G-2-2-03	50	17	1.9	3.81	0.346	0.040	0.057	0.454	16.4	0.75	3.0
13	G-2-2-06	50	17	1.9	3.81	0.346	0.040	0.057	0.454	16.4	0.33	1.3
14	G-2-2-07	50	17	1.9	3.81	0.346	0.035	0.057	0.454	18.8	0.66	2.3
15	G-3-1-03	50	17	1.9	3.81	0.346	0.050	0.033	0.680	13.2	0.50	2.5
16	G-3-1-06	50	17	1.9	3.81	0.346	0.050	0.033	0.680	13.2	0.26	1.3
17	G-3-2-01	50	8	1.3	2.31	0.108	0.040	0.033	0.103	3.5	0.00	0
18	G-3-2-02	50	17	1.3	2.31	0.232	0.040	0.033	0.377	7.6	0.10	0.4
19	G-3-2-03	50	17	1.9	3.81	0.346	0.040	0.033	0.680	16.4	0.30	1.2
20	G-3-2-04	50	8	1.3	2.31	0.108	0.040	0.033	0.103	3.5	0.18	0.7
21	G-3-2-06	50	17	1.9	3.81	0.346	0.040	0.033	0.680	16.4	0.60	2.4
22	G-3-2-07	50	17	1.9	3.81	0.346	0.035	0.033	0.680	18.8	0.34	1.2

Note: Literature data are based on non-homogeneous particle size distribution, same as the scenario in our experiments. However, the existing researches in this field normally used d_{50} as the representative parameter in the analysis. Therefore, we can only collect the d_{50} from literatures. Future work will pay attention to the influence of non-homogeneous particle size distribution on scouring process.

4.2 Collected data from the literature

In addition to the experimental results above, more data are collected from the literature to form a larger dataset for statistical learning. The source of the collected data is given in **Table 5**, while the data are given in the Appendix. All literature datasets obtained are under the wave-only actions, same as the scenario in our experiments.

Table 5. Summary of the data used in the present study.

Dataset	Experiment performer	Data quantity	Literature that provided the data
Training set for the sand model or the generalized model	The present experiment	22 (9 sands+13 silts)	
	Sumer and Fredsøe (1990)	25	Sumer and Fredsøe (1990)
	Sumer and Fredsøe (1996)	19	Cheng et al. (2020)
	Kim et al. (2011)	45	
	Mattioli et al. (2013)	9	
	Zhang et al. (2017)	7	Zhang et al. (2017)
	Dogan and Arisoy (2015)	20	Dogan and Arisoy (2015)
	Zang et al. (2019)	36	Zang et al. (2019)
	Fredsøe et al. (1992)	4	Fuhrman et al. (2014)
	Subtotal	187	
Test set for sands	Mousavi et al., (2009)	9	Kazeminezhad et al. (2010)
	Pu et al., (2001)	8	
	Subtotal	17	
Total		204	

Note: The collected data are detailed in the Appendix. The *S*/*D* of Zhang et al. (2017) is the average of S_{ef}, S_{ep}, and S_{eb} therein.

4.3 Data comparability

To use a collected dataset, it is necessary to evaluate the comparability of data from different works. As the parameters involved are all nondimensional, the comparability is generally good, except for the Shields parameter θ (Eq. 14), which shows a large deviation in different data origins. This deviation can be explained by Eqs. (14-18). A wave friction factor f_{w} is necessary for the computation of θ, but not all the literature chose the same formula for f_{w}. To make all the θ values comparable, θ is recalculated in the present work for the entire dataset.

The judgement of live-bed or clear-water regimes is also considerred here. Sumer and Fredsøe (1990) pointed out that the effect of the Shields parameter on *S* is quite weak for live-bed situations, while Etemad-Shahidi et al. (2011) found that the Shields number is very important in clear-water conditions. For sandy sediments, θ_{cr} is normally calculated with Eqs. (31-32) (Soulsby, 1997):

$$\theta_{\mathrm{cr}} = \frac{0.24}{d_*} + 0.055\,(1 - \exp(-0.02 d_*)), \tag{31}$$

where d_* is the dimensionless diameter of the bed sand:

$$d_* = d_{50}(\frac{(\rho_s-\rho)g}{\rho v^2})^{1/3}, \tag{32}$$

where v=10^{-6} m^2/s is the kinematic viscosity.

It is found that θ_{cr}=0.06 for the dataset in the present study, which is quite consistent with the dividing number of θ_{cr}=0.064 in previous works for sands as referred in the introduction (Eqs. 6-9). However, the criterion of Soulsby (1997) was for sands. As far as the authors' concern, no universal criterion for the entrainment of cohesive sediments is available. But according to the Shields curve, the θ_{cr} for silts should be even smaller than sands, thus most of the cases in the present study are in live-bed regime. Considering the two points mentioned above, we tried to train the model without distinguishing the two regimes, to test if the transport regime is critical.

The model scale of the present experiment was 1:20, while the scales of the other experiments from literature may not be the same. In this inevitable situation of data from different literatures, we tried to use dimensionless parameters. If the finally-derived models choose the dimensionless parameters, one can believe the different scalings in the present study has little impact.

The other parameters collected from the literature show good comparability and therefore are reliable for analysis and modelling. The last point that needs to be declared is that the data used in the present work are all for buried ($e/D < 0$) or in-contact (e/D=0) pipelines, and no suspended pipeline scenarios are included ($e/D > 0$).

5. Data-driven model development

In this section, firstly, different transformations were attempted for all the investigated parameters (independent variables) to obtain their best correlations with the response variable (ln S/D). Secondly, a series of robust regression models were built by covering any permutation and combination of the parameters using the sand dataset, correspondingly, an optimal formula for S/D in sands was obtained based on the AIC value. Outliers in the dataset were detected and handled with a newly-proposed adaptive robust regression to improve the modeling accuracy. Thirdly, the optimal formula for sands was tested for predicting the silt data, but the results were not good enough. Therefore, fourthly, another training was performed with the combined dataset of the silt data and the original sand data. A dummy variable η was introduced to distinguish the sediment types. An optimal prediction formula was finally found for both sandy and silty seabeds.

5.1 Data pre-processing

For nonlinear regression, it is a conventional method to firstly deform the variables in different forms, find the transformed form which has the the maximum correlation coefficient with the dependent variable (ln S/D), then carry out regression modeling. Here, maximum correlation coefficient means the largest absolute correlation coefficient between all kinds of transformed independent variables and the dependent variable. Different transformations (logarithmic and different polynomial transformations with the highest order term from - 1 to 1 with a step of 0.1) were tried for all the seven independent parameters (U_m, wave period T_w, pipe diameter D, e/D, grain size d_{50}, Shields parameter θ, and KC) to find the best correlations with the response (ln S/D) according to the maximum correlation coefficients. The best transformations for the parameters are listed in **Table 6**. The transformed variables are denoted as ln S/D, ln KC, e/D, ln U_m, ln T_w, D^{-1}, ln d_{50}, and $\theta^{0.5}$.

Table 6. Best transformations for the investigated parameters.

Independent variables							
Original parameter	U_m	T_w	D	e/D	d_{50}	θ	KC

Best transformation form	ln U_m	ln T_w	D^{-1}	e/D	ln d_{50}	$\theta^{0.5}$	ln KC
Max correlation coefficient	0.4674	0.5311	0.3642	0.3275	-0.1483	0.4369	0.7132

In addition, for the prediction of silts in Section 5.3, the grain size is divided into two types (sands and silts) according to the criterion of d_{50}=0.0625 mm, and a dummy variable η is introduced as η = 1 for silts and η = 0 for sands.

The correlations between all the investigated factors and the response ln S/D are plotted in **Figure 3**. Note that the numbers on the axis of each box are the value range of any two variables which are marked in the boxes along the diagonal of **Figure 3**; the numbers in each box are the correlation coefficients between any two variables involved, e.g., 0.71 means that the correlation coefficient between ln S/D and ln KC is 0.71.

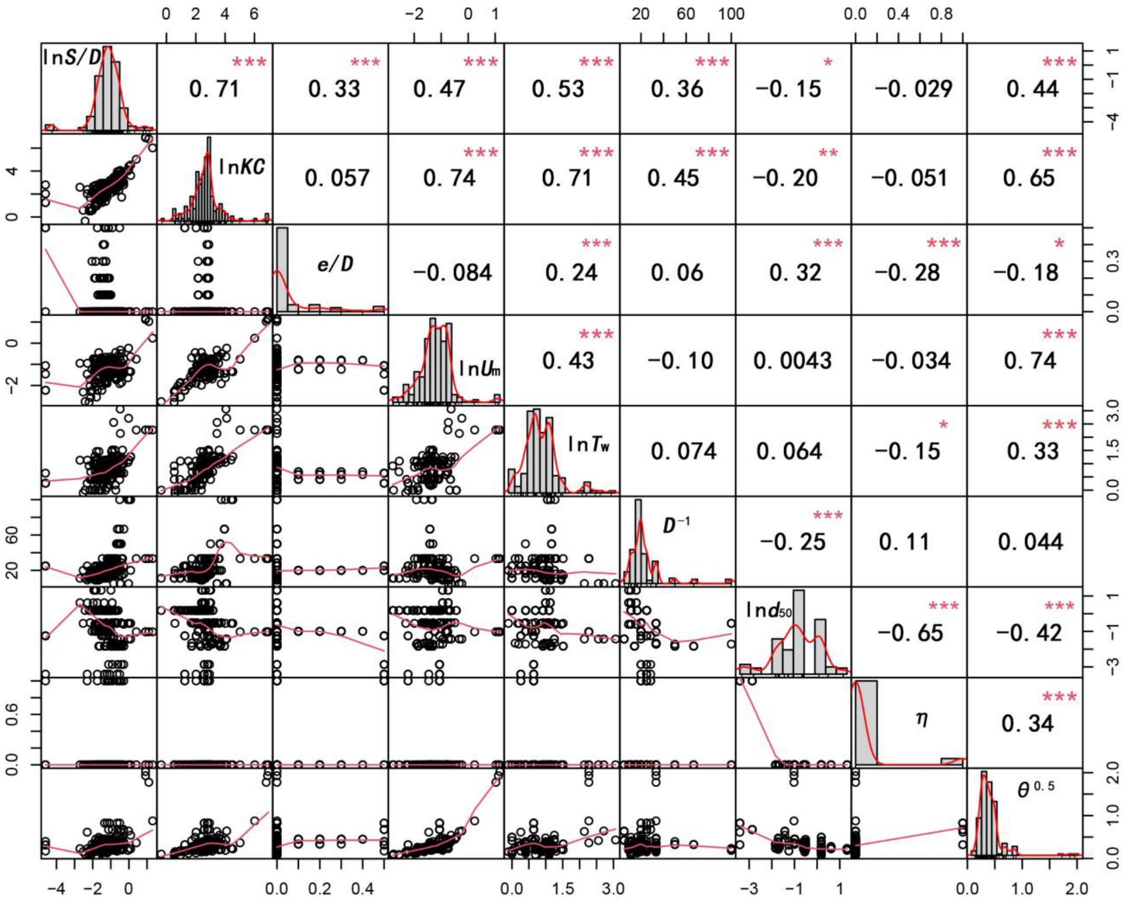

Figure 3. Correlations between the transformed parameters and the response (ln S/D). Note that the numbers on the axis of each box are the value range of any two variables which are

marked in the boxes along the diagonal; the numbers in each box are the correlation coefficients between any two variables involved, e.g., 0.71 means that the correlation coefficient between ln *S*/*D* and ln *KC* is 0.71. η is the dummy variable that is used to identify the type of sediments. The asterisk in the figure represents the significance level of the correlation.

It is worth noting that the specific correlation coefficient between ln *S*/*D* and a specific variable can be changed when more variables are involved while doing the ARR modeling. However, this will not affect the final regression performance in essence, as finding the best transformations between ln *S*/*D* and a specific variable is just a preparatory step. With the respective best transformation form, it is still the scheme that most likely to obtain the best nonlinear regression result while all the independent variables are involved, although the coefficient and contribution of each parameter will be slightly adjusted automatically in the subsequent ARR modeling process.

In the following ARR statistical modelling work, LS approach was firstly used for ordinary regression, then the residuals from the LS approach are examined. If the residuals are normal (i.e., residuals follow a normal distribution), then LS approach is enough to solve this problem. However, if the residual from the LS regression is not normally distributed, i.e., outliers were found by checking with the QQ (quantile-quantile) plot. Our ARR method was used to further diagnose the investigated dataset to obtain an optimal model by handling the outliers.

5.2 Model training with only the sand data

First, the sand dataset was used for the training, and the estimates for the three-parameter model with an ordinary regression method (LS approach) are:

$$\ln\left(\frac{S}{D}\right) = 0.51\ln(KC) + 3.15\frac{e}{D} - 0.14\ln(d_{50}) - 2.49 + \epsilon, \tag{33}$$

with residuals ϵ included. Then, a QQ (quantile-quantile) plot and box plot is used to check the distribution of residuals from formula (33) in **Figure 4**. The quantile-quantile plot is a graphical method for determining whether the data samples come from the same population or not. The order statistics of the sample are plotted against the corresponding standard values from the assumed distribution (Aly and Aydin, 1988). In the left subfigure of Figure 4, when the dots well follow the blue line, it means the residuals follow a normal distribution without any outliers; however, when residuals are larger or smaller, i.e., the dots are far away from the blue line, outliers exist. In addition, in the right subfigure of Figure 4, according to the box plot, it is also apparent that there are many outliers that differ significantly from the rest of the dataset (out of the normal range). Therefore, the new adaptive robust regression described in Section 3 is employed to handle the outliers in the original dataset.

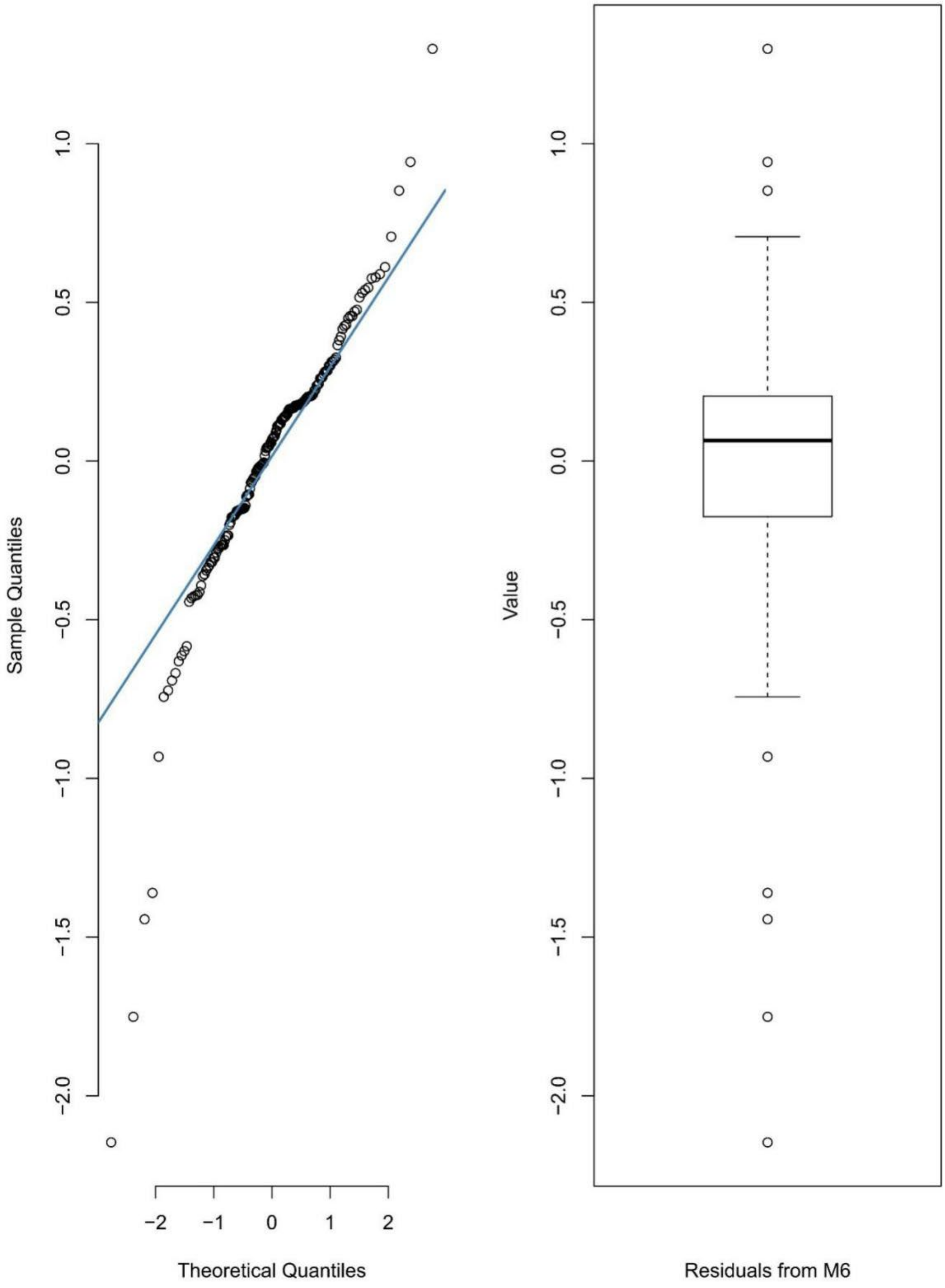

Figure 4. QQ (quantile-quantile) plot and box plot for the residuals from the three-parameter model with the least squares approach. The presence of outliers was clearly detected. The black dots close to the blue (reference) straight line indicate that the corresponding residuals are normally distributed. In contrast, when the black dots are beyond the upper left or under the lower right of the blue line, the dots are suspicious outliers.

Our first attempt was to build a model for sands. Therefore, all combinations of the seven parameters were investigated using sand data. The corresponding estimates for all the models are listed in **Table 7**.

Table 7. Coefficient estimates from the adaptive robust regression for sands.

Model No.	$\boldsymbol{\sigma}$	$\boldsymbol{\tau}$	ln *KC*	*e*/*D*	ln U_m	ln T_w	D^{-1}	ln d_{50}	$\boldsymbol{\theta}^{0.5}$	*c*	AIC
One-parameter Model											
M1	0.30	0.8	0.52	-	-	-	-	-	-	-2.61	490.64
Two-parameter Model											
M2	0.24	0.4	0.52	1.72	-	-	-	-	-	-2.41	667.19
Three-parameter Model											
M3	0.23	0.9	0.56	1.60	-0.11	-	-	-	-	-2.69	500.01
M4	0.24	0.3	0.54	1.77	-	-0.07	-	-	-	-2.41	740.14
M5	0.24	0.4	0.49	1.65	-	-	0.00	-	-	-2.42	668.75
M6	0.26	1.0	0.50	1.95	-	-	-	-0.07	-	-2.43	433.00
M7	0.23	0.4	0.54	1.64	-	-	-	-	-0.17	-2.42	673.51
Four-parameter Model											
M8	0.24	0.4	0.61	1.63	-0.13	-0.10	-	-	-	-2.73	672.22
M9	0.23	0.8	0.54	1.60	-0.08	-	0.00	-	-	-2.62	522.97
M10	0.26	0.9	0.54	1.76	-0.08	-	-	-0.05	-	-2.62	470.55
M11	0.23	0.9	0.56	1.60	-0.12	-	-	-	-0.03	-2.71	505.07
M12	0.24	0.4	0.51	1.70	-	-0.03	-0.00	-	-	-2.41	669.37
M13	0.26	1.0	0.54	2.12	-	-0.11	-	-0.08	-	-2.43	444.15

M14	0.24	0.3	0.57	1.70	-	-0.08	-	-	-0.19	-2.42	739.33
M15	0.26	1.0	0.49	1.88	-	-	0.00	-0.06	-	-2.44	439.06
M16	0.23	0.4	0.50	1.63	-	-	0.00	-	-0.05	-2.42	675.28
M17	0.25	0.9	0.52	1.88	-	-	-	-0.06	-0.09	-2.43	477.18
Five-parameter Model											
M18	0.25	1.0	0.77	1.76	-0.31	-0.24	-0.00	-	-	-3.16	446.98
M19	0.26	1.0	0.59	1.94	-0.10	-0.13	-	-0.06	-	-2.68	442.77
M20	0.25	0.9	0.61	1.72	-0.16	-0.11	-	-	0.06	-2.80	473.10
M21	0.25	0.8	0.51	1.76	-0.04	-	0.00	-0.05	-	-2.52	509.10
M22	0.23	0.5	0.53	1.57	-0.07	-	0.00	-	-0.00	-2.58	620.65
M23	0.26	0.9	0.54	1.75	-0.09	-	-	-0.05	0.01	-2.63	472.76
M24	0.27	0.9	0.52	2.01	-	-0.08	0.00	-0.07	-	-2.43	465.55
M25	0.24	0.4	0.53	1.68	-	-0.05	0.00	-	0.11	-2.42	673.14
M26	0.25	0.9	0.56	2.04	-	-0.11	-	-0.07	-0.12	-2.44	487.34
M27	0.26	1.0	0.49	1.88	-	-	0.00	-0.06	-0.01	-2.44	439.41
Six-parameter Model											
M28	0.25	1.0	0.72	1.92	-0.25	-0.23	-0.00	-0.06	-	-3.02	449.73
M29	0.26	1.0	0.77	1.76	-0.32	-0.24	-0.00	-	0.04	-3.02	443.30
M30	0.26	1.0	0.59	1.93	-0.11	-0.13	-	-0.06	0.02	-2.69	444.16
M31	0.25	0.8	0.51	1.76	-0.04	-	0.00	-0.05	0.01	-2.52	510.87
M32	0.25	0.9	0.54	2.01	-	-0.09	0.00	-0.07	-0.07	-2.44	479.76
Seven-parameter Model											
M33	0.24	0.8	0.72	1.85	-0.25	-0.22	-0.00	-0.05	-0.03	-3.01	517.18

Note: *c* is the constant term of the model; AIC is used in the model selection rather than the determination coefficient R^2. Although a model with more parameters should correspond to a larger R^2, too many parameters may also result in overfitting, thus the generalization of the model would decrease. AIC is used for model selection here as it has two components: likelihood term and penalty term which counterbalance with each other. Physically, minimizing AIC would balance the model complexity and the generalization of the model. For example, introducing more less relevant variables may bring a small improvement of the likelihood term (model performance) but also a large penalty term to make the model clear and simple.

The best robust regression model for the sand data is the three-parameter model M6 with AIC = 433.00 (**Table 7**):

$$\ln\left(\frac{S}{D}\right) = 0.50\ln(KC) + 1.95\frac{e}{D} - 0.07\ln(d_{50}) - 2.43 + 0.26\epsilon, \qquad (34)$$

Eq. (34) is much more concise than Eq. (33), the reason is that Eq. 33 was derived before handling the outliers. Ignoring the residual term and taking the power function with e as the base on both sides of the equation at the same time, Eq. (34) is back-transformed to:

$$\frac{S}{D} = 0.09\cdot\sqrt{KC}\cdot 0.14^{-\frac{e}{D}}\cdot {d_{50}}^{-0.07}, (e/D > \text{-}0.5), \qquad (35)$$

According to the clear relationship in Eq. (35), the form of our model is generally similar to that of SF1990, but *e*/*D* is further incorporated which shows a negative effect on the scour process. The grain size also slightly influences the scouring of sands beneath pipelines.

5.3 Model training with both the sand and silt data

As no specific model is available for predicting the scouring of silts, the optimal model for sands, namely, Eq. (35) was tested for predicting the *S*/*D* of silts with our experimental data. The results are shown together with the difference between the prediction and observation in **Table 8**. The difference can not be ignored, the root mean square error (RMSE) is 0.21 and mean absolute error (MAE) is 0.19, thus the predictions are not satisfactory.

Table 8. Test results of the performance of Eq. (35) for predicting the silt data.

Exp. No. for silts	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
Observed S/D	0.54	0.40	0.75	0.33	0.66	0.52	0.26	0.01	0.10	0.30	0.18	0.60	0.34
Predicted S/D	0.40	0.15	0.44	0.17	0.47	0.41	0.16	0.21	0.31	0.46	0.08	0.17	0.49
Difference	0.14	0.25	0.31	0.16	0.19	0.11	0.1	-0.2	-0.21	-0.16	0.1	0.43	-0.15

Given that silt data are rare at the present stage, it is impossible to build a data-driven model for silts specifically. An alternative is the attempt to build a joint model for sands and silts. Therefore, the whole dataset in the present work (sands+silts) was employed to establish a more generalized model for *S*/*D*. Similar modelling processes were performed again with adaptive robust regression but with the sand+silt dataset for all the permutations and combinations of the parameters as done in Section 5.2. In this round, the optimal model with AIC = 469.99 is found as:

$$\ln\left(\frac{S}{D}\right) = 0.59\ln(KC) + 1.30\frac{e}{D} - 0.17\ln(U_\mathrm{m}) - 0.10\ln(T_\mathrm{w}) - 0.08\ln(d_{50}) + 0.15\sqrt{\theta} + 0.27\epsilon, \tag{36}$$

This model is rather complex for practical application, therefore, further examinations were performed by introducing a dummy variable η to distinguish the type of sediment. With η included in the modelling, the estimates of each model and corresponding AIC value are reported in **Table 9**.

Table 9. Coefficient estimates from adaptive robust regression for sands and silts with the dummy variable included.

Model No.	$\boldsymbol{\sigma}$	$\boldsymbol{\tau}$	ln *KC*	*e*/*D*	ln U_m	ln T_w	D^{-1}	ln d_{50}	$\boldsymbol{\theta}^{0.5}$	η	*c*	AIC
One-parameter Model												
M1D	0.34	1.3	0.54	-	-	-	-	-	-	0.18	-2.57	391.92
Two-parameter Model												
M2D	0.24	0.1	0.51	1.43	-	-	-	-	-	0.47	-2.41	1124.524
Three-parameter Model												
M3D	0.23	0.1	0.58	1.25	-0.14	-	-	-	-	0.48	-2.77	1127.32
M4D	0.24	0.1	0.53	1.45	-	-0.04	-	-	-	0.47	-2.42	1126.17
M5D	0.23	0.4	0.49	1.35	-	-	0.00	-	-	0.48	-2.43	731.912
M6D	0.25	1.0	0.51	1.46	-	-	-	-0.05	-	0.38	-2.45	514.64

M7D	0.23	0.3	0.55	1.34	-	-	-	-	-0.22	0.55	-2.44	799.94
Four-parameter Model												
M8D	0.22	0.3	0.61	1.30	-0.17	-0.08	-	-	-	0.47	-2.81	804.74
M9D	0.22	0.1	0.56	1.26	-0.12	-	0.00	-	-	0.48	-2.70	1130.07
M10D	0.23	0.8	0.57	1.27	-0.13	-	-	-0.02	-	0.44	-2.74	578.32
M11D	0.22	0.1	0.58	1.24	-0.11	-	-	-	-0.14	0.53	-2.69	1129.22
M12D	0.23	0.4	0.49	1.35	-	-0.04	-0.00	-	-	0.48	-2.43	733.69
M13D	0.26	1.0	0.53	1.53	-	-0.06	-	-0.06	-	0.36	-2.45	505.87
M14D	0.24	0.3	0.57	1.37	-	-0.06	-	-	-0.22	0.54	-2.44	792.87
M15D	0.23	0.6	0.49	1.36	-	-	0.00	-0.02	-	0.44	-2.43	644.19
M16D	0.22	0.4	0.51	1.33	-	-	0.00	-	-0.11	0.51	-2.42	675.28
M17D	0.23	0.7	0.54	1.38	-	-	-	-0.03	-0.16	0.47	-2.43	477.18
Five-parameter Model												
M18D	0.23	0.4	0.74	1.29	-0.31	-0.18	-0.00	-	-		-3.16	728.63
M19D	0.24	0.8	0.60	1.34	-0.15	-0.08	-	-0.02	-		-2.80	569.46
M20D	0.22	0.1	0.61	1.29	-0.14	-	-	-	0.10		-	1131.97

						0.09					2.73	
M21D	0.23	0.6	0.55	1.26	-0.11	-	0.00	-0.00	-		-2.69	648.96
M22D	0.22	0.1	0.57	1.25	-0.10	-	0.00	-	-0.12		-2.65	1131.30
M23D	0.23	0.8	0.57	1.27	-0.12	-	-	-0.02	-0.01		-2.73	580.89
M24D	0.25	0.9	0.49	1.42	-	-0.01	0.00	-0.04	-		-2.45	540.48
M25D	0.23	0.4	0.53	1.34	-	-0.02	0.00	-	0.12	0.52	-2.44	730.02
M26D	0.24	0.1	0.58	1.39	-	-0.05	-	-0.01	-0.26	0.54	-2.46	1129.87
M27D	0.23	0.6	0.51	1.35	-	-	0.00	-0.02	-0.08	0.47	-2.45	641.22
Six-parameter Model												
M28D	0.24	0.7	0.73	1.31	-0.30	-0.18	-0.00	-0.01	-	0.43	-3.13	602.82
M29D	0.23	0.4	0.75	1.28	-0.29	-0.18	-0.00	-	0.10	0.50	-3.12	730.12
M30D	0.24	0.8	0.60	1.34	-0.15	-0.08	-	-0.02	0.00	0.42	-2.80	571.36
M31D	0.24	0.8	0.55	1.29	-0.10	-	0.00	-0.02	0.00	0.43	-2.69	578.14
M32D	0.23	0.1	0.55	1.38	-	-0.05	0.00	-0.01	-0.18	0.51	-2.43	1131.84
Seven-parameter Model												
M33D	0.23	0.7	0.74	1.32	-0.29	-0.18	-0.00	-0.01	-0.04	0.44	-3.12	608.74

Note: *c* is the constant term of the model.

According to the smallest AIC value (391.92) shown in **Table 9**, the best model for *S*/*D* is very elegant with only two parameters, ln *KC* and η, and the formulation can be given as:

$$\ln\left(\frac{S}{D}\right) = 0.54\ln(KC) + 0.18\,\eta - 2.57 - 0.34\epsilon, \qquad (37)$$

where η=1 for silts and η=0 for sands. The AIC value of Eq. (37) is 391.92, which is much smaller than that of Eq. (36) (AIC=469.99) before incorporating the dummy variable. Ignoring the residual item, and taking the power function with e as the base on both sides of the equation at the same time, the optimal *S*/*D* model for sands and silts is back-transformed to:

$$\frac{S}{D} = 0.08 \cdot \alpha^{\eta} \cdot KC^{0.54}, \ (e/D > \text{-}0.5), \qquad (38)$$

where α is an empirical coefficient; here, α=1.2 for the silts from the Yellow River Delta. Again, the model was found to have a similar shape to that of SF1990, but the influence of sediment type was successfully detected. Under the same *KC* number, the equilibrium scour depth over silts is generally 1.2 times that of sands for the Yellow River silts. The influence of sediment type is even bigger than that of *e*/*D* in Eq. (38). However, this does not mean that the influence of *e*/*D* is always weaker than sediment types, as only cases with *e*/*D* > -0.5 are involved in the present work. Future works will be extended to deeper burial (*e*/*D* < -0.5) and suspended (*e*/*D* > 0) scenarios.

6. Validation

The optimal models proposed in this paper are tested and validated in this section.

6.1 Test of the proposed sand model for predicting sands

The first validation was performed for sands. To this end, 30 sets of sand test data (cf. Appendix) from Mousavi et al. (2009) and Pu et al. (2011), which were not used in the model training in Section 5, were used to test the performance of the proposed sand model (Eq. 35). Comparisons were conducted with the three process-based models of Sumer and Fredsøe (1990), Cevik and Yuksel (1999), Sumer and Fredsøe (2002) and the two data-driven models of Etemad-Shahidi et al. (2011) and Sharafafi et al. (2018), which have been given in the Introduction section and here abbreviated as SF1990, CY1999, SF2002 and ES2011, Sh2018, respectively. The models of Mousavi et al. (2006) and Pu et al. (2011) were not used as comparison models because the Mousavi et al. (2006) model is valid only for *KC* < 6, and the complete form of the Pu et al. (2011) model is not available, as mentioned in the Introduction. Therefore, the sand data from Mousavi et al. (2006) and Pu et al. (2011) were used as the test set of sands. The test results are shown in **Figure 5**. Note that the data for suspended ($e/D > 0$) pipelines in Pu et al. (2011) were not used here, as the present study concerns only cases in which $e/D \leq 0$ (in-contact or buried pipelines).

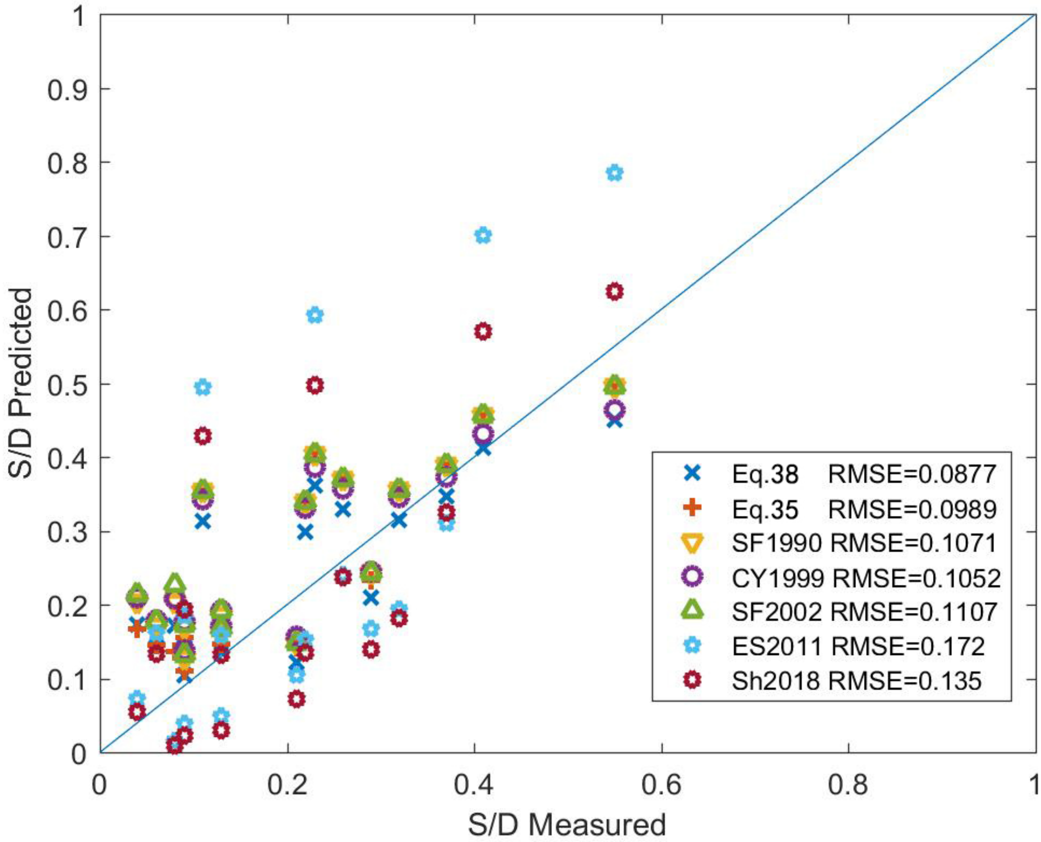

Figure 5. Comparisons of *S*/*D* modelling for the test set of sands (sand data collected from Mousavi et al. (2009) and Pu et al. (2001)). Note that Eq. (35) is our proposed best model for sands. Eq. (38) (with the dummy variable η incorporated) is the best model that we proposed for both sands and silts. Here, η=0 because it is the validation for sands. SF1990, CY1999, SF2002, ES2011 and Sh2018 refer to the three process-based models of Sumer and Fredsøe (1990), Cevik and Yuksel (1999), Sumer and Fredsøe (2002) and two data-driven models of Etemad-Shahidi et al. (2011) and Sharafafi et al. (2018), respectively.

It can be seen from **Figure 5** that Eq. (35) and Eq. (38) are the best two among all seven models. The RMSE of the proposed sand model (Eq. 35) is 0.0989, while the RMSE of the proposed sand/silt model (Eq. 38) is 0.0877, both are smaller than the others. The validation results demonstrate that our proposed models are more effective in predicting scouring in sandy seabeds.

6.2 Test of the proposed generalized model for predicting silts

Secondly, further validation was performed for the silts. As silt data are only available from our experiments, the 13 sets of silt data were used as the test set. Comparisons were also made to the five models above, and the results are shown in **Figure 6**.

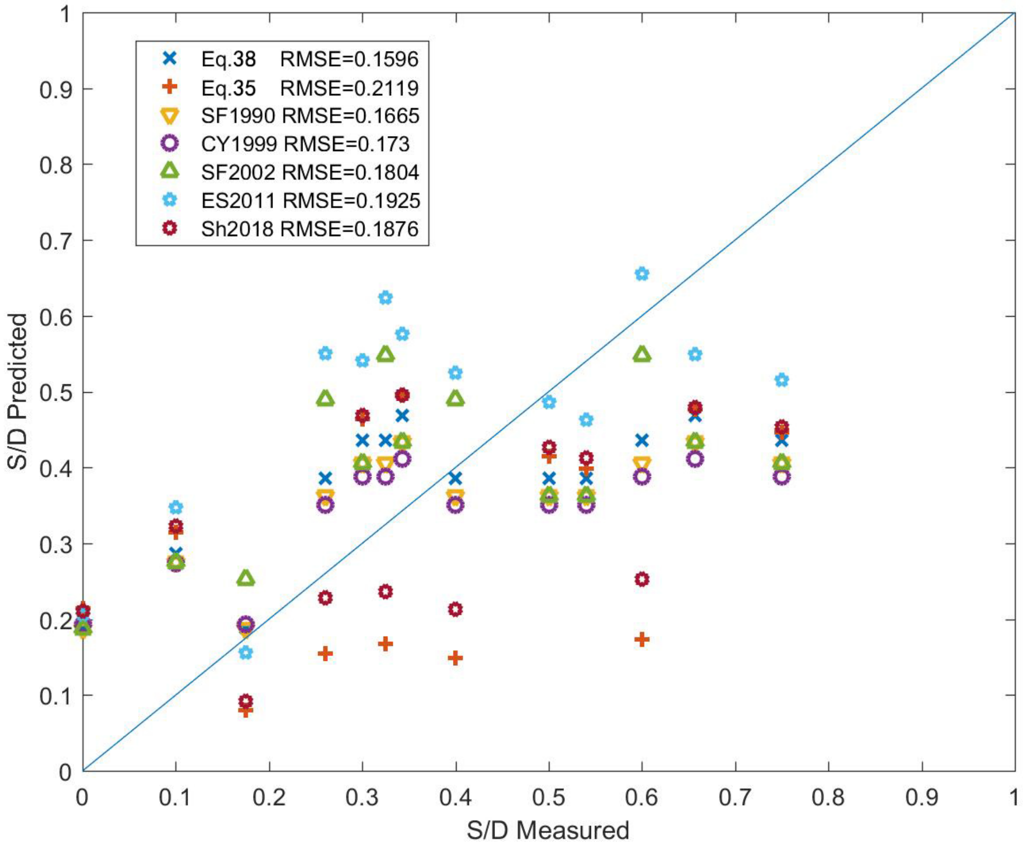

Figure 6. Comparisons of *S*/*D* modelling for the test set of silts (our experimental data). Note that the proposed Eq. (38) (with the dummy variable η incorporated) is the best model for the prediction of scouring in both sands and silts; here, η=1, as it is the validation for silts. Eq. (35) is the proposed best model for predicting the scouring of sands, but it is nevertheless presented here for comparison. SF1990, CY1999, SF2002 and ES2011, Sh2018 refer to the three process-based models of Sumer and Fredsøe (1990), Cevik and Yuksel (1999), Sumer and Fredsøe (2002) and two data-driven models of Etemad-Shahidi et al. (2011) and Sharafafi et al. (2018),

respectively.

Figure 6 shows that although SF1990 is the best process-based model with an RMSE of 0.1665, the proposed Eq. (38) gives the best prediction (with the smallest RMSE of 0.1596) among all seven models. This result demonstrates that the newly-proposed model for both sands and silts (Eq. 38) is the most effective model for predicting scour in silty seabeds. The reason for this better performance is that Eq. (38) distinguishes the sediment type with the dummy variable η. This implicitly indicates a different scour process between sands and silts from the data-driven perspective.

6.3 Test of the proposed generalized model for predicting both sands and silts

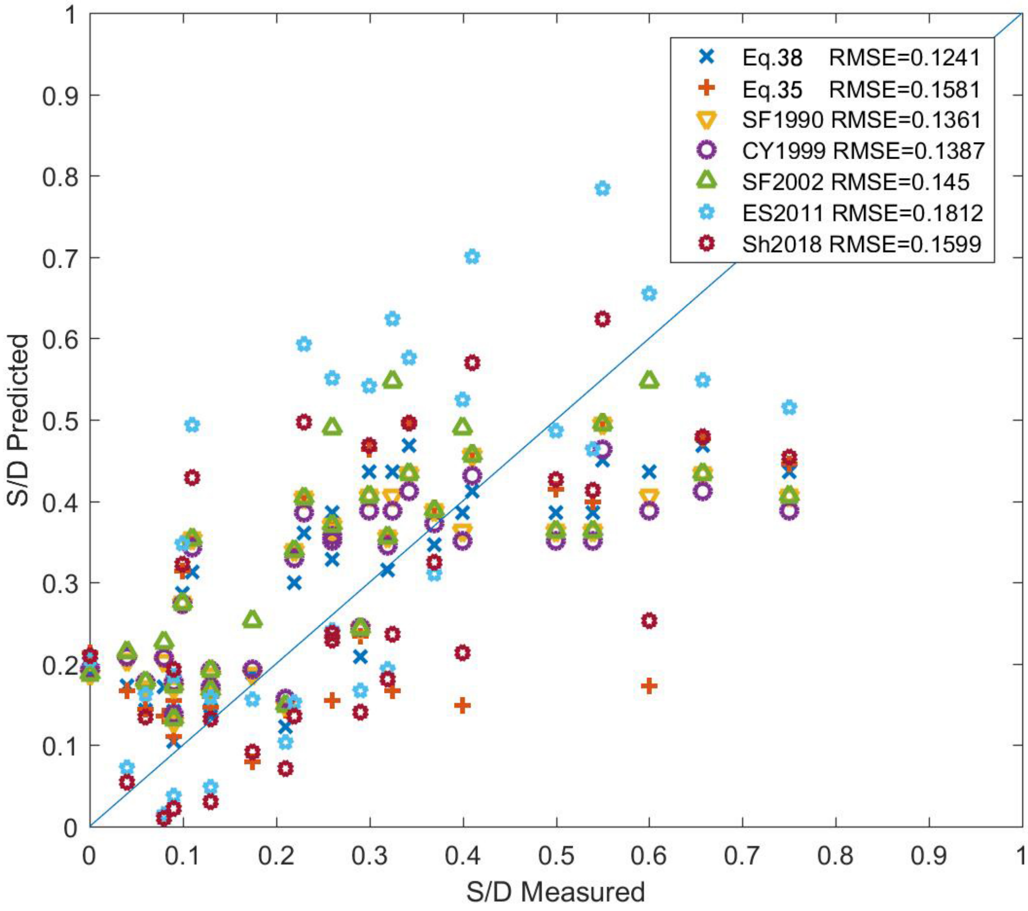

Figure 7. Comparisons of S/D modelling for the combined test set of silts and sands. Note that Eq. (38) (with the dummy variable η incorporated) is the best model that we proposed for predicting the scouring of both sands and silts. Eq. (35) is the best model that we proposed for predicting the scouring of sands only. SF1990, CY1999, SF2002, ES2011 and Sh 2018 refer to the three process-based models of Sumer and Fredsøe (1990), Cevik and Yuksel (1999), Sumer

and Fredsøe (2002) and two data-driven models of Etemad-Shahidi et al. (2011) and Sharafafi et al. (2018), respectively.

When the test sets of sands (17 sets) and silts (13 sets) are combined and used for model validation, Eq. (38) still shows the best performance, with the smallest RMSE of 0.1241 (**Figure 7**).

7. Conclusions

The scour depth beneath submarine pipelines is very important for ocean engineering. However, few quantitative models have been established for silty seabeds compared to noncohesive sands. In this study, firstly, laboratory experiments were conducted for both silty and sandy seabeds; secondly, data from the literature were collected and combined with our experimental data to form a 204-set dataset, which is the most abundant dataset to date on this topic. Based on this dataset, two statistical learning models were established for sands only and sands and silts together for predicting the equilibrium scour depth beneath pipelines under waves. Detailed conclusions derived from the present work are given as follows:

1. Adaptive robust regression is effective in handling the outliers in the dataset, thereby improving the prediction accuracy for the equilibrium scour depth under pipelines. The proposed models not only outperform three commonly-employed process-based models and two data-driven models in accuracy but also show good interpretation in physics.

2. A simple formula for predicting the *S*/*D* beneath pipelines under waves in sandy seabeds is suggested:

$$\frac{S}{D} = 0.09 \cdot \sqrt{KC} \cdot 0.14^{-\frac{e}{D}} \cdot {d_{50}}^{-0.07}, \ (e/D > \text{-}0.5),$$

indicating that the scour in sands is mostly related to the *KC* number and initial pipeline-seabed gap (*e*/*D*) but is also weakly related to the grain size of sediments (d_{50}). Note that here *e*/*D* is negative when the pipeline is buried, not the absolute value.

3. A generalized model for predicting the *S*/*D* beneath pipelines under waves in both sandy and silty seabeds is suggested:

$$\frac{S}{D} = 0.08 \cdot \alpha^{\eta} \cdot KC^{0.54}, \ (e/D > \text{-}0.5),$$

with a dummy variable (η) incorporated to distinguish sands from silts. α is an empirical coefficient, where α=1.2 for the silts from the Yellow River Delta in China. This model indicates that the scour beneath pipelines is mostly related to the *KC* number and sediment type. With the same *KC* number, the equilibrium scour depth in silts is generally 1.2 times that in sands for Yellow River silts.

The present work not only provides more experimental data but also contributes two practical formulas to the pipeline scour community. However, the models proposed in the present paper are limited to $e/D > -0.5$, and future works may extend to $e/D < -0.5$ or suspended scenarios ($e/D > 0$).

Acknowledgments

This work was supported by the National Key R&D Program-Intergovernmental Key Special Project [grant number 2017YFE0133500]; the Natural Science Foundation of Shandong Province [grant number ZR2019BD009]; the National Natural Science Foundation of China [grant numbers 41976198 and 41807229]; the Project of Taishan Scholar [grantee G. Li]; the Study Abroad Program of the Ocean University of China [grantee Y. Zhang] and the Study Abroad Program of Shandong Province (International Exchange Program of Postdoctoral Research Fellows) [grant number 201801026, grantee S. Zhang]. Jinghao Shi, Zhiheng Zhang and Chunyan Zhou are acknowledged for their contributions to the experiments.

Notation

D	pipe diameter
d_{50}	median grain size
h	water depth
e [italic]	clearance between the pipe and undisturbed bed
e	the base of the natural logarithm
KC	Keulegan-Carpenter number
θ	Shields parameter
θ_{cr}	Critical θ for the initiation of the motion of bed-material particles
Ur	Ursell parameter
S	equilibrium scour depth

H_w	wave height
T_w	wave period
L	wavelength
g	acceleration due to gravity
ρ	density of water
ρ_s	density of sediment particles
U_m	maximum water particle velocity on the bed in the absence of the pipe
U_*	friction velocity
τ_w	wave shear stress
f_w	wave friction factor
A	semi-excursion of wave particles at the bottom
d_*	dimensionless diameter of the sands
v	kinematic viscosity
η	dummy variable
c	constant term of the model
m	constant related to bed materials in Pu et al. (2001)
B	function of e/D in Pu et al. (2001)
α	empirical coefficient for the proposed model
k	$=2\pi/L$ is the wavenumber
x_i	independent variable
y_i	dependent variable
R	real number field
n	sample amount of the investigated data set
d	dimension of x_i
ϵ	residual
μ	noise
σ	scale parameter
β	coefficient of the regressor
$\hat{\beta}$	estimation of β

τ	hyper-parameter
$\hat{\tau}^*$	estimated data-dependent tuning optimal hyper-parameter
$\hat{\sigma}$	estimated scale parameter
$\hat{e}_i$	$(y_i - x_i^T\beta)/\hat{\sigma}$
Q	maximum likelihood function value
K	number of independently adjusted parameters
$I(\cdot)$	indicator function
$\Phi(\cdot)$	cumulative probability function of the standardized normal distribution
$\rho(\cdot)$	dispersion function
$\Psi(\epsilon)$	gradient of $\rho(\epsilon)$
T	transpose of matrix

Appendix A: Figure A1

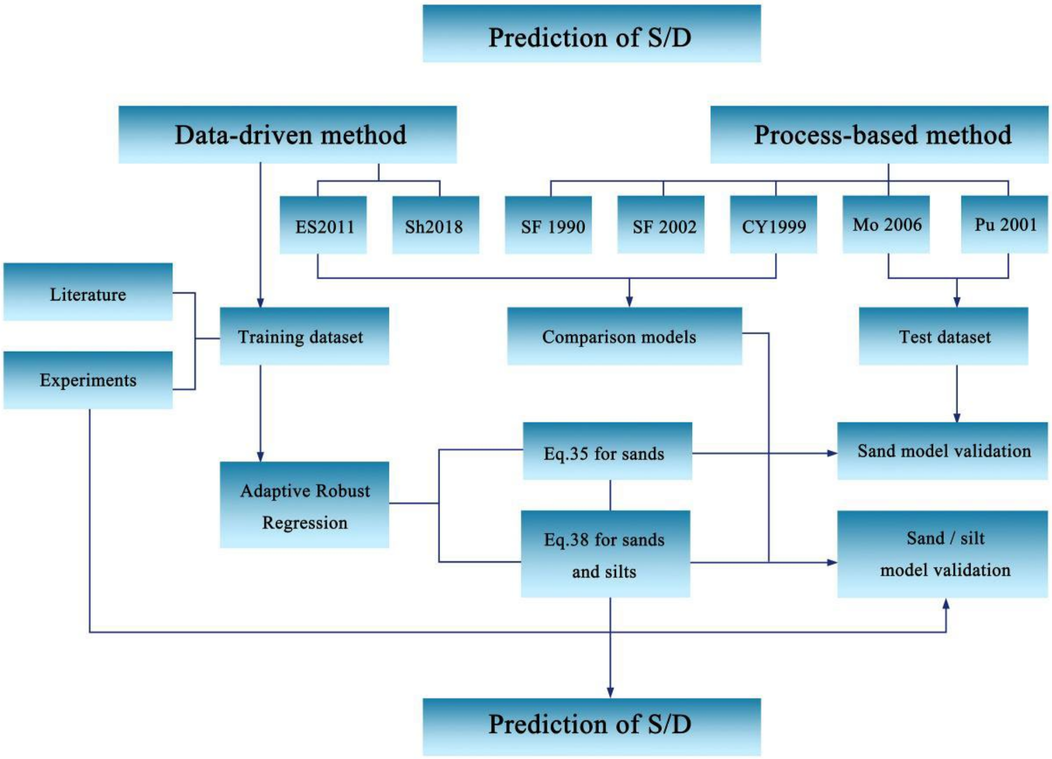

Figure A1. General flow of the present study.

Appendix B: Collected data from the literature

The dataset is attached as an Excel file.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716–723.

Aly, Emad-Eldin AA, and Aydin Öztürk. Hodges-Lehmann quantile-quantile plots. *Computational Statistics & Data Analysis*, 6, no. 2 (1988): 99-108.

Corvaro, S., Marini, F., Mancinelli, A., Lorenzoni, C., Brocchini, M. (2018). Hydro-and morpho-dynamics induced by a vertical slender pile under regular and random waves. *Journal of Waterway, Port, Coastal, and Ocean Engineering*, ASCE, 144(6), 04018018.

Callens, A., Wang, Y.-G., Fu, L., & Liquet, B. (2020). Robust estimation procedure for autoregressive models with heterogeneity. Environmental Modeling & Assessment, 1-11.

Çevik, E., & Yüksel, Y. (1999). Scour under submarine pipelines in waves in shoaling conditions. *Journal of Waterway*, *Port*, *Coastal*, *and Ocean Engineering*, ASCE, 125(1), 9-19.

Cheng, N. S., Wei, M., Xu, P., & Mao, R. (2020). Length scale for evaluating wave-induced pipeline scour. *Ocean Engineering*, *218*, 108153.

Dean, R. G., Dalrymple, R. A. (1991). Water wave mechanics for engineers and scientists (Vol. 2). World Scientific Publishing company.

Zhao, E. J., Dong, Y. K., Tang, Y. Z., Sun, J. K. (2021). Numerical investigation of hydrodynamics and local scour around submarine pipeline under joint effect of solitary wave and current. *Ocean Engineering*, 222, 108553.

Dong, Y., Wang, D., Randolph, M. F. (2017). Investigation of impact forces on pipeline by submarine landslide using material point method. *Ocean Engineering*, 146, 21-28.

Fan N, Jiang J, Dong Y, Guo L, Song L. (2022). Approach for evaluating instantaneous impact forces during submarine slide-pipeline interaction considering the inertial action. *Ocean Engineering*, 245, 110466. DOI: 10.1016/j.oceaneng.2021.110466.

Dogan, M., & Arisoy, Y. (2015). Scour regime effects on the time scale of wave scour below submerged pipes. *Ocean Engineering*, 104, 673-679.

Etemad-Shahidi, A., Yasa, R., & Kazeminezhad, M. H. (2011). Prediction of wave-induced scour depth under submarine pipelines using machine learning approach. *Applied Ocean Research*, *33*(1), 54-59.

Fredsøe, J., Sumer, B.M., Arnskov, M.M., 1992. Time scale for wave/current scour below pipelines. *Int. J. Offshore Polar Eng*. 2, 13–17.

Fu, L., Wang, Y.-G., & Cai, F. (2020). A working likelihood approach for robust regression. *Statistical Methods in Medical Research*, 29(12), 3641–3652.

Fuhrman, D. R., Baykal, C., Sumer, B. M., Jacobsen, N. G., & Fredsøe, J. (2014). Numerical simulation of wave-induced scour and backfilling processes beneath submarine pipelines. *Coastal Engineering*, *94*, 10-22.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: data mining, inference, and prediction. Springer Science & Business Media.

Huang, D., Cabral, R., & De la Torre, F. (2015). Robust regression. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(2), 363–375.

Huber, P. J. (1973). Robust regression: asymptotics, conjectures and monte carlo. *Annals of Statistics*, 1(5), 799–821.

Kazeminezhad, M. H., Etemad-Shahidi, A., & Yeganeh Bakhtiary, A. (2010). An alternative approach for investigation of the wave-induced scour around pipelines. *Journal of Hydroinformatics*, 12(1), 51-65.

Kim, S., Lee, H. J., & Yeon, J. H. (2011). Characteristics of parameters for local scour depth around submarine pipelines in waves. *Marine Georesources and Geotechnology*, *29*(2), 162-176.

Kızılöz, B., Çevik, E., & Aydoğan, B. (2015). Estimation of scour around submarine pipelines with Artificial Neural Network. *Applied Ocean Research*, *51*, 241-251.

Lucassen, R. J. (1984). Scour underneath submarine pipelines. TU Delft, Civil Engineering and Geosciences, Hydraulic Engineering

Maronna, R. A. (1976). Robust m-estimators of multivariate location and scatter. *The Annals of Statistics*, 51–67.

Mattioli, M., Mancinelli, A., Brocchini, M. (2013). Experimental investigation of the wave-induced flow around a surface-touching cylinder. *Journal of Fluids and Structures*, *37*, 62-87.

Mousavi, M. E., Bakhtiary, A. Y., Enshaei, N. (2009). The equivalent depth of wave-induced scour around offshore pipelines. *Journal of Offshore Mechanics and Arctic Engineering*, *131*(2).

Nielsen, P. (2009). Coastal and estuarine processes (Vol. 29). World Scientific Publishing Company.

Nielsen, P., Robert, S., Møller-Christiansen, B., & Oliva, P. (2001). Infiltration effects on sediment mobility under waves. *Coastal Engineering*, 42(2), 105-114.

Postacchini, M., & Brocchini, M. (2015). Scour depth under pipelines placed on weakly cohesive soils. *Applied Ocean Research*, *52*, 73-79.

Pu, Q., Li, K., & Gao, F. (2001). Scour of the seabed under a pipeline in oscillating flow. *China Ocean Engineering*, *15*(1), 129-138.

Sharafati, A., Yasa, R., & Azamathulla, H. M. (2018). Assessment of stochastic approaches in prediction of wave-induced pipeline scour depth. *Journal of Pipeline Systems Engineering and Practice*, 9(4), 04018024.

Soulsby R. Dynamics of marine sands: a manual for practical applications. London: Thomad Telford; 1997.

Sumer, B. M., & Fredsøe, J. (1990). Scour below pipelines in waves. *Journal of Waterway*, *Port*, *Coastal*, *and Ocean Engineering*, ASCE, 116(3), 307-323.

Sumer, B. M., & Fredsøe, J. (1996). *Scour around pipelines in combined waves and current* (No. CONF-9606279-). American Society of Mechanical Engineers, New York, NY (United States).

Sumer, B. M., & Fredsoe, J. (2002). *The mechanics of scour in the marine environment* (Vol. 17). World Scientific Publishing Company.

Wang, N., Wang, Y.-G., Hu, S., Hu, Z.-H., Xu, J., Tang, H., & Jin, G. (2018). Robust regression with data-dependent regularization parameters and autoregressive temporal correlations.

Environmental Modeling & Assessment, 23(6), 779–786.

Wang, Y.-G., Lin, X., Zhu, M., & Bai, Z. (2007). Robust estimation using the huber function with a data-dependent tuning constant. *Journal of Computational and Graphical Statistics*, 16(2), 468–481.

Xu, J., Li, G., Dong, P., Shi, J., 2010. Bedform evolution around a submarine pipeline and its effects on wave-induced forces under regular waves. *Ocean Engineering*, 37 (2–3), 304–313.

Xu J, Pu J, Li G. 2012. Field observations of seabed scour around a submarine pipeline on cohesive bed. In: Advances in computational environment science. Springer Verlag, editor. p. 23–33.

Zang, Z., Tang, G., Chen, Y., Cheng, L., & Zhang, J. (2019). Predictions of the equilibrium depth and time scale of local scour below a partially buried pipeline under oblique currents and waves. *Coastal Engineering*, *150*, 94-107.

Zhang, Q., Draper, S., Cheng, L., & An, H. (2017). Time scale of local scour around pipelines in current, waves, and combined waves and current. *Journal of Hydraulic Engineering*, ASCE, 143(4), 04016093.

Zhang, Y., Zhang, S., & Li, G. (2019). Seabed scour beneath an unburied pipeline under regular waves. *Marine Georesources & Geotechnology*, *37*(10), 1247-1256.

Zhou, C., Li, G., Dong, P., Shi, J., & Xu, J. (2011). An experimental study of seabed responses around a marine pipeline under wave and current conditions. *Ocean Engineering*, 38(1), 226-234.

Zhang, S., Nielsen, P., Perrochet, P., Jia, Y. 2021a. Multiscale superposition and decomposition of field-measured suspended sediment concentrations: implications for extending 1DV models to coastal oceans with advected fine sediments. *Journal of Geophysical Research: Oceans*, 126(3): e2020JC016474.

Zhang, S., Nielsen, P., Perrochet, P., Xu, B., Jia, Y., Wen, M. 2021b. Derivation of settling velocity, eddy diffusivity and pick-up rate from field-measured suspended sediment concentration profiles in the horizontally uniform but vertically unsteady scenario. *Applied Ocean Research*, *107*, 102485.

Zhang, S., Wu, J., Jia, Y., Wang, Y. G., Zhang, Y., Duan, Q. 2021c. A temporal LASSO regression model for the emergency forecasting of the suspended sediment concentrations in coastal oceans: Accuracy and interpretability. *Engineering Applications of Artificial Intelligence*, *100*, 104206.

Zhang, S., Wu, J., Wang, Y., Jeng, D., Li, G. 2022. A physics-informed statistical learning framework for forecasting local suspended sediment concentrations in marine environment. *Water Research*. 118518. https://doi.org/10.1016/j.watres.2022.118518.

Zhang Y. Influence of wave-induced seepage on seabed scour beneath submarine pipeline. Doctoral Dissertation. Ocean University of China. 2019.