Desalination and Water Treatment
www.deswater.com

doi: 10.5004/dwt.2023.30043

314 (2023) 352–358
December

Multiple linear regression analysis of vertical distribution of nearshore suspended sediment

Wenjin Zhu[a], Wei Yu[a], Xiaotian Dong[a,b,*], Zigui Jin[a], Shunqiu Hu[a]

[a]*School of Civil and Ocean Engineering, Jiangsu Ocean University, Lianyungang 222000, China, email: dongxt@jou.edu.cn*
[b]*Observation and Research Station of East China Coastal Zone, Ministry of Natural Resources, Nanjing 210005, China*

Received 11 December 2023; Accepted 30 September 2023

ABSTRACT

Many scholars use the diffusion theory and the Rouse equation to describe the vertical distribution of sediment in the nearshore waters of estuaries. This article analyzes several assumptions for the establishment of the equation to test how well it can be applied in the marine nearshore zone. Six independent variables were selected based on the traditional Rouse distribution, the linear form of the Rouse equation, and the characteristics of water transported sediment. Then, they were analyzed for their impacts on sediment concentration through multiple linear regression. According to the results, the concentration value at the reference point had the most significant impact on the calculation of the sediment concentration. In comparison, the sediment concentration calculated based on the relative water depth only had a correlation of 0.3 with the measured values. Among the six independent variables, flow velocity's sensitivity (especially that of the variable ln$u$) was much lower than the sensitivity of the reference point concentration. Thus, flow velocity was decided as a non-sensitive factor. After excluding it from the independent variables, the equation contained five independent variables. Then we used the equation to calculate the sediment concentration at each point. The correlation coefficients between the results and the measured values reached around 0.8. This proves that the method adopted by this paper can reflect the vertical concentration distribution of fine particle sediment in nearshore waters under complex dynamic conditions.

Keywords: Rouse equation; Multiple linear regression; Vertical distribution of suspended sediment

1. Introduction

Suspended sediment concentration is an important parameter for describing sediment dynamics in coastal and marine environments, such as sediment transport and sedimentation [1–3]. The theories of diffusion [4], mixing, energy dissipation, similarity and stochastic models have been applied to study suspended-sediment concentration distributions. Starting from different perspectives, these theories have produced a wide range of results [5]. We have analyzed these theories in a comprehensive way and found that the results of all the theories shared a similar pattern: they were close to or in the form of diffusion equations. The only differences were the diffusion coefficients. Therefore, scholars directly use the diffusion theory in their study of suspended fine sediments in estuaries and coastal areas [6–9]. Using the Rouse formula, Li et al. [10] simulated the concentration of suspended sediments at each water layer. Under the condition that the vertical mixing of sediment particles was sufficient, the results were relatively accurate. Lin et al. [11] used the Rouse formula to fit and determine fine suspended sediment's settling velocity, which was correlated with the vertical distribution curve of suspended sediment concentration. According to the results, the equation of the settling velocity in estuaries is proportional to the exponential power of suspended sediment concentration, salinity, and frictional velocity. Zhao et al. [12] found that the suspended

* Corresponding author.

sediment concentrations follow the Rouse distribution only at the break of neap tide. Most of the time, that is, during neap tide and spring tide, they deviate from it.

There are strict limitations when the Rouse equation is to be applied to the unsteady flow of the estuaries that are linked to the nearshore ocean. Before the flow velocity's vertical distribution is determined clearly, the Rouse equation can only be used when the water flows in a stable way without density stratification or significant wind-generated waves. The dynamic field in nearshore waters is unstable and involves multiple factors. Therefore, the vertical concentrations of sediments in nearshore waters are too complex to be described by any available methods. This article proposes a new method that combines traditional Rouse distribution with multiple linear regression. The new method has introduced multiple independent variables for linear regression in order to analyze the characteristics of the vertical distribution of nearshore suspended sediment.

2. Method

2.1. Applicability analysis of Rouse distribution in nearshore areas

According to the diffusion theory, sediment suspension occurs as a result of turbulent diffusion and gravity. When sediment particles reach balance in the water flow, their physical state satisfies the mass conservation equation or the continuous equation. The Rouse equation describing the vertical distribution of suspended sediment (the coordinate origin was placed at the bottom of the water and the water surface $z = h$) is as follows:

$$\frac{c(z)}{c_a} = \left[\frac{a(h-z)}{z(h-a)}\right]^{\omega_s/\kappa u^*} \tag{1}$$

where c_a is the concentration of sediment at the reference point. The Rouse equation is established based on the following assumptions: (1) The time-averaged velocity of vertical flow is 0; (2) The assumptions of the mixing length model, which is a method attempting to describe momentum transfer by turbulence (i.e., the logarithmic law of the velocity distribution); (3) Sediment diffusion coefficient ε_z equals the water flow's momentum exchange coefficient ε; (4) The sedimentation rate is a constant that does not vary with the water depth. However, most nearshore waters do not conform to these assumptions.

First, sedimentation rate ω_s is not a constant for nearshore waters because the fine sediment particles have different settling velocities in fresh water and saline water, respectively. According to Qin and Wan [13], when the overall salinity is small, the average settling velocity of flocs will increase with salinity. Once the salinity has surpassed a certain threshold, any further increase in salinity will have little impact on the velocity. Besides, as the sediment concentration at the bottom of the nearshore water is high, the settling of particles is reduced due to return flow, particle collisions, and increased mixture viscosity. This process is referred to as "hindered settling". That's why the sedimentation rate is no longer a constant.

Secondly, as the horizontal time-averaged flow velocity in the nearshore waters of estuaries is influenced by both the accelerating flow and the decelerating flow, the velocity distribution does not follow the logarithmic law (Fig. 1). The sediment diffusion coefficient ε_z along the water depth no longer follows parabolic distribution (Fig. 2). Stratification in water makes the distribution of sediment diffusion coefficient even more complex. Currently, only a few studies focus on the vertical mixing and stratification of estuaries. Most of previous studies believe that the vertical distribution of sediment's diffusion coefficient is complex, especially when the water is highly stratified. Instead of following a simple parabolic distribution, the distribution of the sediment's vertical diffusion coefficient is also influenced by the vertical gradient of water density and flow velocity.

Therefore, before the vertical distribution of flow velocity is clearly determined, the Rouse equation can only be

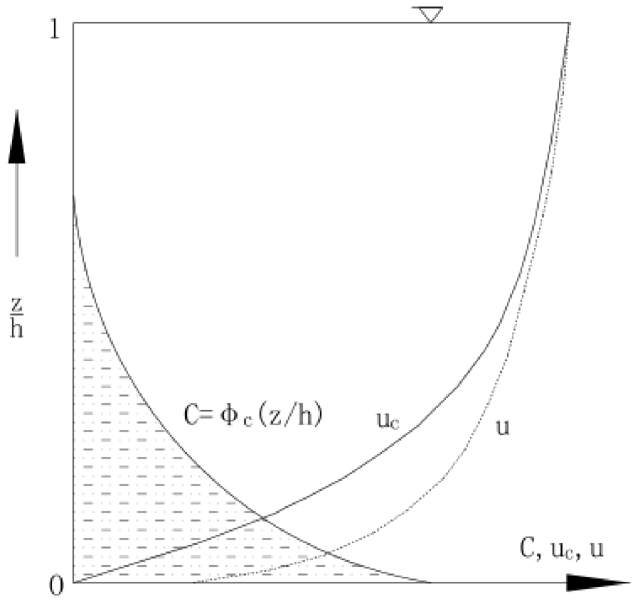

Fig. 1. Vertical distribution of concentration and velocity (Uc is the velocity of sediment-laden flow).

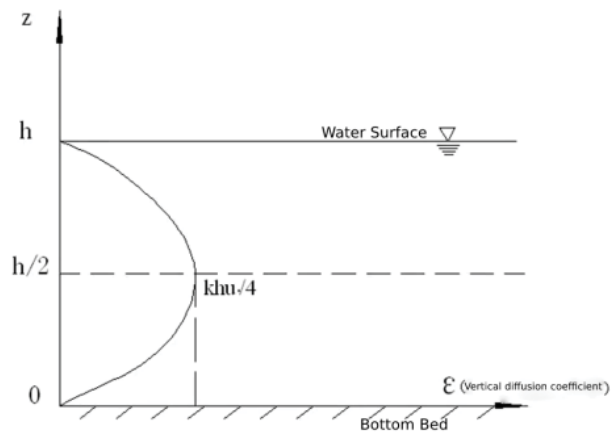

Fig. 2. Distribution map of vertical diffusion coefficient at steady flow.

used with limitations in non-stratified waters with small wind waves and slow flow. Otherwise, significant errors will occur.

2.2. Linear model of Rouse distribution

The logarithms on both sides of Eq. (1) are taken, and the equations are rewritten as linear ones, so that the vertical distribution curve of suspended sediment concentration can be more easily calculated:

$$\ln c = a_0 + a_1 \ln\frac{z}{h} + a_3 \ln c_a \quad (2)$$

where the undetermined coefficient is obtained through regression of the measured data. The undetermined parameters are used to improve the correlation between the sediment concentration at each layer and the sediment concentration at the reference point. Eq. (2) reflects the correlation between the sediment concentration at each point of the vertical line and the variables – water depth and the concentration at the reference point. Using vertical sediment concentrations at six measured points in Hai'an Bay of Qiongzhou Strait in May 2008, we analyzed the correlation between the measured data with the independent variables ln(z/h) and lnc_a (Fig. 3). The results are shown in Fig. 4. Although we can use these two independent variables to calculate the sediment concentration, the results' correlation coefficient with measured value is only between 0.5 and 0.6. This indicates that for fine-grained sediment in Hai'an Bay, under complex dynamics, neither the Rouse Eq. (1) nor the linear Eq. (2) can accurately reflect the vertical distribution of sediment concentration.

The samples used in this study are from the year of 2008. Water level data: SC1, SC2, SC3, SC4, and SC5 were observed continuously for 15 d during May 13–May 28. Flow velocity data: the neap tide lasted from 10:00 on May 14 to 14:00 on May 16 (52 h in total). The spring tide lasted from 17:00

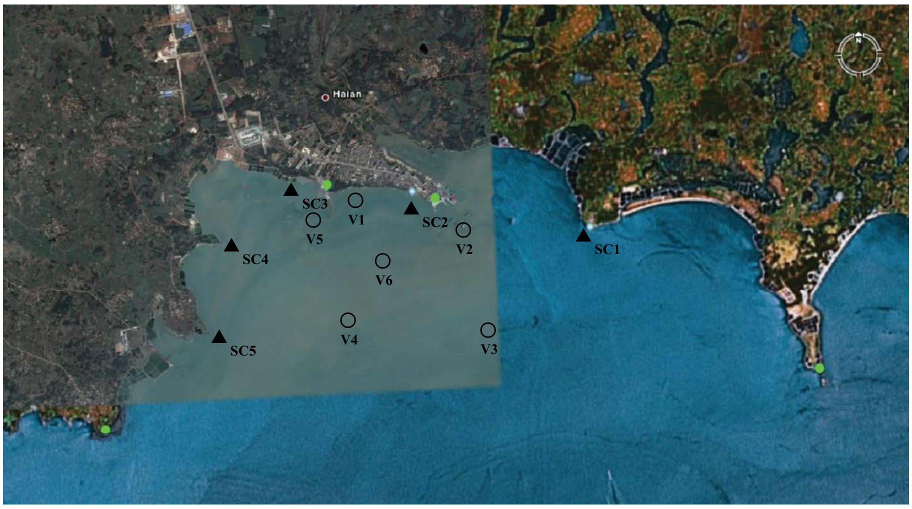

Fig. 3. General situation of Hai'an Bay Sea area and hydrometric station location.

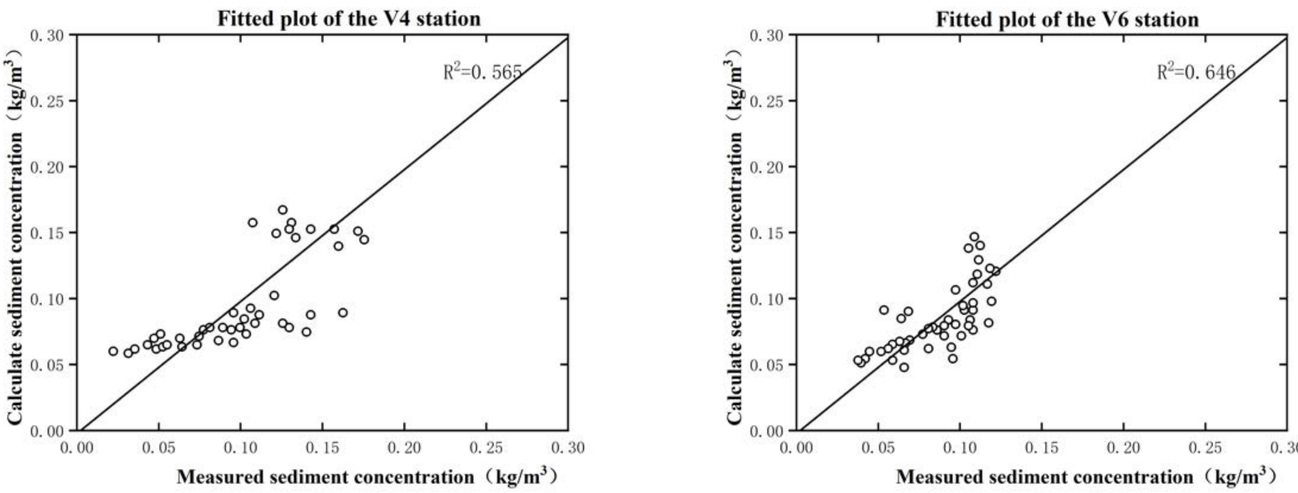

Fig. 4. Correlation between calculated and measured values of vertical sediment concentration when ln(z/h) and lnc_a are considered.

on May 20 to 22:30 on May 22 (53 h in total). Sediment data: data was measured at the same time with the observation of flow velocity, and sampling was conducted for every 2 h.

3. Improvement

3.1. Multiple linear regression method

Multiple linear regression analysis is a well-known method for multivariate statistical analysis. It can predict one or more corresponding variables (dependent variables) from a collection of independent variables. Multiple linear regression analysis can provide the predicted values and trends of the dependent variables, allowing managers a theoretical foundation on which to base their decisions. If the dependent variable Y satisfies a linear relationship with multiple independent variables $X_1, X_2, \ldots, X_k$, the following multiple linear regression model can be created:

$$\begin{cases} Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k \\ \varepsilon - N\left(0, \sigma^2\right) \end{cases} \quad (3)$$

where Y is an observable random variable; ε is an unobservable random error, and $\beta_0, \beta_1, \ldots, \beta_k$ is an unknown number that is independent of $X_1, X_2, \ldots, X_k$.

3.2. Regression analysis of vertical sediment concentration

According to the turbulent diffusion theory, the changing pattern of vertical sediment concentration is in line with that of water flow's turbulence intensity. Moreover, there is a correlation between the concentration of suspended sediment and water flow velocity and velocity gradient. The measured value of the flow velocity at each layer vertically in the nearshore waters is influenced by various dynamic factors (mainly by the waves and tides). The measured flow velocity can not only reflect the strength of the tides, but also reflect the stage of the water flow (rising tide, falling tide, or slack tide). The measured flow velocity can be used to indirectly reflect the influence of waves and currents and to further correct the distribution of vertical sediment concentration.

Under the influence of sedimentation at the upper layers and turbulent diffusion at the lower layers, there is a good correlation between sediment concentrations at adjacent layers. Another good correlation exists among various layers during the neap tide and spring tide in the Hai'an Bay. Table 1 shows the correlation coefficients between the bottom layer, 0.6th layer, and average sediment concentration at the water surface.

According to the analysis, the relative water depth (z/h), the sediment concentration at the reference point c_a, and

Table 1
Correlation of sediment concentration between layers

Stations	Relative depth position	Spring tide				Neap tide			
		0.0	0.6	1.0	Average	0.0	0.6	1.0	Average
Point 6	0.0	1	0.496	0.124	0.734	1	0.858	0.605	0.809
	0.6	–	1	0.506	0.851	–	1	0.788	0.908
	1.0	–	–	1	0.518	–	–	1	0.917
	Average	–	–	–	1	–	–	–	1
Point 5	0.0	1	0.548	0.196	0.550	1	0.732	0.103	0.506
	0.6	–	1	0.699	0.915	–	1	0.394	0.718
	1.0	–	–	1	0.767	–	–	1	0.663
	Average	–	–	–	1	–	–	–	1
Point 4	0.0	1	0.055	0.013	0.427	1	0.715	0.624	0.821
	0.6	–	1	0.442	0.545	–	1	0.927	0.949
	1.0	–	–	1	0.525	–	–	1	0.934
	Average	–	–	–	1	–	–	–	1
Point 3	0.0	1	0.521	0.378	0.708	1	0.944	0.861	0.961
	0.6	–	1	0.719	0.896	–	1	0.958	0.994
	1.0	–	–	1	0.848	–	–	1	0.963
	Average	–	–	–	1	–	–	–	1
Point 2	0.0	1	0.799	0.604	0.876	1	0.394	0.170	0.603
	0.6	–	1	0.816	0.952	–	1	0.562	0.763
	1.0	–	–	1	0.856	–	–	1	0.665
	Average	–	–	–	1	–	–	–	1
Point 1	0.0	1	0.676	0.426	0.862	1	0.871	0.705	0.913
	0.6	–	1	0.542	0.882	–	1	0.855	0.972
	1.0	–	–	1	0.717	–	–	1	0.908
	Average	–	–	–	1	–	–	–	1

the flow velocity u are three crucial values to calculate the vertical distribution of sediment's concentration. This article considers sediment concentration as a random variable, and treats relative water depth (z/h), reference point sediment concentration c_a, and flow velocity u as independent variables. To determine the forms of the three independent variables in the formula, we introduced the linear form of the Rouse formula and the six independent variables, which were (z/h), $\ln(z/h)$, c_a, $\ln c_a$, u, and $\ln u$, to construct formula 4 (Table 2 for a list of variables).

$$\ln c = a_0 + a_1 \ln \frac{z}{h} + a_2 c_a + a_3 \ln c_a + a_4 u + a_5 \ln u + a_6 \frac{z}{h} \tag{4}$$

Table 2
List of variables

Dependent variable	$\ln c$
Constant A_0	—
Independent variable a_1	$\ln(z/h)$
Independent variable a_2	c_a
Independent variable a_3	$\ln c_a$
Independent variable a_4	u
Independent variable a_5	$\ln u$
Independent variable a_6	z/h

In the regression analysis, we added 4 more variables (C_a, u, $\ln u$, (z/h)) to the original 2 variables ($\ln(z/h)$ and $\ln c_a$) in the Rouse equation. Table 3 shows that the concentration at the reference point has the most significant effect on the sediment concentration. The method that only adopted the relative water depth and the actual measured value to calculate the sediment concentration generate a small correlation of about 0.3.

Table 3 shows the results of correlations after combining independent variable 1 with independent variable 2, independent variable 3, independent variable 4, independent variable 5, and independent variable 6, respectively.

As shown in Table 3, the combination of "independent variable 1 + independent variable 2" generate the highest correlation coefficient than others. In other words, the combination of "independent variable 1 + independent variable 2" has a crucial impact on the vertical distribution of sediment in the Hai'an Sea area. Therefore, it is regarded as the basic combination to be further integrated with other independent variables. As a result, a total of 16 correlation coefficients were calculated under different combinations. After the basic combination was further integrated with "independent variable 4" and "independent variable 5", the new group (Group 14) has a slightly lower effect than when the basic combination was integrated with "independent variable 6" (Group 13). For Hai'an Bay, factors of relative water depth, reference point concentration, and flow velocity all have a certain impact on the vertical distribution of fine-grained sediment concentration. According to Tables 1 and

Table 3
Correlation between vertical sediment concentration calculated under the combination of variables and measured values

Order number	Combination of independent variables	V_1	V_2	V_3	V_4	V_5	V_6
1	Independent variable 1	0.225	0.180	0.370	0.555	0.360	0.388
2	Independent variable 1 + Independent variable 2	0.786	0.782	0.713	0.562	0.783	0.635
3	Independent variable 1 + Independent variable 3	0.756	0.800	0.384	0.565	0.815	0.389
4	Independent variable 1 + Independent variable 4	0.394	0.227	0.424	0.572	0.417	0.646
5	Independent variable 1 + Independent variable 5	0.380	0.235	0.383	0.560	0.408	0.389
6	Independent variable 1 + Independent variable 6	0.240	0.197	0.738	0.741	0.445	0.464
7	Independent variable 1 + Independent variable 2 + Independent variable 3	0.787	0.810	0.740	0.567	0.816	0.646
8	Independent variable 1 + Independent variable 2 + Independent variable 4	0.808	0.785	0.713	0.578	0.824	0.647
9	Independent variable 1 + Independent variable 2 + Independent variable 5	0.801	0.790	0.767	0.567	0.821	0.647
10	Independent variable 1 + Independent variable 2 + Independent variable 6	0.801	0.808	0.713	0.748	0.868	0.711
11	Independent variable 1 + Independent variable 2 + Independent variable 3 + Independent variable 4	0.810	0.827	0.740	0.580	0.831	0.657
12	Independent variable 1 + Independent variable 2 + Independent variable 3 + Independent variable 5	0.802	0.846	0.740	0.570	0.828	0.656
13	Independent variable 1 + Independent variable 2 + Independent variable 3 + Independent variable 6	0.802	0.838	0.794	0.763	0.901	0.722
14	Independent variable 1 + Independent variable 2 + Independent variable 3 + Independent variable 4 + Independent variable 5	0.815	0.865	0.740	0.619	0.831	0.657
15	Independent variable 1 + Independent variable 2 + Independent variable 3 + Independent variable 4 + Independent variable 6	0.832	0.849	0.797	0.769	0.911	0.725
16	Independent variable 1 + Independent variable 2 + Independent variable 3 + Independent variable 4 + Independent variable 5 + Independent variable 6	0.836	0.889	0.799	0.798	0.913	0.726

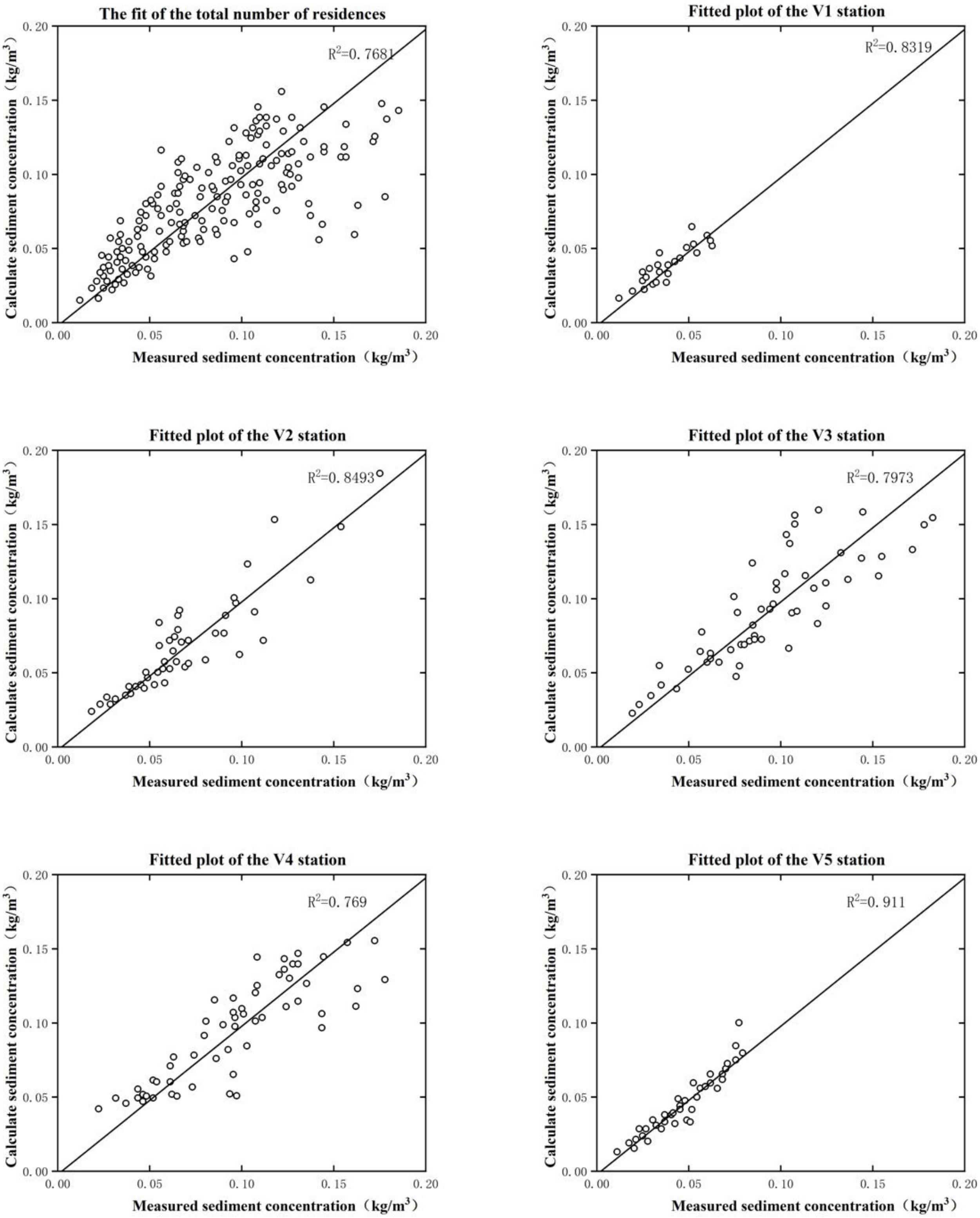

Fig. 5. Correlation between calculated and measured sediment vertical concentration.

3, concentrations at reference points have the highest influence, followed by the relative water depth and flow velocity. After comparing the results in the Groups 14, 15, and 16, we found that the influence of "independent variable 5" is relatively small. Therefore, we removed this variable and only adopted the combination of "independent variable 1 + independent variable 2 + independent variable 3 + independent variable 4 + independent variable 6" to calculate the vertical distribution of sediment in Hai'an Bay (Fig. 5). The combination of the five independent variables above can generate correlation coefficients as high as 0.8 for almost all stations, and the result from the V_5 station can even reach 0.91. Therefore, using the Rouse equation and multiple linear regression Eq. (5) to analyze the vertical distribution of fine sediment concentration in Hai'an Bay can achieve satisfactory accuracy.

$$\ln c = a_0 + a_1 \ln \frac{z}{h} + a_2 c_a + a_3 \ln c_a + a_4 u + a_6 \frac{z}{h} \tag{5}$$

4. Conclusion

Due to complex dynamic conditions, the use of the Rouse equation has to follow strict limitations and few basic theories can be relied on to determine the vertical distribution of fine sediment concentration in nearshore waters of estuaries. This article used the theory of sediment dynamics and multiple linear regression to solve this problem. Based on the linear form of the Rouse equation and considering the sediment transport characteristics in nearshore waters, it treated sediment concentration as a "random variable" and selected $\ln(z/h)$, $\ln c_a$, c_a, u, $\ln u$, (z/h) to be independent variables. Adopting multiple linear regression, it analyzed the impacts of the six independent variables on sediment concentration (the random variable) and used the significance test to eliminate the factor ($\ln u$) that has the smallest influence. As indicated by the correlation coefficient between the predicted sediment concentration and the measured sediment concentration, the independent variables are able to reflect the vertical distribution of the nearshore fine sediment's concentration under complicated dynamic conditions.

Acknowledgments

This paper was supported by Observation and Research Station of East China Coastal Zone, Ministry of Natural Resources (ORSECCZ2022204).

References

[1] R.J. Li, Y.X. Yan, Z.Y. Song, Calculation of suspended sediment transport in Taiping waterway, J. Sediment Res., 4 (2003) 46–51.
[2] F. Xing, Y.P. Wang, J.H. Gao, X.Q. Zou, Seasonal distributions of the concentrations of suspended sediment along JiangSu coastal sea, J. Oceanol. Limnol., 41 (2010) 459–468.
[3] S. Dey, S.Z. Ali, E. Padhi, Advances in analytical modeling of suspended sediment transport, J. Hydro-Environ. Res., 20 (2018) 110–126.
[4] G.Q. Wang, X.D. Fu, Mechanisms of particle vertical diffusion in sediment-laden flows, Chin. Sci. Bull., 49 (2004) 1086–1090.
[5] J.R. Ni, Y.J. Hui, Various theories and their relationships on the vertical distribution of suspended sediment concentration, Water Resour. Transp. Sci. Res., 1 (1988) 83–97.
[6] R.J. Li, F. Luo, W.J. Zhu, The suspended sediment transport equation and its near-bed sediment flux, Sci. China (Series E), 52 (2009) 387–391.
[7] J.P.-Y. Maa, J.-I. Kwon, Using ADV for cohesive sediment settling velocity measurements, Estuarine Coastal Shelf Sci., 73 (2007) 351–354.
[8] C.J. Wang, River Dynamics, Communications Press, Beijing, China, 2001, pp. 59–67.
[9] F.M. Henderson, Open Channel Flow, MocMillan Co., New York, 1966, pp. 85–86.
[10] Z.H. Li, S. Gao, H.T. Shen, Suspended sediment concentration profiles and grain size distribution patterns over the Dafeng Tidalflat, Probl. Rechn. NanosovSilt Study, 1 (2006) 62–70.
[11] W.B. Lin, D.Y. Kong, F. Luo, Study on determination method for settling velocity of fine sediment in Oujiang River Estuary, J. Hydroelectric Eng., 32 (2013) 114–119.
[12] J.W. Zhao, Z.H. Li, S. Xu, E.F. Zhang, H.Q. Cheng, Vertical profile of suspended sediment concentration in the upper reach of north channel in the Changjiang estuary during the dry season, Resour. Environ. Yangtze Basin, 28 (2019) 2207–2218.
[13] N. Qin, Z.H. Wan, Mechanics of Sediment Transport, Science Press, Beijing, China, 1983, pp. 72–78.