



Evaluation of Lateral Displacement Induced by Jet Grouting using Improved Random Forest

Pierre Guy Atangana Njock · Ning Zhang ·
Annan Zhou · Shui-Long Shen

Received: 24 November 2021 / Accepted: 26 July 2022 / Published online: 29 August 2022
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2022

Abstract This study presents an improved random forest (IRF) model to evaluate the ground displacement caused by jet grouting. The proposed IRF model integrates a new hybrid particle swarm optimization-simulated annealing (PSO-SA) algorithm into random forest. The performance of the PSO-SA optimizer is investigated using a set of 8 benchmark functions; whereas, three benchmark models are used to test the IRF model. The results show that PSO-SA has much better searching and convergence abilities than its counterparts (PSO and SA). When the IRF was applied to predict ground displacement, the IRF model outperforms the benchmarks models. Moreover, the out-of-bag variable importance analysis shows that the ground lateral displacement resulting from jet grouting operations can be controlled

through two operating parameters, namely, the retract rate of monitor and injected fluid pressure. The high accuracy and flexibility of the proposed model make it suitable for other engineering applications regardless of their complexity.

Keywords Random Forest · Jet Grouting · Artificial Intelligence · Ground Displacement · Optimization

1 Introduction

The jet grouting is a ground treatment technique that enjoys an overwhelming popularity in geotechnical engineering owing to its efficiency, versatility, and cost-effectiveness. This method consists of constructing *in-situ* soil-cement columns through the consecutive mechanisms of drilling, jetting and grouting (Kimpritis et al. 2018; Shen et al. 2020; Stark et al. 2009, 2012). The jetting grouting involves the interaction between high pressure jetting fluid and soil, which generates considerable disturbances in the surrounding medium and enhances the ground stiffness by forming jet grouted column. However, the jet grouted column installation process tends to cause substantial ground movements that can jeopardize the good functioning of surrounding structures as a result of additional earth pressures imposed on them (Wang et al., 2018). For instance, lateral displacement is an issue often encountered in tunnelling or excavation bottom plugs (Eramo et al. 2012; Ochmański et al.

P. G. Atangana Njock · N. Zhang · S.-L. Shen
Department of Civil and Environmental Engineering,
College of Engineering, Shantou University,
Shantou 515063, Guangdong, China
e-mail: zhangning@stu.edu.cn

A. Zhou
Discipline of Civil and Infrastructure, School
of Engineering, Royal Melbourne Institute of Technology
(RMIT), Victoria 3001, Australia
e-mail: annan.zhou@rmit.edu.au

S.-L. Shen (✉)
MOE Key Laboratory of Intelligent Manufacturing
Technology, College of Engineering, Shantou University,
Shantou 515063, Guangdong, China
e-mail: shensl@stu.edu.cn

2015b). Quantifying and or predicting these effects have remained an important hurdle to the efficient implementation of jet grouting. To date, only few models have been developed to estimate the ground displacement caused by jet grouting (Chai et al., 2009, 2005; Wu et al., 2016). Although these models are fairly efficient, they suffer from some limitations. Many existing models (see Table 1) do not integrate the eventuality of system disruption; and thus will be hard to reflect some fundamental soil behaviours arising from external and or complex factors. For instance, the excess pore pressure fluctuation which is a catalyst of upward and lateral ground movements, is strongly correlated to the injection pressure and installation sequence. Besides, it has been speculated that quick gelling binders may play a determinant role in the fluctuation of excess pore pressure. Neither the existing field investigations (Poh and Wong, 2001; Wang et al., 2013) nor theoretical models (Wang et al., 2018) could incorporate these considerations in their models. In addition, their complexity is still a hindrance to their effective implementation by practitioners, which may lack relevant mathematical background.

Due to their prominent ability to deal with non-linear problems as well as their advantageous practicality, deep learning models are increasingly adopted in the field of geotechnical engineering (Ochmański et al. 2015a; Wei et al., 2021). Among these models, the random forest method is known as one of the most powerful machine learning tools for regression analyses (Liu et al., 2020). To date, there is no records of studies that capitalized the random forest method to predict the ground displacement induced by jet grouting. However, the few applications of random forest to

other engineering problems (Coulston et al., 2016; Couronné et al., 2018; Palmer et al., 2007; Segal, 2004) has allowed delineating some limitations. In particular, the topology design of the random forest method has remained problematic, and it has usually required the use of optimization techniques. Then again, because these optimizers are sometimes prone to suboptimal solutions due to limited abilities, hybridization strategies are gradually espoused as alternatives.

The basic idea of hybridization of two methods lies in their reciprocal improvement for achieving (near-) optimal solutions. Recent efforts have supplied two categories of models. First, hybrid models achieved via locally embedding SA into the main operative searching process of PSO or vice versa (Dong et al., 2010; Meng et al., 2016; Wang and Li, 2004). Second, models stemming from the all-inclusive hybridization of PSO and SA (Javidrad and Nazari, 2017; Shieh et al., 2011; Sudibyo et al., 2015). Most of these models share operational and conceptual resemblances as well as performance level. However, improvements are still in demand regarding a number of hitches. For instance, the proper characterization of the influence of ‘individuals transferred from PSO’ on the SA’s population has remained difficult to achieved for many models (Meng et al., 2016; Shieh et al., 2011). Moreover, many hybrid models have been merely tested on low-dimensional spaces (*i.e.*, less than 10-D) and/or on unimodal functions (Dong et al., 2010; Javidrad and Nazari, 2017; Sudibyo et al., 2015; Wang and Li, 2004). In other words, the benchmarks testing of many existing PSO-SA algorithms has arguably remained inconclusive. Therefore, there is a room for development of new hybrid PSO-SA model.

Table 1 Equations for calculating lateral displacement induced by the construction of jet grouted column

Solutions	Idealized displacement medium		References
	Elastic zone ($r > R_p$)	Plastic zone ($R_u \leq r \leq R_p$)	
Vesic	$\delta = \frac{R_p^2}{r^3} \delta_p$	$\delta = \frac{2R_p + \delta_p}{2r + \delta_p R_p / r} \delta_p$	Chai et al. (2005)
Verruijt	Indirect solution (see references)	$\begin{cases} u_{xA} = u_p \frac{2R_p + u_p}{2r_A + u_p R_p / r_A} \frac{x}{r_A} \\ u_{yA} = u_p \frac{2R_p + u_p}{2r_A + u_p R_p / r_A} \frac{d_c + y}{r_A} \end{cases}$	Wang et al. (2018)

The objective of this article is to propose a robust machine learning method for predicting complex geotechnical problems in general and the ground displacement caused by hydraulic jetting in particular. In Sect. 2, we proposed an improved random forest (IRF) model that combines the advantages of random forest and robustness of a novel hybrid particle swarm optimization-simulated annealing (PSO-SA) algorithm. In Sect. 3, the proposed improved random forest (IRF) model was applied to the estimation of the lateral ground displacement induced by horizontal jet grouting during a tunnel construction in Shanghai. The prediction performance of the proposed IRF model is presented and discussed and how the soil and jetting parameters affect lateral ground displacement is also analyzed by using the proposed IRF model. Some key findings in this research are summarized and presented as conclusions.

2 Improved Random Forest (IRF) Model

This section thoroughly describes the different aspects of hybridization strategies undertaken to construct the proposed improved random forest model (IRF). Given that, the nature of the problem investigated (prediction of jet grouting-induced displacement), a random forest model was considered in this study. The basic idea behind the proposed IRF is to maximize the predicting performance of the random forest, while ensuring the efficiency of the training process. As such, to make full use the predicting power of the random forest technique, the model has been methodically defined in manner to: (i) minimize the variance of the predictions, (ii) maximize the convergence and accuracy of the model. To be more specific, the proposed IRF was obtained by crossbreeding the random forest model with a hybrid PSO-SA. The objective of the latter was mainly to define the optimum random forest hyper-parameters.

2.1 Random Forest Model

The random forest is a bagging technique that operates by building an assembly of decision trees at training stage and outputting the average prediction of the individual trees. That is, considering an assembly of decision trees $\{r(x, \varphi_k), t = 1, 2, 3, \dots, N\}$, where x is an input variable, φ_k is an independent random forest

variable, and N is the total number of decision trees, the random forest algorithm returns the average score of k decision trees $r(x, \varphi_k)$ as the final prediction result (Coulston et al., 2016). Although this algorithm works in a “black box” fashion, satisfactory prediction results can be obtained if the model is suitably designed. In our study, the random forest model was constructed following 3 main steps which can be discretized as it follows:

Step 1: Bootstrapping was applied for reducing the variance of the random forest predictions. The basic idea was to produce k separate training sets (bootstrap samples) $\varphi_1, \varphi_2, \dots, \varphi_k$ by repeatedly sampling data with replacement from the original training. Subsequently, k decision trees $r(x, \varphi_1), r(x, \varphi_2), \dots, r(x, \varphi_k)$ were generated, and then, methodically split into subtrees. At this step, the number of features to be searched at each split point was itemized as a parameter to the system.

Step 2: The out-of-bag (OOB) estimate of performance was carried out to check the performance of each model. Indeed, each bootstrap sample (previously generated) consisted of approximately 2/3 of the training data, while the remaining portion of samples (around 1/3 of the training data) were not included. These samples are generally called out-of-bag (OOB) samples. Interestingly, by averaging the performance of each model on its OOB, it is possible to obtain an estimated accuracy of the random forest model. This performance was determined by calculating the OOB error.

Step 3: The importance of variables (the most or the least relevant to the problem) was identified. In fact, as the ‘random forest’ is built up, it is necessary to evaluate the error function residual value for a variable at each split point. The greater the value of residual at the moment of choosing the variable, the greater the importance. In this case, adopting the mean square error (MSE) as error function the variable importance measure (VIM) was calculated using Eq. 1 (Liu et al., 2020).

$$VIM(MSE) = \left[\frac{1}{\lambda} \sum_{j=1}^{\lambda} (MSE_{oob} - MSE_{oobi}) \right] / \sigma_{SD}, (1 \leq i \leq K) \quad (1)$$

where λ is the number of decision trees; MSE_{oob} denotes the mean square error value of the OOB; MSE_{oobi} is the updated MSE (i.e., after randomly replacing the target variables x_i); σ_{SD} represents the standard deviation standardization; and K is the

number of features in a specific tree. The aforementioned steps allowed identifying three principal hyper-parameters that may greatly affect the performance of random forest models, including, the number of trees to be used in the forest ('numTrees'); the number of features to consider when looking for the best split ('maxFeatures'); and the minimum number of samples required in a leaf node ('minLeaf'). The suitable selection of these hyper-parameter values is pivotal as it governs the model performance. This state of play thus substantiates the use of a robust parameter tuning strategy such as PSO-SA.

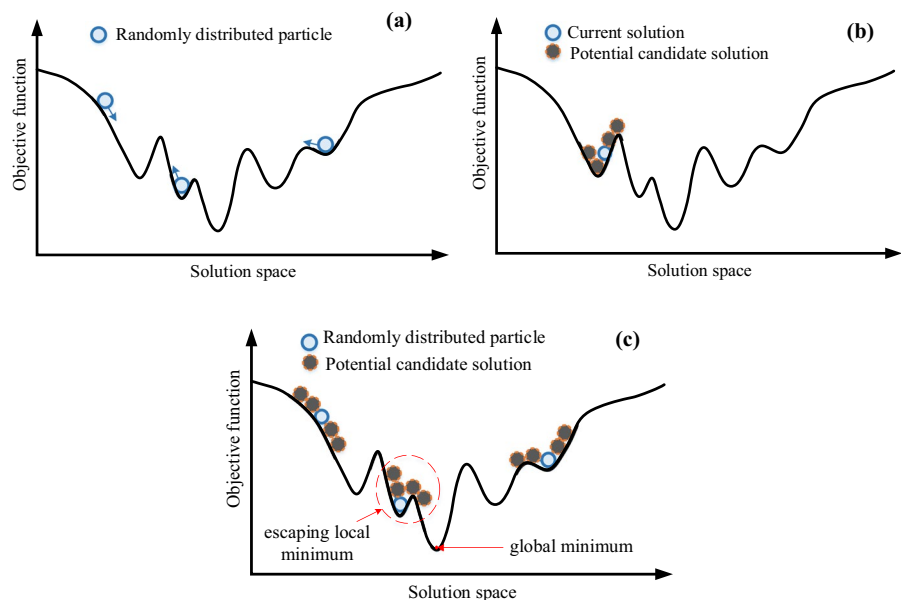
2.2 Proposed Hybrid PSO-SA Optimization Method

It is widely accepted that the PSO possesses a relatively effective global search mechanism. On unimodal functions, it can converge rapidly to the global optimum and result in high accuracy. However, in the case of multimodal functions, the 'pbests' are very likely to get stuck into a local optimum, especially if the latter is far from the global optimum (Gong et al. 2016). On the contrary, SA owns much effective local exploration ability that enables locating the global optimum even for multimodal functions. Then again, this approach intrinsically abides by the theory of the finite Markov chains, *i.e.*, a succession of trials, where the probability of the outcome of a particular trial is governed by the outcome of the preceding trial. As

a result, when performed iteratively, the computation process can be rather slow (Blum and Roli, 2003). For brevity reasons, this section does not review the theoretical formulation and implementation procedures of PSO and SA. The reader can refer to the original articles by Kennedy and Eberhart (1995) and Kirkpatrick et al. (1983) for more information on PSO and SA algorithms, respectively. In addition, for some practical implementation hints on these optimization algorithms, the reader can refer to the works of Yang (2014) and Yin et al. (2018). In short, based on the review, it has been confirmed that PSO and SA have their strengths and weaknesses, which can be mutualized to enhance the overall performance.

The proposed hybrid PSO-SA technique shares some conceptual similarities with the existing approaches that espoused the complete hybridization of PSO and SA. On the other hand, to tackle the drawbacks discussed above, the proposed strategy preconizes some improvements regarding the algorithm structure and search process. As illustrated in Fig. 1, the proposed strategy capitalizes on the PSO for performing a generalized exploration of the search space, while SA is mainly used to refine the searching process. That is, changing the state of the global best swarm member g^{best} , when its performance gets stacked in the iteration process (prevent the particle getting trapped in a local optimum). Moreover, unlike, Javidrad and Nazari (2017) that focused on

Fig. 1 Schematic illustration of the proposed strategy: a global random search (PSO); b local random search (SA); c refined random search (PSO-SA)



the specific improvement of the random search in PSO, the proposed model (in addition) has privileged the refined searching process, given its importance for escaping local optimum (enhancing the global solution). To this end, a genetic SA was engrafted to PSO procedure. Specifically, a mutation process that creates new neighbours from the initial population was defined for establishing a strong consistency between PSO and SA. This operation is vital as it creates and maintains the diversity of the population, which is essential for solving complex multimodal problems. Besides, a new parameter “nMove” was defined to limit the number of neighbours per PSO individual so as to concomitantly control the amount of refinement and the computation effort. The code of the hybrid PSO-SA procedure is given in Appendix, while the flow chart of the proposed method is presented in Fig. 2.

The performance of the proposed hybrid PSO-SA optimization algorithm was evaluated using a set of 8 benchmark functions known for their complexity

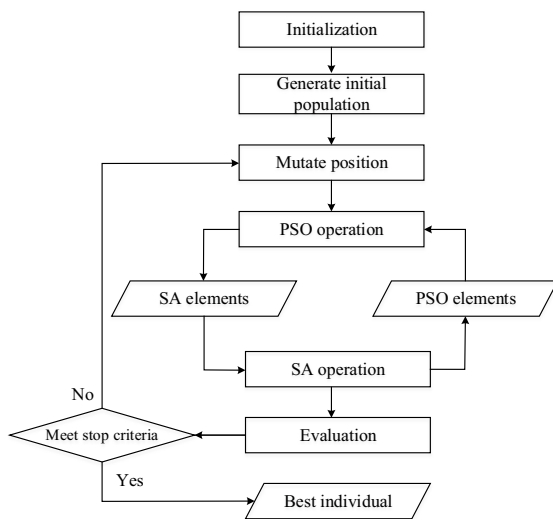


Fig. 2 Framework of proposed hybrid PSO-SA optimization algorithm

Table 2 Parameters settings of algorithms

Algorithms	Parameters
PSO	$w = 1; c_1 = 2; c_2 = 2; wdamp = 0.99$ (damping ratio of inertia coefficient)
SA	$T = 0.1; \alpha = 0.99; nMove = 10$
PSO-SA	$w = 1; c_1 = 2; c_2 = 2; wdamp = 0.99; T = 0.1; \alpha = 0.99; nMove = 20; \mu = 0.5$ (mutation rate)

(Dieterich and Hartke, 2012; Jamil and Yang, 2013; Yao et al. 1999). The experiment was conducted on high dimension (30D) multimodal functions. A comparison with the two previous state-of-the-art optimization methods (PSO and SA) was carried out. The parameter settings of these models are presented in Table 2. The PSO-SA parameters were determined empirically via a systematic trial-and-error process. The population size was taken equal to 50 for all the models, whereas the maximum number of iterations was set to 1000. The aforementioned algorithms were consecutively tested using a MATLAB code ran on an Intel(R) Core(TM) i7-8750H CPU at 2.20 GHz with 16.0 GB of RAM. Table 3 give the statistics (mean value, best, and standard deviation) of the error values achieved by the aforementioned optimization algorithms. On the other hand, the Fig. 3 provide a visual comparison of behaviour of the 3 optimizers on higher dimensional space (i.e., D30) benchmarks functions. This dimension can arguably reflect degree of complexity of some multifaceted real-world problems (than cannot be solved on a lower dimension). All the figures are plotted using a logarithmic scale along the y-axis to clearly appreciate some fundamental difference among the tested algorithms. It should be noted that these statistics were processed based upon 30 repeated runs of each algorithm.

The Fig. 3 shows that the PSO-SA is the most efficient for locating the global optimum of the tested functions. Besides, based on the results provided in Table 3 (best performances are represented in bold), it can be observed that the proposed PSO-SA outperformed both PSO and SA in terms of accuracy. Then again, in terms of convergence, the PSO-SA is by far more effective than the other algorithms. Besides, it can be seen that for the vast majority of investigated functions, the performances of the SA were found superior to that of PSO. These results are consistent with the work of (Gong et al. 2016) that stressed the limitations (premature convergence) of PSO in locating the global optimum of multimodal functions. For

Table 3 Statistical results of high dimension multimodal functions

Function		PSO	SA	PSO-SA
f_1	Mean	0	0	0
	Best	0	0	0
	Std	0	0	0
f_2	Mean	-6.42E-07	-2.49E-08	-7.99E-15
	Best	-3.0981E-10	-1.9655E-08	-7.99E-15
	Std	1.50E-06	2.68E-09	1.58E-30
f_3	Mean	-3.76E-02	-4.59E-03	-3.12E-03
	Best	-0.030404	-2.28E-03	-0.0026112
	Std	7.23E-03	1.76E-03	5.26E-04
f_4	Mean	-4.37E+01	-2.08E+01	-6.28E+00
	Best	-17.9443	-1.92E+01	-5.37E+00
	Std	2.63E+01	8.30E-01	1.10E+00
f_5	Mean	-7.85E-01	-3.67E-01	-2.86E-01
	Best	-6.36E-01	-3.00E-01	-2.00E-01
	Std	7.73E-02	4.71E-02	3.50E-02
f_6	Mean	-4.34E+03	-2.43E+03	-1.03E+03
	Best	-3.30E+03	-1.80E+03	-5.92E+02
	Std	5.40E+02	3.70E+02	3.72E+02
f_7	Mean	0	0	0
	Best	0	0	0
	Std	0	0	0
f_8	Mean	1.17E+03	1.04E+03	1.01E+03
	Best	1.15E+03	9.91E+02	9.77E+02
	Std	1.03E+01	2.91E+01	2.24E+01

instance, for functions with several deep local optima like *Salomon function* f_5 , or *Schwefel function* f_6 , PSO remained trapped into local minima. Conversely, the proposed PSO-SA achieved the lowest error values for these complex multimodal functions, owing to its ability to maintain larger population diversity. This advantageous characteristic can arguably be credited to the mutation operation (diversification of the searching population) proposed in the proposed algorithms. Moreover, in terms of convergence ability (see Fig. 3), it is apparent that the proposed PSO-SA performs the best on the high dimension multimodal functions. Considering its encouraging performances (robust searching ability and rapid convergence), the proposed PSO-SA paradigm is bound to enhance the topology selection of machine learning models in general, and that of random forest in particular.

2.3 Hybridization Random Forest- PSO-SA (Improved Random Forest)

For real world geotechnical problems, the use of traditional machine learning methods in general, and random forest can be hampered by low performances. Thus, the proposed approach crossbreeds the standard random forest model and the previous PSO-SA for achieving optimal performances. The benchmark functions are first replaced by the random forest fitness function, then the PSO-SA model is used to find the values of random forest's hyper-parameters that achieve the lowest fitness value. These hyper-parameters include: the number of trees ('numTrees'), number of features ('maxFeatures'), and number of leaf node ('minLeaf'). The flow chart of IRF is shown in Fig. 4. The process can be divided into six main steps:

Step 1: Load and preprocess data.

Step 2 Initialize the position of the PSO-SA individual vector X , which is composed of the random forest parameters to be tuned $X = [\text{numTrees}, \text{maxFeatures}, \text{minLeaf}]$.

Step 3: Define the objective function to be minimized. In this case we opted for the well-known root mean square error (RMSE) error function.

Step 4: Set the IRF parameters, including the maximum number of iteration and the PSO-SA parameters discussed in Sect. 2.2.

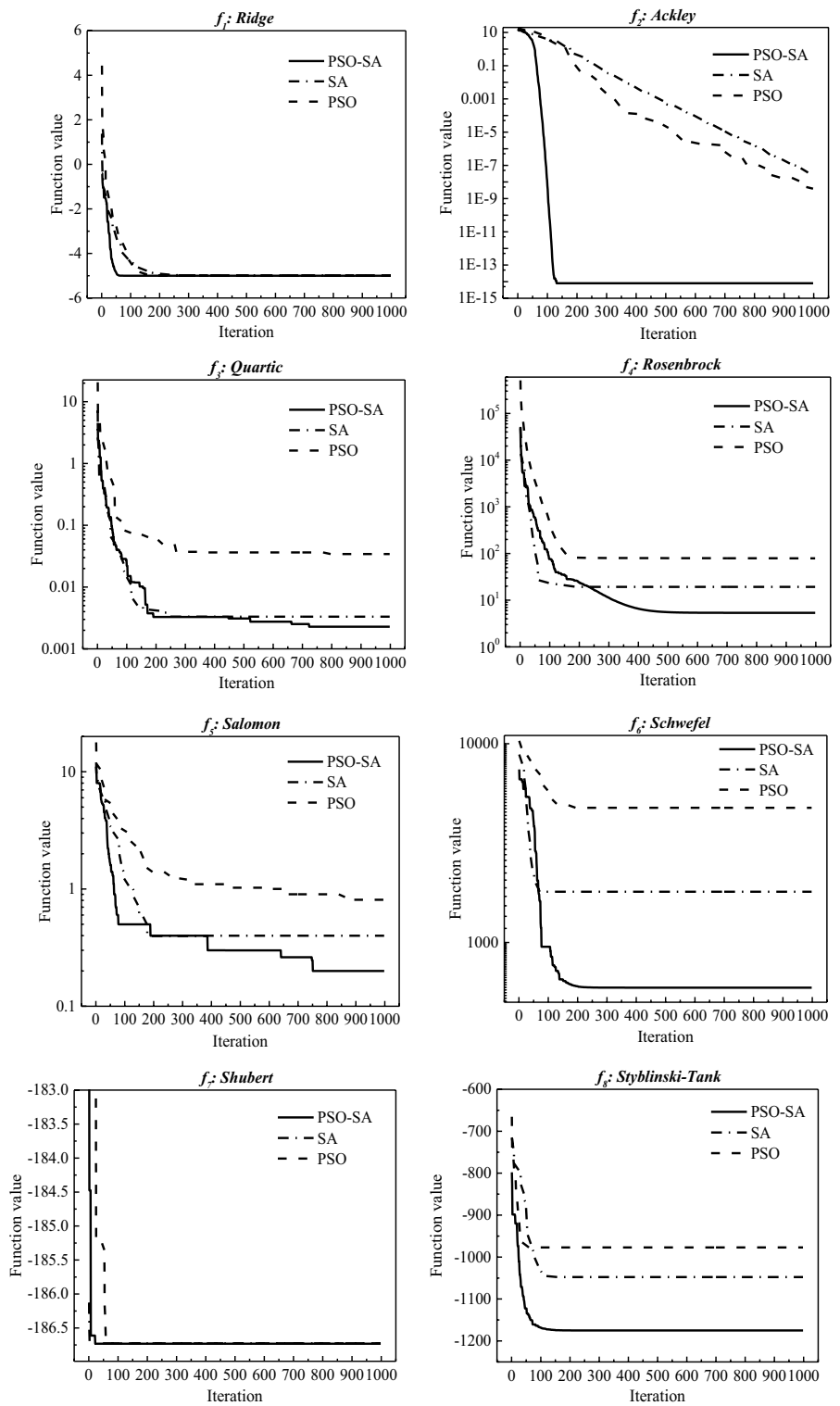
Step 5: Iteratively evaluate (minimize) the fitness function to find the best position of the individual X , via the PSO-SA process discussed previously (i.e., Sect. 2.2).

Step 6: Output the optimum PSO-SA individuals (that satisfied the stopping criteria), i.e., the best hyper-parameters to be assigned to the IRF model.

3 Case study

The proposed improved random forest (IRF) model was applied to the estimation of the lateral ground displacement induced by horizontal jet grouting during a tunnel construction in Shanghai. This section only provides the general background information pertaining to this project; complementary details such as monitoring settings can be found in Wang et al. (2013).

Fig. 3 Testing results of high dimensional multi-modal functions



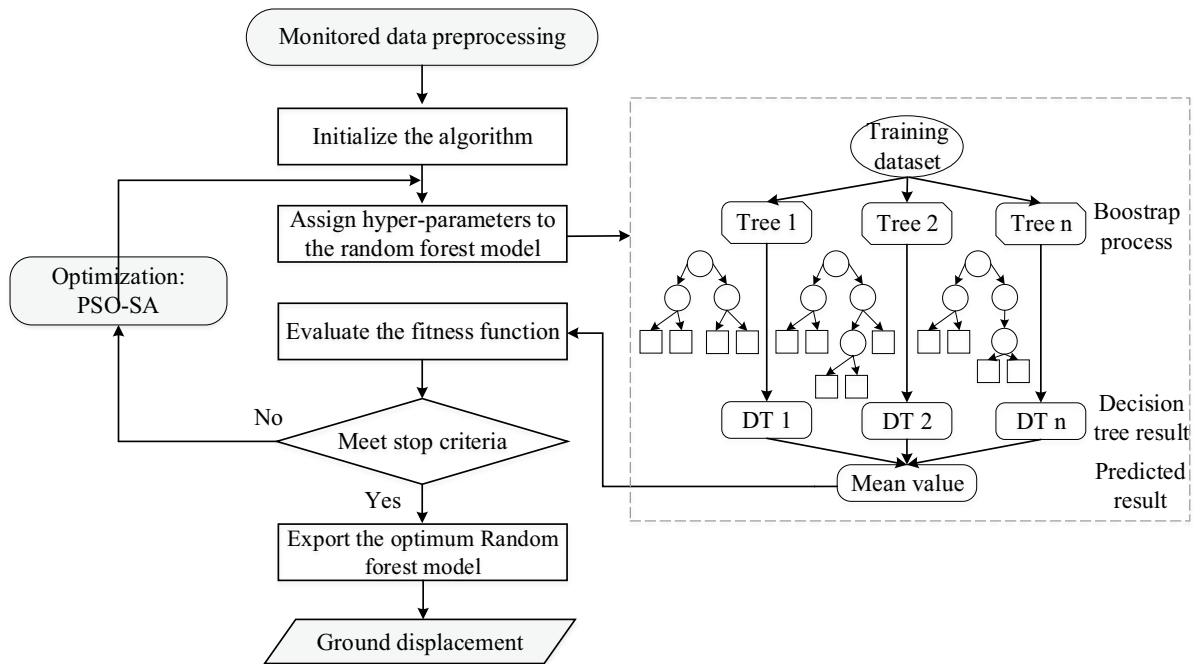


Fig. 4 Simplified illustration of the proposed IRF model

3.1 Site Conditions

The aforementioned tunnel project targeted the creation of the Shanghai Metro Line No. 11. However, the present study only focused on the construction of a metro station located on this line (the ‘*Baili Road*’ metro station). As part of its construction process, a series of horizontal jet grouting operations was undertaken at the entrance shaft of the said metro station. A set of 5 horizontal jet grouted columns were constructed through the tunnel entrance, slightly above the tunnel centreline, which was located 11.65 m beneath the ground surface. The dimensions of the entrance shaft were 24.2 m and 18.9 m for the width and length, respectively. As shown in this figure, a water pipe was detected at a depth of 1.5 m below the ground surface. Furthermore, geotechnical and hydrogeological surveys revealed that the groundwater level varied between 1 and 2 m below the surface of the ground. As shown in Fig. 5, the subsoil profile was mainly composed of 4 layers, including backfill, clayey silt, soft clay and soft silty clay. The unconfined compressive strengths of these soils were found to fluctuate between 30 to 55 kPa, whereas their water content varied between 30 and 50%. Interestingly, the soil layer corresponding to the implementation of

jet grouting was mainly low strength silty clay, with a high time-dependent compressibility characteristic (Axtell et Stark, 2008; Yin et al., 2016). The horizontal jet grouting process was suitably operated using the parameters summarized in Table 4.

3.2 Extraction of Analysis Variables

The present investigation mainly targeted the lateral ground displacement data (see Fig. 6). The proposed IRF was applied for mapping the complex nonlinear relationships between the response variable and input variables. Indeed, the field investigations have allowed delineating 11 input variables and one response variable (lateral ground displacement). As shown in Table 5, five of these input variables relate to the ground conditions, while the remaining input variables relate to the operational jet grouting parameters.

4 Results and Discussions

This section reports and discusses the prediction performance of the proposed IRF model, and also

Fig. 5 Geotechnical profile and soil properties at the construction site (after Wang et al., 2013)

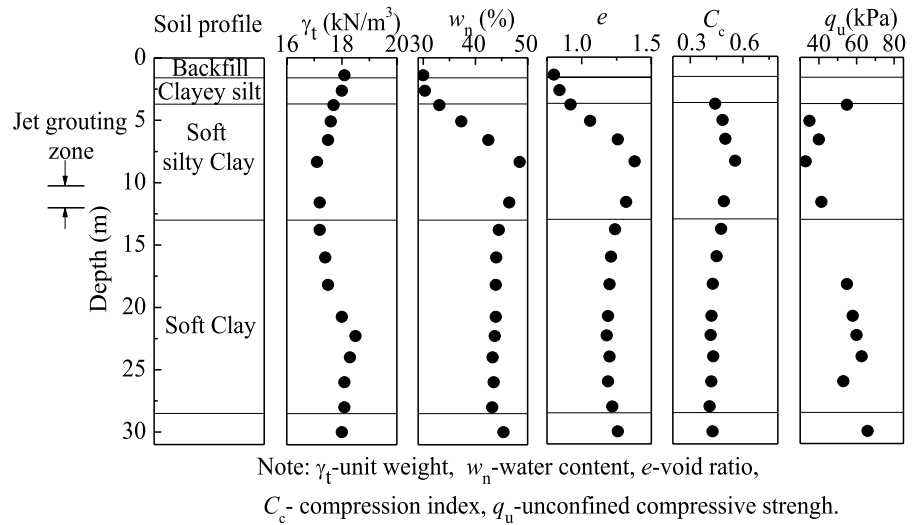


Table 4 Jet grouting construction parameters

Column ID	Air		Cement slurry		Sodium silicate		Jetting parameters	
	Flow rate (m³/min)	Pressure (MPa)	Flow rate (m³/min)	Pressure (MPa)	Flow rate (m³/min)	Pressure (MPa)	Retract rate (m/min)	Rotation rate (rot/min)
1	6	0.7	15	98.01	6	23.1	0.7	15
2	6	0.7	12~15	85.12~98.01	6	23.1	0.7	15
3	6	0.7	10~15	80.12~98.01	3.6	18	0.7	15
4	6	0.7	10~12	85.12~80.92	7.8	26.5	0.6	15
5	6	0.7	12	80.12~98.01	6	23.1	0.6	15

Fig. 6 Lateral displacement observed after the installation of jet grouted columns (After Wang et al. 2013)

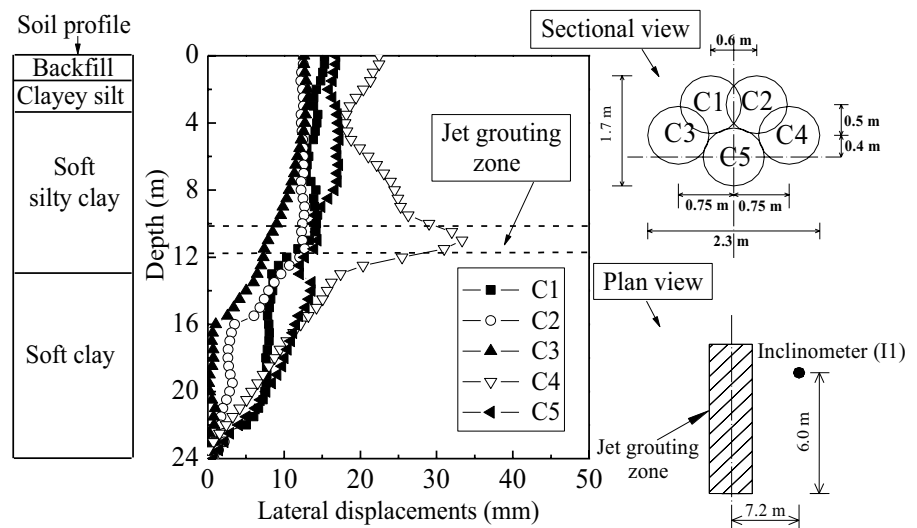


Table 5 Statistic summary of response variable and input variables

	Soil conditions			Jet grouting operational parameters						Lat. Disp.(mm)	
	γ_i (KN/m ³)	w_n (%)	C_c	q_u (kPa)	$P_{c\text{ement}}$ (MPa)	$Q_{c\text{ement}}$ (m ³ /min)	P_{ss} (MPa)	Q_{ss} (m ³ /min)	R_{rate} (m/min)		EPP (kPa)
Mean	17.70	40.70	0.45	51.06	12.35	87.77	5.88	22.76	0.66	17.28	10.74
Std	0.56	5.72	0.018	8.44	1.71	6.31	1.33	2.72	0.05	2.39	6.66
Min	17	30	0.43	37	10	80.12	3.60	18	0.60	14	0.06
Max	18.85	49	0.50	62	15	98.01	7.80	26.50	0.70	21	33.35
Count	245	245	245	245	245	245	245	245	245	245	245

q_u = unconfined compressive strength, C_c = compression index; w_n = water content; γ_i = dry unit weight; $P_{c\text{ement}}$ = cement pressure; $Q_{c\text{ement}}$ = cement flow rate, P_{ss} = sodium silicate pressure; Q_{ss} = sodium silicate flow rate; R_{rate} = retract rate; EPP = excess pore pressure; Lat. Disp. = lateral displacement

presents analyses on how soil and jetting parameters affect lateral ground displacement.

4.1 Benchmark Random Forest Models

Together with the analytical results reported by Wang et al. (2018), two optimized random forest models were adopted to serve as baselines of comparison for the proposed IRF. Specifically, the random forest-particle swarm optimization (RF-PSO) (Malik et al., 2011) and the random forest-genetic algorithm (RF-GA) models were employed. The second optimized model (RF-GA) was developed by the authors for comparison purposes. This algorithm abides by the same principles (hybridization and implementation) as the proposed IRF. In this case, we adopted the following parameters for the genetic algorithms: mutation rate = 0.9; crossover rate = 0.1; population size = 50; Generation = 20. Besides, for all the algorithms, the lower and upper bounds of the hyper-parameters were initialized as follows: numTrees = [10, 150]; numFeatures = [1, 10], numLeaf = [1, 25]. It should also be noted that 80% of the dataset was used to train the proposed model, while the remaining portion (20%) was used for testing.

4.2 Performance of the IRF Model

Table 6 summarizes the simulated results (best performances are represented in bold). These statistics were compiled on the basis of 15 consecutive runs of each algorithm. Generally speaking, it can be seen that all the optimized random forest models performed quite well, with an average correlation coefficient bigger or equal to 0.96. Although the proposed RF-GA presents a higher average coefficient of correlation, its best value was found smaller than the best value of RF-PSO. Compared with the other techniques, the proposed IRF brings about a substantial increase in the prediction ability. This improvement can be attributed to the proposed optimization strategy, namely PSO-SA, which tends to perform better than the others. Indeed, the root mean square error on the test data was found the smallest for the proposed IRF, followed by the RF-GA and RF-PSO.

Moreover, the comparison with the theoretical model by Wang et al. (2018) tends to confirm the efficacy of the proposed approach. As shown in Figs. 7, and 8, the proposed IRF model agrees well with the

Table 6 Evaluation of performance of improved random forest models

Algorithms	Elements of comparison	Mean	Best value	Standard deviation
RF-PSO (Malik et al., 2011)	numTrees	66.667	123	40.925
	numFeatures	4.750	7	0.354
	numLeaf	2.333	2	0.471
	r-squared	0.960	0.9662	0.007
	RMSE	1.303	1.22	0.240
RF-GA (This study)	numTrees	91.000	100	30.474
	numFeatures	3.133	2	0.660
	numLeaf	1.667	2	0.471
	r-squared	0.962	0.9634	0.003
	RMSE	1.220	1.26	0.071
IRF (This study)	numTrees	97.667	93	40.958
	numFeatures	2.333	2	0.471
	numLeaf	2.667	6	2.357
	r-squared	0.979	0.9894	0.008
	RMSE	1.048	1.05	0.003

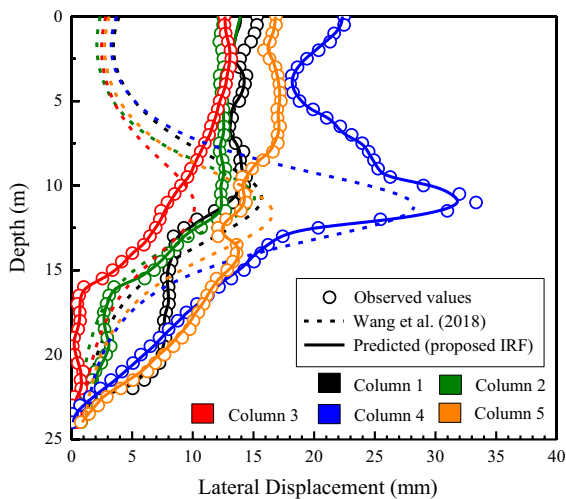


Fig. 7 Prediction performance of proposed IRF model

observed data. On the contrary, the results obtained using Wang et al.’s method reveal the limitations of this analytical approach. In fact, some obvious discrepancies can be observed between the predicted values and the monitored displacements. These discrepancies stem from the inability of this theoretical model to strongly integrate the nonlinearity of the problem, *i.e.*, external factors like overburden stress. Then again, it can be seen that there is a slight disparity between the observed and predicted data (obtained using the proposed model) near the ground surface

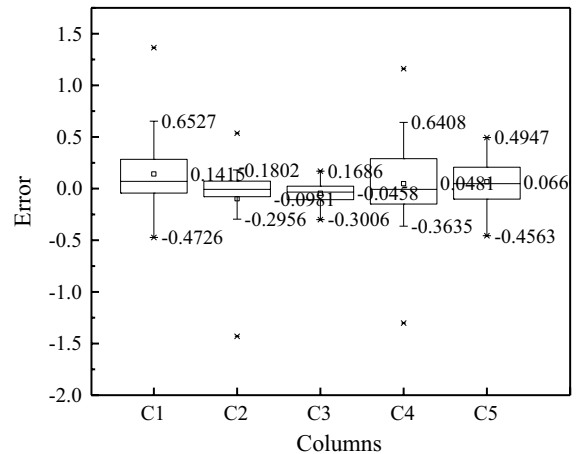


Fig. 8 Residual values achieved by each column

for the columns 1 and 2. This can be ascribed to the strong variability of the input variables of these two columns near the ground surface. This effect is further discussed in Sect. 4.4.

4.3 Effect of Soil and Jetting Parameters on Lateral Ground Displacement

The influence of soil properties and jetting parameters on ground responses was investigated using variable importance evaluation. The random forest is considered as one of the most reliable approaches

for achieving this task due to its voting process and out-of-bag error calculation. Fig. 9 gives the variable importance of the predictors investigated in this study. This plot represents the increase in *VIM* (MSE) (i.e., Eq. 1) resulting from the permuting of out-of-bag observations across each input variable and averaging it over all trees in the forest (overall interpretation). The larger is the out-of-bag variable importance, the more important is the variable.

It can be observed that the ground depth and the retract rate have the highest influence on the lateral displacement during jet grouted column construction. The reason is that the soil properties vary substantially with depth, and as a result, significantly affects the soil erosion process. Then, achieving the targeted jet grouted column diameter in this condition arguably requires a longer exposure time (fluid jet –soil interaction). This also explains the observed relatively large effect of retract rate on the lateral ground movement. The second duet (soil properties—jetting parameters) of input variables with a large variance importance was found to be the soil strength and the sodium silicate pressure. Again, the soil strength in this case can be associated with the *in-situ* geological profile (soil layers), in the sense that hard layers will be more difficult to erode than softer layers. That is to say, the jetting effect on the surrounding soil will be more pronounced in hard layers than softer layers due to the exposure time. Moreover, the sodium silicate tends to play a significant role in the ground movement process, which observation is consistent with the suspicions of Wang et al. Future research

will monitor and consider the sodium silicate gelling time to provide deeper insights on excess pore pressure generation and ground movement. Overall, the pressure and flow rate of injected fluids, as well as the retract rate were found to be the key operating parameters to control in order to reduce environmental effects due to jet grouting operations.

4.4 Interpretation of IRF predictions

To better understand the physics behind the proposed IRF as well as understand the disparities in the lateral displacements following the chronological installation of columns, a local interpretation of the proposed model is performed. That is, for each data point and associated IRF prediction, the contributions from each feature are determined (via inspecting the prediction paths and decomposing the predictions) to explain the said prediction. Mathematically, this translate into the following Equation, where N is the number of features, c_{full} is the bias (value at the root of the node) and $contrib(x, i)$ is the contribution from the i -th feature in the feature vector x .

$$f(x) = c_{full} + \sum_{i=1}^N contrib(x, i) \quad (2)$$

Figure 10 illustrates the contribution of each feature to the model predictions. Particularly, it explains the sources of the strong disparities observed for the columns No 1 and No 2. It can be seen that the strong disparities in predictions (for these two columns)

Fig. 9 Effect of input variables on the ground lateral displacement

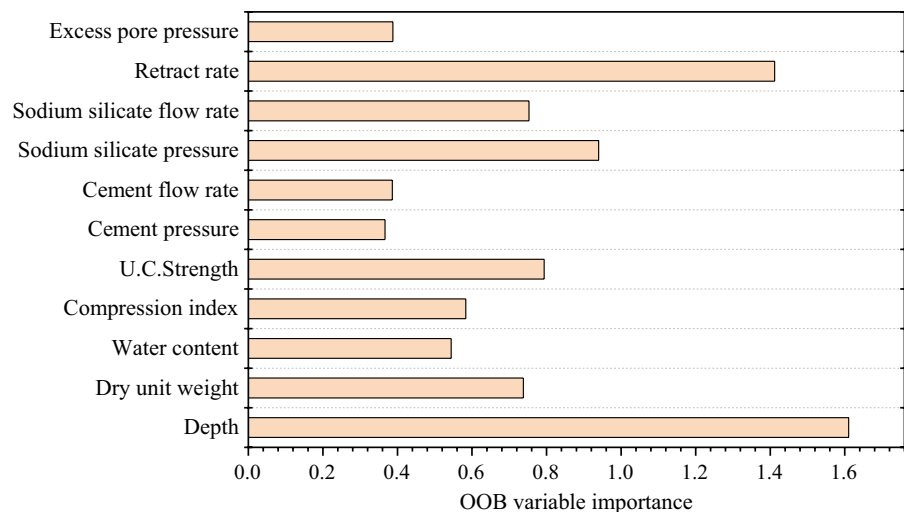
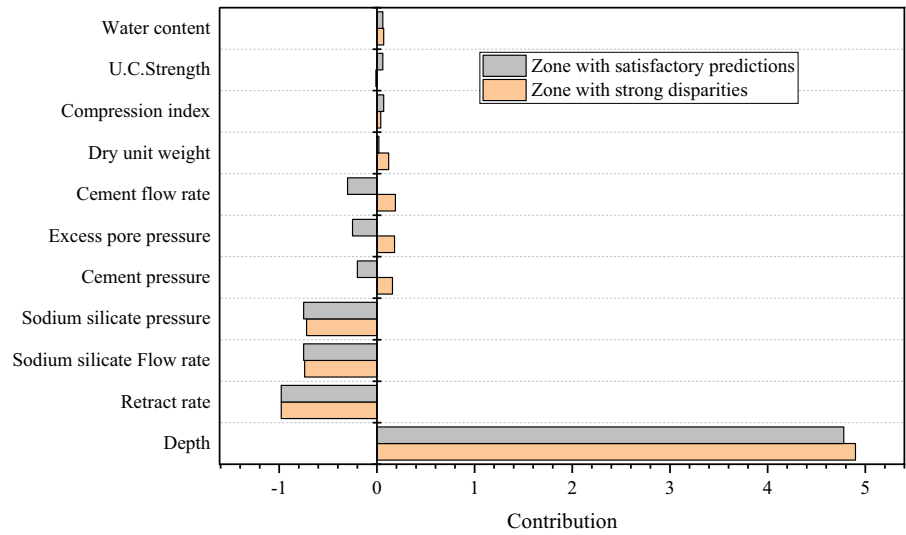


Fig. 10 Effect of different features on IRF model predictions

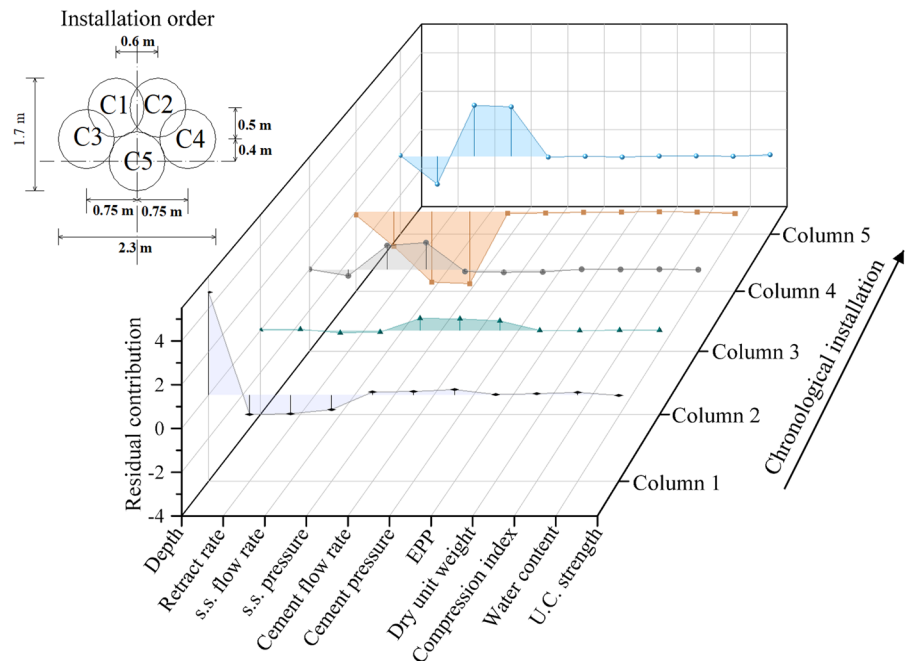


near the ground surface stems mainly from the positive and unexpected contributions of cement pressure, excess pore water pressure and cement flow rate. This decomposition was easily verified via checking that contributions and bias sum up to the predicted value

Further, Fig. 11 exhibits the residual contributions of different parameters throughout the sequential columns' construction, which arguably explain why predicted displacement are different from one column

installation to another. For example, the residual contribution between column 3 and column 4 implies that one reason for the strong disparities between the lateral displacement of columns 4 and the others is the significant negative contribution of retract rate, sodium silicate flow rate and sodium silicate pressure. While on the other hand, the other reason can be ascribed to the distance separating these two columns.

Fig. 11 Residual features contribution to lateral displacement caused by earlier column installation



5 Conclusions

This study proposed a novel intelligent approach, improved random forest (IRF) model, for predicting the ground displacement caused by jet grouting. A new hybrid metaheuristic (PSO-SA) algorithm was developed and tested against a set of 8 well-known benchmark functions. Then, the PSO-SA was incorporated into the random forest model for optimizing its topology to build up the IRF model, the performance of which was then investigated through benchmark models. The main limitation of the proposed model could be its constraint to a specific problem domain. That is, any update of information shall trigger the update of the model topology (number of trees, etc.). Then again, the latter is efficiently tackled by the proposed approach. Overall, the following conclusions can be drawn.

- (1) The proposed hybrid metaheuristic PSO-SA algorithm was found to have significantly better searching and convergence abilities than its counterparts (PSO and SA). These prominent qualities have been substantiated by its ability to timely obtain (near-) optimal solutions when tested on high dimensional complex multimodal benchmark functions. The robustness of the proposed optimization strategy makes it suitable for complex engineering problems.
- (2) The proposed IRF model has shown a substantial improvement in the accuracy compared to the other methods. Besides, the statistical results have demonstrated its stability over consecutive runs. The proposed IRF model can provide an encouraging prediction of the jetting-induced lateral displacement, outperforming the existing analytical methods.
- (3) The variable importance analysis has allowed delineating the key factors affecting the ground lateral displacement resulting from jet grouting operation. The environmental effects of such operation can be controlled via two main operating parameters that include the retract rate of the monitor and the injected fluid pressure.

Acknowledgements The research work described herein was funded by “The Pearl River Talent Recruitment Program” in 2019 (Grant No. 2019CX01G338), Guangdong Province and the Research Funding of Shantou University for New Faculty Member (Grant No. NTF19024-2019). These financial supports are gratefully acknowledged.

Declarations

Conflict of interest The authors declare that there is no conflict of interest.

Appendix

See Table 7.

Table 7 Implementation code of the proposed PSO-SA

```

1: %* Initialization *%
2: for  $i = 1$  to  $nPop$  do
3:   Randomly initialize  $v_i$  and  $x_i$ ;
4:   Evaluate  $f(x_i)$ ;
5:    $i^{best} = x_i$ ;
6: end
7: Set  $g^{best}$  to the current best position of particles;

8: %* Main Loop *%
9: for  $It = 1$  to  $MaxIt$  do
10:  %* Create and evaluate new solutions*%
11:  for  $subIt = 1$  to  $MaxSubIt$  do
12:    for  $i = 1$  to  $nPop$  do
13:      for  $j = 1$  to  $nMove$  do
14:        Create neighbors  $x_{i,j}$  via Mutation of focal particle's positions  $x_i$ 
15:        Evaluate  $f(x_{i,j})$ ;
16:      end
17:    end
18:    for  $i = 1$  to  $nPop$  do
19:      %* Update velocity and position*%
20:       $v_i = w \cdot v_i + c_1 \cdot rand_1(i^{best} - x_i) + c_2 \cdot rand_2(g^{best} - x_i)$ ;
21:       $x_i = x_i + v_i$ ;
22:      Evaluate  $f(x_i)$ ;
23:      %* Update personal best*%
24:      if  $f(x_i)$  is better than  $f(i^{best})$ 
25:         $i^{best} = x_i$ ;
26:         $f(i^{best}) = f(x_i)$ ;
27:      %*Update global best
28:      if  $f(i^{best})$  is better than  $f(g^{best})$ ;
29:         $g^{best} = i^{best}$ ;
30:      end
31:    end
32:    %* Refine the exploration when PSO ceases improving *%
33:    if  $f(x_{i,j})$  is better than  $f(x_i)$ 
34:       $x_i = x_{i,j}$ ;
35:       $f(x_i) = f(x_{i,j})$ ;
36:    else
37:      Accept as new solutions with probability  $P(T, f(x_{i,j}), f(x_i))$ 
38:       $x_i = x_{i,j}$ ;
39:       $f(x_i) = f(x_{i,j})$ ;
40:    end
41:    Update best solution ever found  $g^{best}$ 
42:  end
43:  Update the damping inertia coefficient  $w = w \times wdamp$ ;
44:  Update the temperature  $T = alpha \times T$ ;
45: end

```

References

- Axtell PJ, Stark TD (2008) Increase in shear modulus by soil mix and jet grout methods. *Deep Found Inst J* 2(1):11–21. <https://doi.org/10.1179/dfi.2008.002>
- Blum C, Roli A (2003) Metaheuristics in combinatorial optimization: overview and conceptual comparison. *ACM Comput Surv* 35:268–308. <https://doi.org/10.1145/937503.937505>
- Chai J, Carter JP, Miura N, Zhu H (2009) Improved prediction of lateral deformations due to installation of soil-cement columns. *J Geotech Geoenviron Eng* 135:1836–1845. [https://doi.org/10.1061/\(ASCE\)GT.1943-5606.0000155](https://doi.org/10.1061/(ASCE)GT.1943-5606.0000155)
- Chai JC, Miura N, Koga H (2005) Lateral displacement of ground caused by soil-cement column installation. *J Geotech Geoenviron Eng* 131:623–632. [https://doi.org/10.1061/\(ASCE\)1090-0241\(2005\)131:5\(623\)](https://doi.org/10.1061/(ASCE)1090-0241(2005)131:5(623))
- Coulston JW, Blinn CE, Thomas VA, Wynne RH (2016) Approximating prediction uncertainty for random forest regression models. *Photogramm Eng Remote Sens* 82(3):189–197. <https://doi.org/10.14358/PERS.82.3.189>
- Couronné R, Probst P, Boulesteix AL (2018) Random forest versus logistic regression: a large-scale benchmark experiment. *BMC Bioinf* 19(1):270. <https://doi.org/10.1186/s12859-018-2264-5>
- Dieterich JM, Hartke B (2012) Empirical review of standard benchmark functions using evolutionary global optimization. *Appl Math* 3:1552–1564. <https://doi.org/10.4236/am.2012.330215>
- Dong X, Ouyang D, Cai D, Zhang Y, Ye Y (2010) A hybrid discrete PSO-SA algorithm to find optimal elimination orderings for Bayesian networks. In: 2010 2nd international conference on industrial and information systems, IIS 2010, 1: 510–513. <https://doi.org/10.1109/INDUSIS.2010.5565797>
- Eramo N, Modoni G, Arroyo M (2012) Design control and monitoring of a jet grouted excavation bottom plug. Proc. of the 7th Int. In: Symposium on geotechnical aspects of underground construction in soft ground, TC28 IS Rome, Viggiani ed., Taylor & Francis Group London, 16–18 May 2011, ISBN 978-0-415-66367-8, pp.611–618
- Gong YJ, Li JJ, Zhou Y, Li Y, Chung HSH, Shi YH, Zhang J (2016) Genetic learning particle swarm optimization. *IEEE Trans Cybern* 46:2277–2290. <https://doi.org/10.1109/TCYB.2015.2475174>
- Jamil M, Yang XS (2013) A literature survey of benchmark functions for global optimisation problems. *Int J Math Model Numer Optim* 4:150–194. <https://doi.org/10.1504/IJMMNO.2013.055204>
- Javidrad F, Nazari M (2017) A new hybrid particle swarm and simulated annealing stochastic optimization method. *Appl Soft Comput J* 60:634–654. <https://doi.org/10.1016/j.asoc.2017.07.023>
- Kennedy J, Eberhart R (1995) Particle swarm optimization. In: Proceedings of ICNN'95 - international conference on neural networks, 4:1942–1948. doi:<https://doi.org/10.1109/ICNN.1995.488968>
- Kimpritis T, Standing JR, Thurner R (2018) Estimating column diameters in jet-grouting processes. *Proc Inst Civ Eng* *Ground Improv* 171(3):148–158. <https://doi.org/10.1680/jgrim.17.00001>
- Kirkpatrick S, Gelatt CD Jr, Vecchi MP (1983) Optimization by simulated annealing. *Sci* 220(4598):671–680. <https://doi.org/10.1126/science.220.4598.671>. PMID: 17813860
- Liu D, Fan Z, Fu Q, Li M, Faiz MA, Ali S, Li T, Zhang L, Khan MI (2020) Random forest regression evaluation model of regional flood disaster resilience based on the whale optimization algorithm. *J Clean Prod* 250:119468. <https://doi.org/10.1016/j.jclepro.2019.119468>
- Malik AJ, Shahzad W, Khan FA (2011) Binary PSO and random forests algorithm for PROBE attacks detection in a network. In: 2011 IEEE congress of evolutionary Computation, CEC 2011. New Orleans, US, pp. 662–668. <https://doi.org/10.1109/CEC.2011.5949682>
- Meng Q, Zhang L, Fan Y (2016) A hybrid particle swarm optimization algorithm for solving job shop scheduling problems. *Commun Comput Inf Sci* 644:71–78. https://doi.org/10.1007/978-981-10-2666-9_8
- Ochmański M, Modoni G, Bzówka J (2015a) Prediction of the diameter of jet grouting columns with artificial neural networks. *Soils Found* 55(2):425–436. <https://doi.org/10.1016/j.sandf.2015.02.016>
- Ochmański M, Modoni G, Bzówka J (2015b) Numerical analysis of tunnelling with jet-grouted canopy. *Soils Found* 55(5):929–942. <https://doi.org/10.1016/j.sandf.2015.08.002>
- Palmer DS, O'Boyle NM, Glen RC, Mitchell JBO (2007) Random forest models to predict aqueous solubility. *J Chem Inf Model* 47(1):150–158. <https://doi.org/10.1021/ci060164k>
- Poh TY, Wong IH (2001) A field trial of jet-grouting in marine clay. *Can Geotech J* 38:338–348. <https://doi.org/10.1139/t00-093>
- Segal MR (2004) Machine learning benchmarks and random forest regression. UCSF: Center for bioinformatics and molecular biostatistics. Retrieved from <https://escholarship.org/uc/item/35x3v9t4>
- Shen SL, Atangana Njock PG, Zhou A, Lyu HM (2020) Dynamic prediction of jet grouted column diameter in soft soil by using Bi-LSTM deep learning. *Acta Geotech* 16(1):303–315. <https://doi.org/10.1007/s11440-020-01005-8>
- Shieh HL, Kuo CC, Chiang CM (2011) Modified particle swarm optimization algorithm with simulated annealing behavior and its numerical verification. *Appl Math Comput* 218:4365–4383. <https://doi.org/10.1016/j.amc.2011.10.012>
- Stark TD, Axtell PJ, Lewis RJ, Dillon JC, Empson WB, Topi JE, Walberg FC (2009) Soil inclusions in jet grout columns. *Deep Foundation Institute Journal* 3(1):44–55. <https://doi.org/10.1179/dfi.2009.004>
- Stark TD, Axtell PJ, Walberg FC, Dillon JC, Bellow GM, Matthews DL (2012) Jet grouting and safety of tuttle creek dam. *Deep Foundation Inst J* 6(1):3–20. <https://doi.org/10.1179/dfi.2012.001>
- Sudibyo S, Murat MN, Aziz N (2015) Simulated annealing-particle swarm optimization (SA-PSO): Particle distribution study and application in neural wiener-based NMPC. In: 2015 10th Asian control conference: emerging control

- techniques for a sustainable world, ASCC 2015. <https://doi.org/10.1109/ASCC.2015.7244567>
- Wang XH, Li JJ (2004) Hybrid particle swarm optimization with simulated annealing. In: Proceedings of 2004 international conference on machine learning and cybernetics, 4: 2402–2405. Doi: <https://doi.org/10.1109/icmlc.2004.1382205>
- Wang ZF, Shen SL, Ho C-E, Kim Y-H (2013) Investigation of field-installation effects of horizontal twin-jet grouting in Shanghai soft soil deposits. *Can Geotech J* 50:288–297. <https://doi.org/10.1139/cgj-2012-0199>
- Wang ZF, Shen SL, Cheng WC (2018) Simple method to predict ground displacements caused by installing horizontal jet-grouting columns. *Math Probl Eng*. <https://doi.org/10.1155/2018/1897394>
- Wei X, Zhang LL, Yang HQ, Zhang LM, Yao YP (2021) Machine learning for pore-water pressure time-series prediction: application of recurrent neural networks. *Geosci Front* 12(1):453–467. <https://doi.org/10.1016/j.gsf.2020.04.011>
- Wu YD, Diao HG, Ng CWW, Liu J, Zeng CC (2016) Investigation of ground heave due to jet grouting in soft clay. *J Perform Constr Facil* 30(6):06016003. [https://doi.org/10.1061/\(ASCE\)CF.1943-5509.0000910](https://doi.org/10.1061/(ASCE)CF.1943-5509.0000910)
- Yang X (2014) *Nature-inspired optimization algorithms*, 1st Edition. Elsevier, London, p 300
- Yao X, Liu Y, Lin G (1999) Evolutionary programming made faster. *IEEE Trans Evol Comput* 3:82–102. <https://doi.org/10.1109/4235.771163>
- Yin ZY, Wu ZX, Hicher PY (2018) Modeling Monotonic and Cyclic Behavior of Granular Materials by Exponential Constitutive Function. *J Eng Mech* 144(4):04018014. [https://doi.org/10.1061/\(ASCE\)EM.1943-7889.0001437](https://doi.org/10.1061/(ASCE)EM.1943-7889.0001437)
- Yin ZY, Huang HW, Hicher PY (2016) Elastoplastic modeling of sand-silt mixtures. *Soil Found* 56(3):520–532. <https://doi.org/10.1016/j.sandf.2016.04.017>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.