Journal of Zhejiang University-SCIENCE A (Applied Physics & Engineering)
ISSN 1673-565X (Print); ISSN 1862-1775 (Online)
www.jzus.zju.edu.cn; www.springerlink.com
E-mail: jzus@zju.edu.cn

Estimation of spatiotemporal response of rooted soil using a machine learning approach*

Zhi-liang CHENG[1], Wan-huan ZHOU[†‡1,2], Zhi DING[1,3], Yong-xing GUO[1,4]

[1]*State Key Laboratory of Internet of Things for Smart City and Department of Civil and Environmental Engineering, University of Macau, Macau SAR 999078, China*

[2]*Zhuhai UM Science and Technology Research Institute, Zhuhai 519000, China*

[3]*Department of Civil and Environmental Engineering, School of Engineering, Zhejiang University City College, Hangzhou 310015, China*

[4]*Key Laboratory of Metallurgical Equipment and Control Technology of Ministry of Education, Wuhan University of Science and Technology, Wuhan 430081, China*

[†]E-mail: hannahzhou@um.edu.mo

Received Oct. 29, 2019; Revision accepted Apr. 24, 2020; Crosschecked May 23, 2020

Abstract: In this study, a machine learning method, i.e. genetic programming (GP), is employed to obtain a simplified statistical model to describe the variation of soil suction in drying cycles using five selected influential parameters. The data used for model development was recorded by an in-situ experiment. The image processing technology is used to quantify several tree canopy parameters. Based on four accuracy metrics, i.e. root mean square error (RMSE), mean absolute percentage error (MAPE), coefficient of determination (R^2), and relative error, the performance of the proposed GP model was evaluated. The results indicate that the model can give a reasonable estimation for the spatiotemporal variations of soil suction around a tree with acceptable errors. Global sensitivity analysis for the statistical model obtained using limited data of a specific region demonstrates the drying time as the most influential variable and the initial soil suction as the second most influential variable for the soil suction variations. A case study was conducted using a set of assumed input variable values and validated that the simplified GP model can be used to estimate and predict the spatiotemporal variations of soil suction in rooted soil at a certain range.

Key words: Genetic programming (GP); Simplified statistical model; Spatiotemporal variations; Soil suction
https://doi.org/10.1631/jzus.A1900555 **CLC number:** TU413.7

1 Introduction

Soil suction plays an important role in shallow soil layers, because it is closely related to unsaturated soil characteristics, such as soil water retention property, soil shear strength in the unsaturated zone, and effective stress (Fredlund and Xing, 1994; Fredlund et al., 2002; Kim et al., 2004; Hossain and Yin, 2010; Zhou et al., 2013; Nowamooz et al., 2016; Tan et al., 2016, 2018; Crawford et al., 2019), which can reflect the behaviors of green geotechnical infrastructures. It is well known that plant is generally applied to some green geotechnical infrastructures (Leung et al., 2015; Ni et al., 2018; Gadi et al., 2019; Zhou and Qi, 2019). Therefore, in the unsaturated zone of green geotechnical infrastructures, soil suction suffers serious impact by the change of local weather and plant coverage under natural environmental and climatic conditions (Lee et al., 2009; Cui et al., 2010; Yang et al., 2015; Zhou et al., 2017, 2020; Gadi et al., 2018; Zhu et al., 2018; Garg et al.,

‡ Corresponding author

* Project supported by the National Key R&D Program of China (No. 2019YFB1600700), the Science and Technology Development Fund of Macau (Nos. SKL-IOTSC-2018-2020 and 0193/2017/A3), and the University of Macau Research Fund (No. MYRG2018-00173-FST), China

ORCID: Zhi-liang CHENG, https://orcid.org/0000-0002-0607-8912; Wan-huan ZHOU, https://orcid.org/0000-0001-5183-9947

2019a, 2019b). Accordingly, it is significant to develop a computational model to estimate and predict the spatiotemporal variations of soil suction considering the effects of plant and weather, which will contribute to stability analysis, design, and maintenance of some green geotechnical infrastructures.

Soil suction is usually expressed as a function of soil water content (Fredlund et al., 2002). Some methods can be used for prediction of soil suction, such as analytical method and finite element method (Prasad, 1988; Totoev and Kleeman, 1998; Fredlund et al., 2011; Hemmati et al., 2012; Too et al., 2014; An et al., 2017; Qi and Zhou, 2017; Qi et al., 2017; Gamse et al., 2018; Zhou et al., 2018; Feng S et al., 2019). However, these analytical or empirical solutions are usually integrated with a series of assumption and setting of boundary conditions, which leads to difficulty in accurately estimating the variations of soil suction under complicated environmental conditions. With the rapid development of smart monitoring techniques, variations of soil suction under different natural environmental conditions can be recorded for a long time. Besides, the corresponding influential factors, such as covered plant activities and local weather, can also be obtained at the same time interval. The measured data can directly reflect the complicated relationships between soil suction and those influential factors. In recent years, machine learning techniques have attracted much attention and are used to analyze complicated relationships between different types of variables by developing a statistical mathematical model based on sample data, known as "training data," which can also be used to make predictions or decisions (Bishop, 2006). Many researchers have employed machine learning techniques to analyze the relations among the soil parameters and different types of influential factors (Ahmad et al., 2010; Samui and Sitharam, 2011; Srivastava et al., 2013; Brungard et al., 2015; Karandish and Šimůnek, 2016; Yin et al., 2017, 2018; Pham et al., 2018; Feng Y et al., 2019; Gopal et al., 2019; He et al., 2019; Jin et al., 2019a, 2019b; Cheng et al., 2020; Wang, 2020; Zhang et al., 2020). Machine learning is a branch of artificial intelligence based on the biological learning process (Lary et al., 2016), and is involved in several artificial intelligence methods, one of which, genetic programming (GP) has been applied to many studies in geotechnical engineering (Whigham and Crapper, 2001; Javadi et al., 2006; Makkeasorn et al., 2006; Parasuraman et al., 2007; Rezania and Javadi, 2007; Alavi and Gandomi, 2011; Gandomi and Alavi, 2012; Kisi et al., 2012; Roushangar et al., 2014; Pétrowski and Ben-Hamida, 2017). Johari et al. (2006) used the GP method to obtain a mathematical model of soil-water characteristic curve considering the effects of some soil physical properties, such as initial void ratio and clay content. Garg et al. (2015) developed the GP model to depict the relationship between unsaturated soil properties and several vegetation-related parameters. In some cases, the GP model can give a better performance than other machine learning methods. Zhou et al. (2016) demonstrated that the performance of the GP model was the best in curve-fitting of soil water retention properties, comparing with other two machine learning methods, artificial neural network (ANN) and support vector regression (SVR). Alemdag et al. (2016) also came to the same conclusion, which refers to that the GP model gave the best performance by comparison of different deformation modulus values calculated using different approaches. Prevailing studies related to estimation/prediction of soil suction are mainly based on soil water characteristic curves (SWCCs), and fewer studies are performed to estimate/predict field-monitored soil suction only using vegetation and atmosphere parameters (Johari et al., 2006; Garg et al., 2015). Accordingly, the significance of this study lies in providing a way to estimate/predict field-monitored soil suction variations using a set of simple input variables that are related to soil water content but not soil water content itself.

In this study, the working principle of GP and two methods for global sensitivity analysis are introduced first. Then, a field monitoring test was performed for quantification of soil parameters and corresponding local plant and weather parameters. Based on the settings of variables and parameters for GP, a simplified statistical model was obtained using the data collected from the field monitoring for estimating and predicting soil suction values of in-situ experiment considering the effects of local plant coverage and weather. The performance of the proposed simplified statistical model was evaluated based on four accuracy metrics, i.e. the root mean square error (RMSE), mean absolute percentage error (MAPE), coefficient of determination (R^2), and relative error. The importance of each input variable on soil suction was investigated based on two variance-based global sensitivity analysis methods. A case study was performed for validating the ability of the proposed GP model in the estimation and prediction of the spatiotemporal variations of soil suction.

2 Methodology

2.1 Genetic programming

GP is an evolutionary algorithm technique based on Darwin's theory of "survival of the fittest" (Poli et al., 2008), and it was invented by Cramer (1985). As a type of machine learning approach, GP can automatically generate an optimal structured representation for a set of input/output variables according to user's settings of required parameters without any assumption of the solution structure in advance (Johari et al., 2006; Shahin, 2015). GP programs can be described using syntax trees, which mainly include functional and terminal nodes, as shown in Fig. 1 (Cheng et al., 2020). For different problems, the user can give a different setting of the functional and terminal modes according to corresponding requirements.

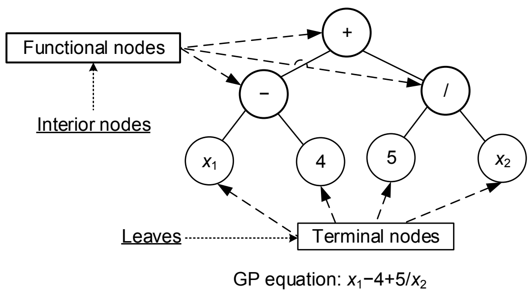

Fig. 1 Typical GP syntax tree. Reprinted from (Cheng et al., 2020), Copyright 2020, with permission from Elsevier

The working flow chart is shown in Fig. 2. For a particular problem, the corresponding data is prepared for model development. The function and terminal settings can be made by the requirements or trial-and-error method. The initial population can be generated randomly according to parameter settings, which represents the starting of the modelling by GP. Then, the generated population can be evaluated according to the fitness function that usually refers to the error between the fitted and actual output values. If the performance of the model does not satisfy the requirement, and it does not reach the maximum number of the generations, the new population would be generated by three basic operations: reproduction, crossover, and mutation. Until the model fitness satisfies the requirement or it reaches the maximum number of the generations, the GP would not be ended. A commonly used function for evaluating the performance of a model is the RMSE:

$$\mathrm{RMSE}=\sqrt{\frac{\sum_{i=1}^{n}\left|P_i-A_i\right|^2}{n}}, \tag{1}$$

where P_i and A_i represent the fitted and measured values of the ith data point, respectively, and n is the total number of the data points.

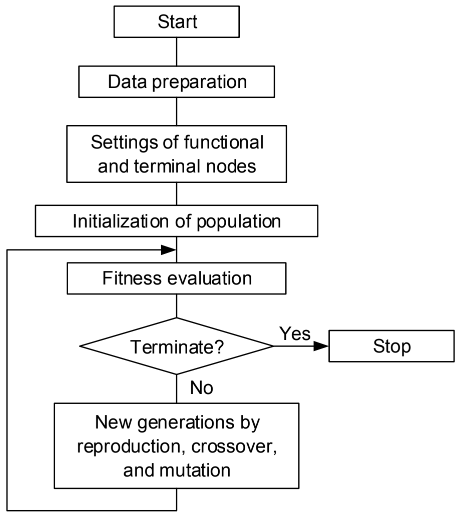

Fig. 2 Working flow chart of GP

2.2 Global sensitivity analysis

In this study, a sensitivity analysis was performed to investigate how different input variables affect the variations of the output variable, which can provide a further understanding of the relationships between independent and dependent variables in a computational model. Because the obtained model is relatively complicated, global sensitivity analysis (GSA) is a better choice to reflect different sensitivity indexes over the entire input variable space (Saltelli and Sobol', 1995; Saltelli et al., 2010; Cannavó, 2012; Ni et al., 2018). As one of the GSA methods, the variance-based sensitivity analysis has been considered as an effective approach to provide not only the ranking of different variables' importance, but also quantitative sensitivity coefficients for revealing the effect of one variable or the coupled effect of multiple variables (Dai and Ye, 2015; Ni et al., 2018). With respect to variance-based global sensitivity analysis, the Fourier amplitude sensitivity test (FAST) and the Sobol' sensitivity analysis are two robust approaches and were used for global sensitivity analysis in this study (Cukier et al., 1973; Sobol', 1990; Cannavó, 2012).

The variance-based global sensitivity analysis aims to quantify the contribution of each input variable to the total variance of the output variable. It is assumed that a function $y=f(\boldsymbol{X})$ describes the relation between one dependent variable y and the independent variable $\boldsymbol{X}$, where $\boldsymbol{X}$ is the vector of the m-input variables, $\boldsymbol{X}=(x_1, x_2, \ldots, x_m)$. The first-order sensitivity coefficient obtained by variance-based global sensitivity analysis is described as

$$S_i = \frac{D(E(y \mid x_i))}{D(y)}, \tag{2}$$

where x_i is the ith independent variable, and D refers to variance calculation. The conditional expectation value $E(y|x_i)$ equals the mean value of all possible values of y that are calculated using all x_m ($m\neq i$) except for x_i variable. Thereby, the variance calculation $D(E(y|x_i))$ expresses the conditional variance related to x_i. $D(y)$ represents the total variance of y. Similarly, the second-order coefficient of variance-based sensitivity analysis is described as

$$S_{ij} = \frac{D(E(y \mid x_{ij}))}{D(y)}, \quad i \neq j. \tag{3}$$

The higher-order sensitivity coefficient can be calculated according to the aforementioned similar regulation. For the global sensitivity analysis calculation, $E(y)$ can be calculated using the following integral:

$$E(y) = \int_{\boldsymbol{I}^m} f(x)\mathrm{d}x, \tag{4}$$

where $\boldsymbol{I}^m$ represents the unit hypercube in the m-dimensional spaces. In order to calculate the integral efficiently, a commonly used decomposition, analysis of variance (ANOVA), is used as follows (Cannavó, 2012):

$$f(\boldsymbol{X}) = f_0 + \sum_{k=1}^{m} \sum_{i_1<i_2<\cdots<i_k} f_{i_1 i_2 \cdots i_k}(x_{i_1}, x_{i_2}, \cdots, x_{i_k}), \tag{5}$$

where f_0 is the expected value of this function when the independent variables are uniformly distributed, and the second term consists of all combinations of k (k=1, 2, …, m) independent variables.

The integral in Eq. (4) can be calculated using the FAST method according to the ergodic theorem (Weyl, 1938). Thus, the variance values are obtained based on a series of Fourier coefficients as

$$D\{f_{i_1 i_2 \cdots i_k}\} = \sum_{t_{i_1}=-\infty}^{\infty} \sum_{t_{i_2}=-\infty}^{\infty} \cdots \sum_{t_{i_k}=-\infty}^{\infty} |\boldsymbol{C}_{t_{i_1} t_{i_2} \cdots t_{i_k}}|, \tag{6}$$

where $\boldsymbol{C}_{t_{i_1} t_{i_2} \cdots t_{i_k}}$ is the Fourier coefficient. Eq. (6) can be solved by translating $f(\boldsymbol{X})$ into the Fourier series as

$$f(\boldsymbol{X}) = \sum_{t_1=-\infty}^{\infty} \sum_{t_2=-\infty}^{\infty} \cdots \sum_{t_m=-\infty}^{\infty} \boldsymbol{C}_{t_1 t_2 \cdots t_m} \mathrm{e}^{2\pi \mathrm{j}(t_1 x_1 + t_2 x_2 + \cdots + t_m x_m)}, \tag{7}$$

with Fourier coefficient

$$\boldsymbol{C}_{t_1 t_2 \cdots t_m} = \int_{\boldsymbol{I}^m} f(\boldsymbol{X}) \mathrm{e}^{-2\pi \mathrm{j}(t_1 x_1 + t_2 x_2 + \cdots + t_m x_m)} \mathrm{d}\boldsymbol{X}, \tag{8}$$

where j is the imaginary unit, and t_1–t_m refer to different coefficients generated during Fourier translating for different input variables. In Eq. (5), the component $f_{i_1 i_2 \cdots i_k}(x_{i_1}, x_{i_2}, \cdots, x_{i_k})$ can be substituted by the form of Fourier series according to the ANOVA decomposition method (Cannavó, 2012).

The coefficient of variance-based global sensitivity analysis by the Sobol' sensitivity (Sobol', 2001) is defined as the ratio of one variance to the total variance:

$$\boldsymbol{S}_{i_1 i_2 \cdots i_k} = \frac{\boldsymbol{D}_{i_1 i_2 \cdots i_k}}{\boldsymbol{D}}, \tag{9}$$

where

$$\begin{cases} \boldsymbol{D}_{i_1 i_2 \cdots i_k} = \int_0^1 f_{i_1 i_2 \cdots i_k}^2 \mathrm{d}x_{i_1} \mathrm{d}x_{i_2} \cdots \mathrm{d}x_{i_k}, \\ \boldsymbol{D} = \int_0^1 f^2(\boldsymbol{X}) \mathrm{d}\boldsymbol{X} - f_0^2 = \sum_{k=1}^{m} \sum_{i_1 < i_2 < \cdots < i_k} \boldsymbol{D}_{i_1 i_2 \cdots i_k}. \end{cases} \tag{10}$$

Sobol's global sensitivity analysis can be carried out integrating with the Monte Carlo method (Kucherenko and Shah, 2007). Detailed calculation procedures can be found in (Sobol', 1990, 2001; Chan et al., 2000; Kucherenko and Shah, 2007).

3 Field monitoring

A site covered with trees in Macau, China was selected for recording the variations of soil suction and corresponding plant and weather parameters. In the test site, the selected objective tree is called as *Elaeocarpus apiculatus* Master. The soil in the unsaturated zone of the in-situ test site is classified as well-graded sand (SW) based on the Unified Soils Classification System (ASTM, 2011). More details regarding the site information can be found in (Zhou et al., 2020).

In this study, the data of six monitoring points labelled as A_1, A_2, B_1, B_2, C_1, and C_2 (Fig. 3) was used for model development. The six monitoring points were distributed at different distances, 0.5 m, 1.5 m, and 3.0 m, from the selected tree and at different depths, 0.2 m and 0.4 m, and the detailed locations of them are shown in Fig. 3. Different types of sensors (soil water potential sensor, soil water content, air temperature/relative humidity sensor, rainfall gauge, etc.), data logger, and a drone (Fig. 4) were employed to quantify the soil parameters and plant and weather parameters. The data record interval is half an hour. The monitoring period in this study is from 21st Nov. 2017 to 11th June 2018. In this monitoring period, the drone was used to take the images of the tree canopy for two times, on 22nd Dec. 2017 and 19th Mar. 2018. ImageJ software was employed to quantify the tree canopy area and its radius, as shown in Fig. 5. The tree canopy area at different times can be quantified first, and then the average radius of tree canopy can be calculated based on the relationship between radius and area of a circle by assuming that the tree canopy is circular (Wikipedia, 2019). Due to no significant variation of the tree canopy in this monitoring period (the area of tree canopy varies from 5.998 m^2 to 5.975 m^2), the radius of tree canopy used in the following modeling analysis is the mean value of 1.380 m.

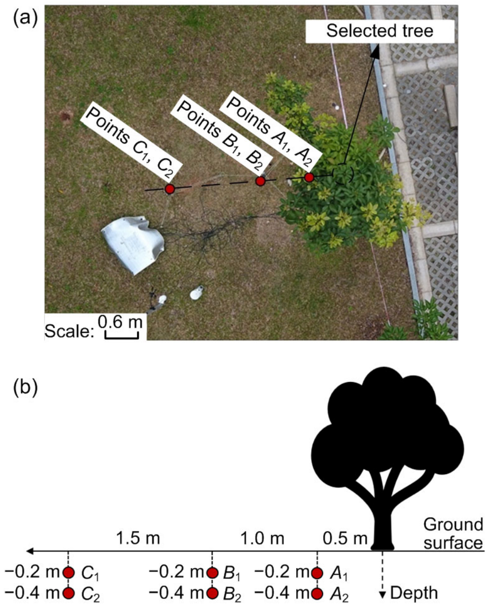

Fig. 3 Distribution of six monitoring points considered in the field monitoring

Fig. 4 Different types of sensors and equipment used

4 Parameter settings

4.1 Variable settings for modeling

This study aims to develop a statistical model for estimation and prediction of spatiotemporal variations of soil suction values in drying cycles collected from the in-situ test under the natural environmental

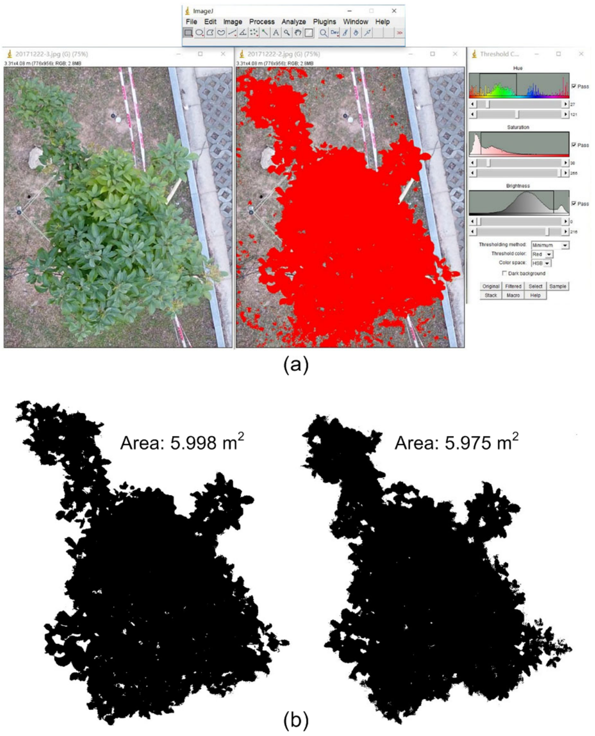

Fig. 5 Image processing for the tree canopy area and radius: (a) operation interface; (b) results of image processing

conditions. Accordingly, the data of the soil suction in drying cycles at different distances from the tree (horizontal direction) and different depths (vertical direction), and the corresponding influential parameters were selected for the model development. The measured soil suction values in drying cycles were set as the dependent variable, y. Initially, some atmosphere and vegetation parameters that affect the variations of soil suction, such as air temperature, relative humidity, soil temperature, and drying time, were all considered as the independent variables. In the regression model, the multicollinearity testing was performed for the selection of input variables by the variance inflation factor. After multicollinearity estimation for different input variables, five independent variables, initial soil suction (x_1), air relative humidity (x_2), drying duration (x_3), the ratio of the distance from the tree, dist., to the radius of the tree canopy, r (x_4), and the depth (x_5), were selected as the input variables for model development. Among them, the ratios of the distances from tree to the radius of tree canopy at six monitoring positions are 0.36, 1.09, and 2.17, respectively, according to the results of image processing for tree canopy. Three indices, variance inflation factor (VIF), coefficient of correlation R, and the tolerance T, of each two independent variables are calculated for multicollinearity evaluation, as shown in Table 1. The evaluation criteria is VIF=1/T<5, which indicates that there is no multicollinearity problem between different variables (Kalnins, 2018).

Table 1 Multicollinearity evaluation of input variables used for drying-cycle model

Variable	Multicollinearity diagnostic		
	R	T	VIF
x_1 & x_2	0.106	0.989	1.011
x_1 & x_3	−0.173	0.970	1.031
x_1 & x_4	−0.359	0.871	1.148
x_1 & x_5	−0.301	0.909	1.100
x_2 & x_3	−0.475	0.775	1.291
x_2 & x_4	0.000	1.000	1.000
x_2 & x_5	0.000	1.000	1.000
x_3 & x_4	0.000	1.000	1.000
x_3 & x_5	0.000	1.000	1.000
x_4 & x_5	0.000	1.000	1.000

Fig. 6 shows the variation curves of the selected input and output variables for model development. The soil suction values used in this study include the selected drying cycles of five time periods of six monitoring points at different distances from the tree and different depths. Based on the selected soil suction values of five drying cycles, the corresponding input variables for each cycle were prepared. For the six monitoring points of the same time period, the relative humidity and the drying time are the same, as shown in Figs. 6b and 6c. It is noted that in order to better display the drying time, the unit of the drying time in Fig. 6c is transformed into day, but in model development, the drying time uses the data with the unit of hour. The selected drying cycles of five time periods include a total of 13 620 data values of soil suction. From Fig. 6f, it can be seen that the soil suction of the monitoring points at the closer distance, 0.5 m, from the tree or at the shallower depth, 0.2 m, shows more visible variation, and the difference of soil suction of different monitoring points at a depth of 0.2 m is more significant.

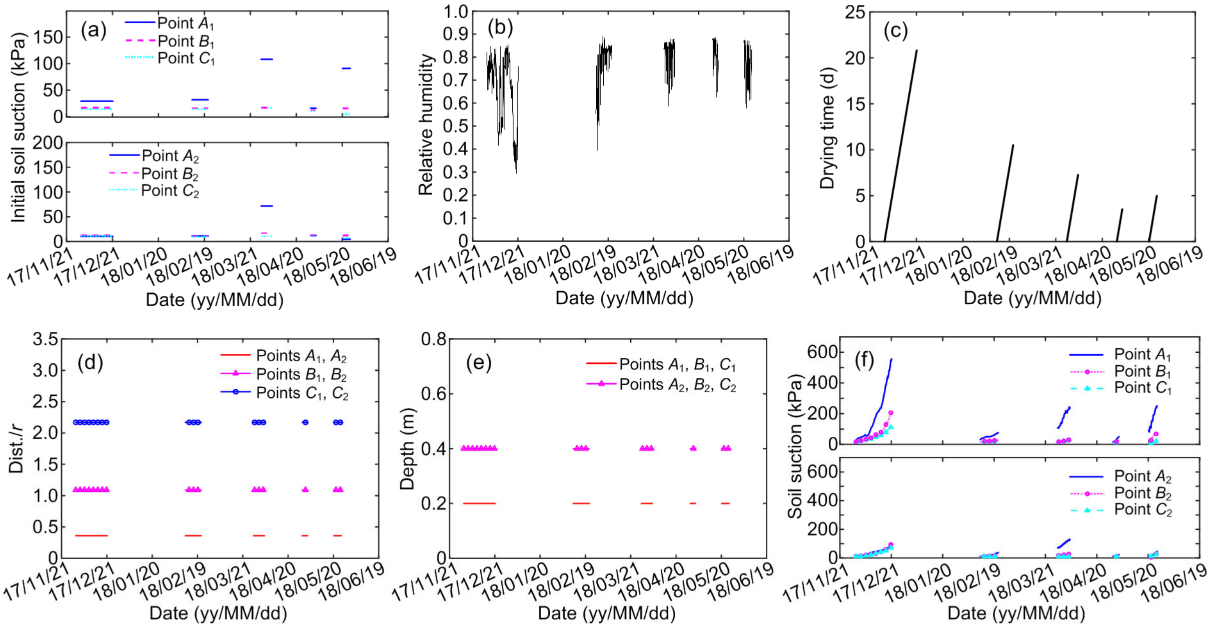

Fig. 6 Measurements of input-output variables: (a) initial soil suction (x_1); (b) air relative humidity (x_2); (c) drying time (x_3); (d) dist./r (x_4); (e) depth (x_5); (f) soil suction in drying cycles (y)

4.2 Parameter settings for GP

Among the total 13 620 data values, the data of previous drying cycles of three time periods (the number of the data is 11 148) was used as the training data for model development, and the remained data of drying cycles of two time periods (the number of the data is 2472) was used as the testing data for testing the robustness of the obtained model. A toolbox GPTIPS (Searson et al., 2010) was employed for the model development based on user's settings of GP parameter by a series of GP operations. At a certain range, more complicated parameter settings of GP usually provide a model with smaller error, but the obtained model is usually complicated, which is not helpful for analysis on relations between variables or importance of input on output. Accordingly, in this study, authors tried to find an optimal model with a simplified structure based on simplified parameter settings for GP. The parameter settings of the GP method for model development are shown in Table 2 (Cheng et al., 2020). The probability rates of three basic GP operations, crossover, mutation, and reproduction, were set for obtaining an optimal model efficiently (Ackora-Prah et al., 2015; Mehr and Nourani, 2018). Relevant GP parameter settings are based on the objective of finding an optimal model with acceptable performance and simplified model structure by trial-and-error method.

Table 2 Parameter settings of the GP method for model development (Cheng et al., 2020)

Parameter	Description
Number of runs	20
Population size	1000
Number of generations	300
Max genes	5
Function set	Times, minus, plus, division
Terminal set	[−100, 100]
Reproduction probability rate	0.05
Crossover probability rate	0.85
Mutation probability rate	0.10

5 Results and performance analysis

5.1 Obtained GP model

Based on the prepared data and parameter settings, a simplified statistical GP model is obtained as follows:

$$y=\frac{1}{x_5}\left[2.9\times10^{-6}x_1(x_3+61.13)\left(\frac{x_3}{x_4}+\frac{290.32}{x_2}\right)\right]+9.15, \tag{11}$$

where y indicates soil suction (kPa), x_1 indicates initial soil suction (kPa), x_2 indicates relative humidity, x_3 indicates drying time (h), x_4 indicates dist./r, and x_5 indicates depth (m). As shown in Fig. 7, the actual measured values of soil suction in drying cycles are collected from the field monitoring, and the estimated values are calculated using the proposed computational model (Eq. (11)). It is obvious that the proposed GP model has captured the field-monitored soil suction variations in drying cycles of six monitoring points. The statistical model can provide a reasonable and reliable estimation of soil suction values.

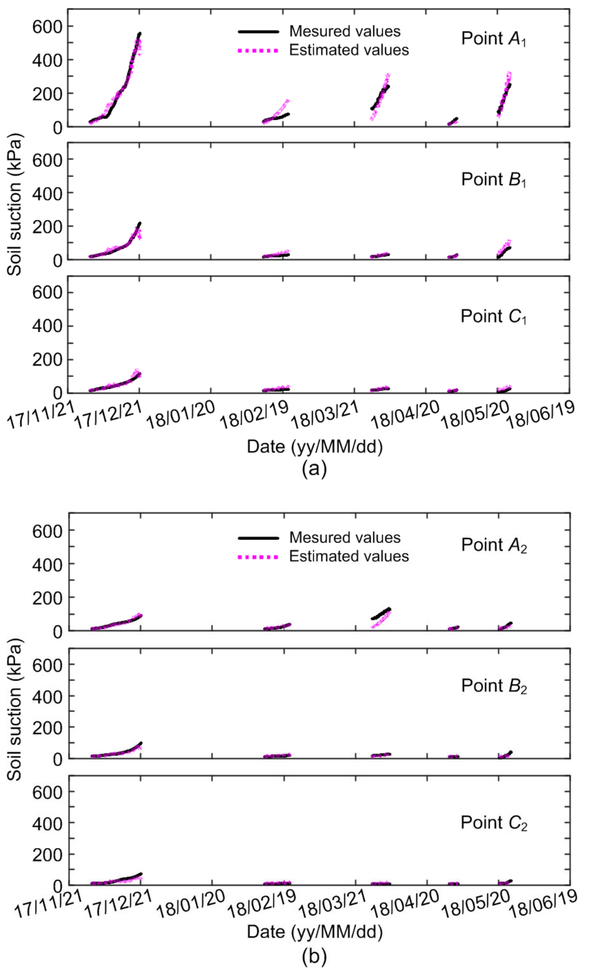

Fig. 7 Comparison between measured and estimated values of soil suction in drying cycles at six monitoring points: (a) depth of 0.2 m; (b) depth of 0.4 m

5.2 Performance analysis

The results of performance evaluation for the obtained drying-cycle model on the training and testing data are illustrated in Fig. 8. Three accuracy metrics are also applied to the performance evaluation of the obtained GP model, i.e. MAPE, relative error, and R^2:

$$\mathrm{MAPE}=\frac{1}{n}\sum_{i=1}^{n}\left|\frac{A_i-P_i}{A_i}\right|\times 100\%, \tag{12}$$

$$\text{Relative error}=\frac{|P_i-A_i|}{A_i}\times 100\%, \tag{13}$$

$$R^2=\left[\frac{\sum_{i=1}^{n}(A_i-\bar{A}_i)(P_i-\bar{P}_i)}{\sqrt{\sum_{i=1}^{n}(A_i-\bar{A}_i)^2\sum_{i=1}^{n}(P_i-\bar{P}_i)^2}}\right]^2, \tag{14}$$

where A_i and P_i indicate the measured and estimated values, respectively, $\bar{A}_i$ and $\bar{P}_i$ indicate the corresponding mean values of A_i and P_i, and n indicates the total number of training or testing data points.

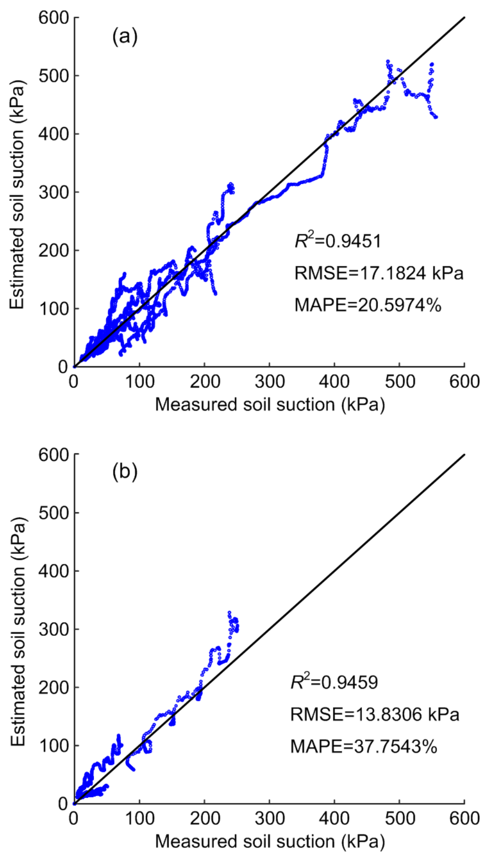

Fig. 8 Performance evaluation of the obtained computational model: (a) training data; (b) testing data

As shown in Fig. 8, the closer the data point is to the diagonal line, the better the performance of GP model is. The coefficients of determination on training and testing data are 0.9451 and 0.9459, respectively, which means that the selected independent variables can provide a good estimation of soil suction variation. The calculation results of RMSE and MAPE indicate that the obtained GP model (Eq. (11)) can be used for estimating the soil suction variations collected from an in-situ test with acceptable errors.

The analysis of relative error for training and testing data of six monitoring points is shown in Table 3. The relative errors on training and testing data of six monitoring points range from 0.004% to 112.463% and from 0.008% to 258.646%, respectively. However, the average values of relative errors on training and testing data of six monitoring points are not larger than 35.454%, which demonstrates that the obtained GP model has grasped the fundamental variation regulation of soil suction in drying cycles with acceptable errors.

Table 3 Relative error analysis

Point	Index	Relative error (%)		
		Min.	Max.	Mean
A_1	Training data	0.023	112.463	27.247
	Testing data	0.008	42.718	18.240
B_1	Training data	0.005	80.290	19.696
	Testing data	7.142	258.646	35.454
C_1	Training data	0.014	99.869	18.248
	Testing data	0.124	183.276	31.910
A_2	Training data	0.014	71.881	15.678
	Testing data	0.015	67.619	20.590
B_2	Training data	0.004	46.502	15.007
	Testing data	0.082	161.518	34.608
C_2	Training data	0.005	65.719	27.776
	Testing data	0.062	49.500	16.276

5.3 Sensitivity analysis

The global sensitivity analysis using aforementioned two approaches was performed to analyze the importance of each independent variable to the dependent variable. A program toolbox (Cannavó, 2012) was employed for calculation of the relevant sensitivity indices. From Fig. 9, it is obvious that the FAST method gives almost the same results with the Sobol' sensitivity method (Chan et al., 2000). For FAST method, the importance of five input variables on the output variable for the drying-cycle model is ranked as follows: drying time (x_3)>initial soil suction (x_1)> dist./r (x_4)>relative humidity (x_2)>depth (x_5), while for the Sobol' method, the rank is drying time (x_3)>initial soil suction (x_1)>relative humidity (x_2)> depth (x_5)>dist./r (x_4). The analysis results confirmed the drying time as the most important independent variables for estimation and prediction of the soil suction values in drying cycles based on the obtained model using measured data in this field monitoring test. According to comparative analysis, for soil suction variations, the initial soil suction is the second most important parameter. Although the sensitivities of the other three input variables, relative humidity (x_2), dist./r (x_4), and depth (x_5), are different from two different methods, the values of sensitivity indices are approximately 4%. These three input variables give a similar contribution to the variance of the output variable, which is also nonnegligible. Especially, the results of the sensitivity analysis are only reasonable for the formula developed by GP based on user's settings of variables in this study. It is obvious that the soil suction variation in drying cycles is dependent on time. The initial suction is quite different between different monitoring points, which causes an important influence on soil suction variation. In fact, the

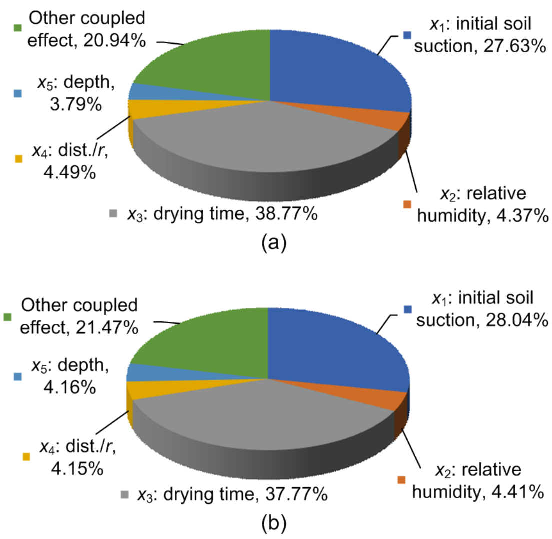

Fig. 9 Global sensitivity analysis with the FAST and Sobol' sensitivity methods for the obtained computational model: (a) FAST method; (b) Sobol' sensitivity method

larger the difference between different initial suction is, the greater the effect of initial suction is, which is dependent on the selection of starting time of each drying cycle, and the effect of initial suction is closely related to the plant-related factors.

5.4 Spatiotemporal prediction of soil suction

The proposed GP model in this study formulates the relations between soil suction and selected five input variables that include the drying time (x_3) and the depth from the ground surface (x_5), and therefore it is able to estimate and predict soil suction changes with time in a certain spatial range around a tree under natural environmental conditions. In order to validate the feasibility and reasonability of the application of the proposed GP model for estimation and prediction of the spatial distribution of soil suction in different drying times, a case study is conducted based on a set of assumed values for input variables for the obtained mathematical GP model. Table 4 shows the assumed settings of five input variables in this case study. Total 20 monitoring points at different distances from the tree, 0.2 m, 0.5 m, 1.5 m, 3.0 m, and 5.0 m (the corresponding ratios of dist./r are 0.14, 0.36, 1.09, 2.17, and 3.62) and at different depths, 0.1 m, 0.2 m, 0.4 m, and 0.6 m, are considered in the case study. The initial value of soil suction in each drying cycle is set as 10 kPa, and the relative humidity varies from 0.822 to 0.791 in assumed 30-d drying period. Based on the assumed settings of five input variables, the spatiotemporal variations of soil suction in drying cycles with time (30-d drying period) at different spatial locations around a tree can be estimated and predicted using the obtained GP model, as shown in Fig. 10. The linear interpolation method is used for obtaining a continuous variation of soil suction at different spatial locations. It is obvious that for the same time, the closer the distance from the tree is in the horizontal and vertical directions, more visible increase the estimated/predicted soil suction has.

Table 4 Assumed settings of input variables

Input variable	Assumed value
x_1 (kPa)	10
x_2	0.822–0.791
x_3 (d)	0–30
x_4	0.14, 0.36, 1.09, 2.17, 3.62
x_5 (m)	0.1, 0.2, 0.4, 0.6

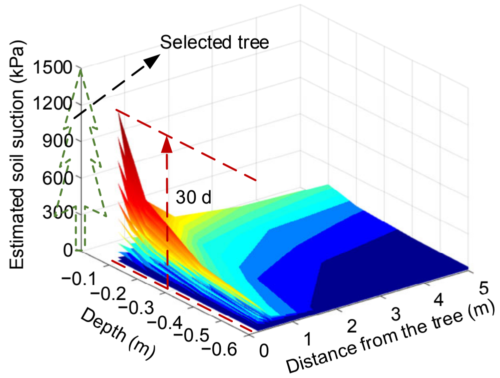

Fig. 10 Estimated soil suction using the obtained model based on a set of input values

Fig. 10 includes 11 three-dimensional curved surfaces of different drying times (0–30 d, the interval is 3 d) obtained using the GP model based on aforementioned set of input variables. Among them, the cross-sections of four different drying times (3 d, 12 d, 21 d, and 30 d) are selected for a better analysis of spatiotemporal variations of soil suction around a tree (Fig. 11). The soil suction increases with drying time, and the soil suction is larger than 1200 kPa after 30 d. It can be seen that the dividing lines between different colors used for different-level soil suction value bulge towards the tree and ground surface, which indicates that if the location is closer to the tree or shallower, the soil suction becomes larger under the same other conditions, while at the locations away from the tree or at deeper locations, the soil suction value is smaller. In drying cycles, the soil suction is mainly affected by the tree root water uptake and the evaporation, and the calculated results show a coincidence with existed studies related to the relationship between root water uptake and depth (Landsberg, 1999; Sun et al., 2011; Guo et al., 2016).

If it is assumed that the effect of the tree on soil suction at same distances in the horizontal direction of 360° is the same and the soil physical parameters at a certain site around the tree is also the same, the spatial distribution of soil suction around a tree can be obtained. As shown in Fig. 12, the spatial distributions of estimated soil suction in four selected drying times (3 d, 12 d, 21 d, and 30 d) are given. In addition, based on the obtained GP model, the soil suction at any location and in different drying times in principle can be estimated. However, the obtained GP model is

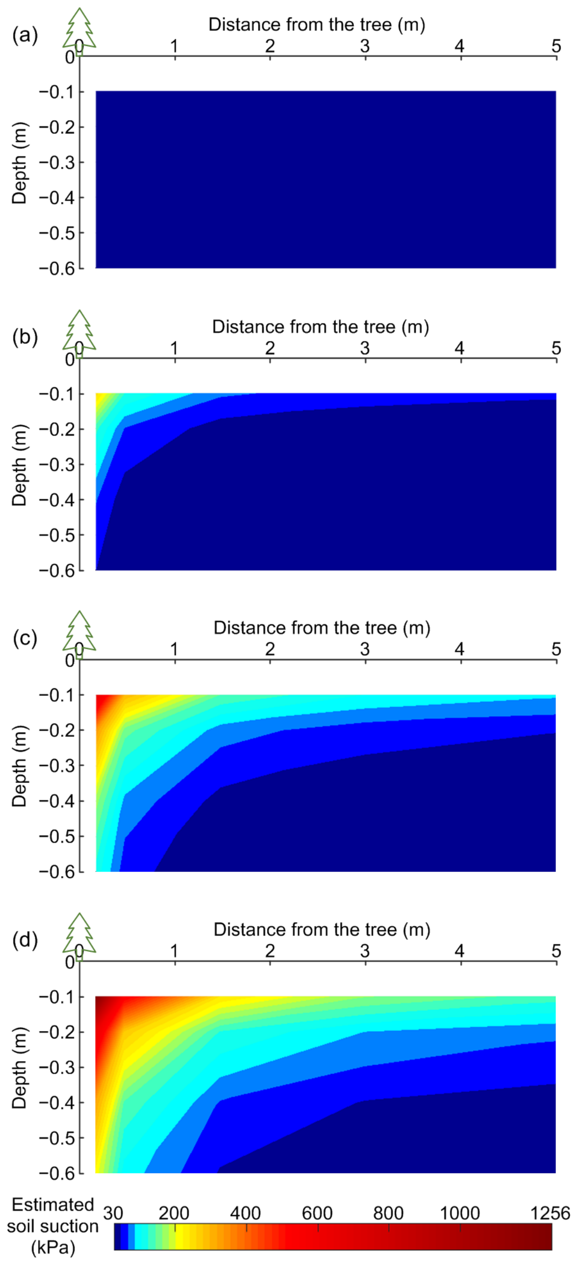

Fig. 11 Cross-section for estimated soil suction in different drying times: (a) 3 d; (b) 12 d; (c) 21 d; (d) 30 d

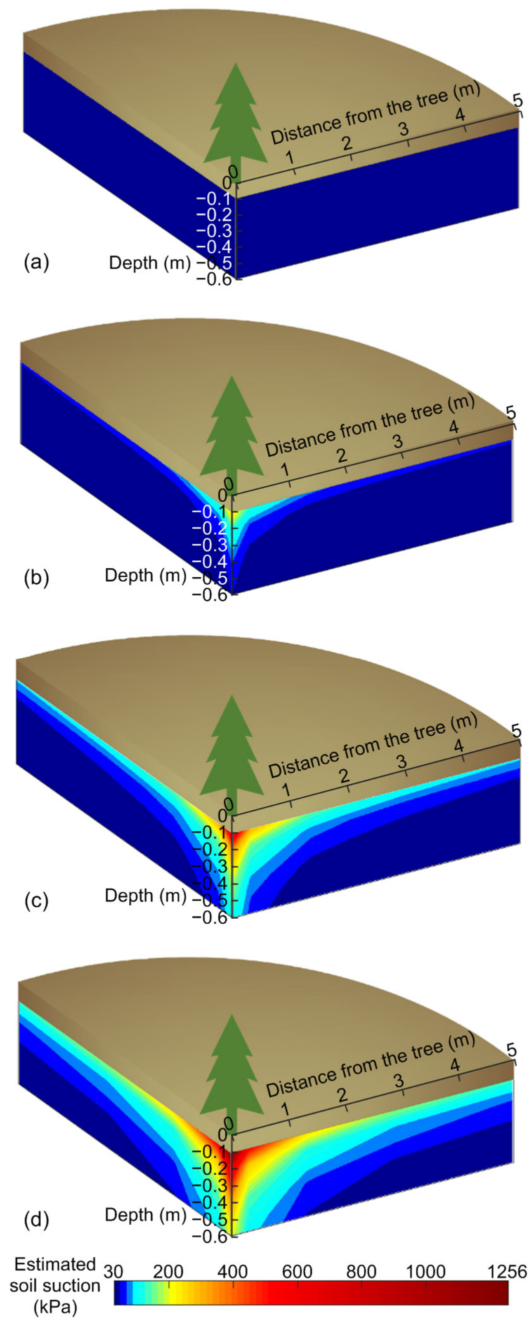

Fig. 12 Three-dimesional profiles of estimated soil suction in different drying times: (a) 3 d; (b) 12 d; (c) 21 d; (d) 30 d

developed using the data of six monitoring points (at different distances from the tree, 0.5 m, 1.5 m, and 3.0 m (the corresponding ratio of dist./r is 0.36, 1.09, and 2.17), and at different depths, 0.2 m and 0.4 m). In this monitoring range, the estimated and predicted values of soil suction are more credible. For the estimated values outside that range, it is necessary to conduct a validation. Besides, the effect range of the tree needs to be considered. Accordingly, if there is a need of better estimation/prediction of soil suction in a larger range, the GP model needs to be improved by model training using more data collected from different locations and at different times, and even several models need to be developed and integrated for estimation/prediction of soil suction with regard to more complicated environmental conditions.

6 Conclusions

Based on the data collected from a field monitoring test from 21st Nov. 2017 to 11th June 2018 in Macau, a simplified statistical model is developed by a machine learning method, GP, for estimating the spatiotemporal variations of soil suction in drying cycles around a tree using selected five influential parameters. The capability of the proposed GP model for estimating soil suction was thoroughly analyzed by different accuracy metrics. The global sensitivity analysis methods are proposed to analyze the importance of each input variable on output variable. Besides, the reliability and feasibility of the obtained GP model in estimation and prediction of spatiotemporal variation of soil suction in drying cycles are validated by a case study based on a set of assumed input variable values. There are some meaningful observations in this study which are presented as follows:

1. The machine learning method GP can efficiently generate an optimal computational model based on user's settings of parameter for GP and the variables. Comparison with some other machine learning methods, the GP method can generate an explicit mathematic formula expressing the relationships among different variables. Four accuracy metrics, i.e. RMSE, MAPE, R^2, and relative error, indicated a good coincidence of measured values and estimated values of soil suction.

2. The variance-based global sensitivity analysis results depicted the drying time as the most important variable and the initial soil suction as the second most important variable. The other three variables, i.e. the relative humidity, the ratio of the distance from the tree to the radius of the tree canopy, and the depth, also give the certain effects on soil suction variations, which cannot be ignored. However, the results of the sensitivity analysis are only reasonable for the formula developed by GP based on user's settings of variables in this study. Especially, the initial suction is quite different between different monitoring points, which causes an important influence on soil suction variation. If the difference among different initial suction is larger, the effect of initial suction is greater, which is dependent on the selection of the starting time of each drying cycle.

3. Based on the results of a case study, it is validated that the closer to the tree or the shallower the location is, the soil suction becomes larger under the same other conditions, while at the locations away from the tree or at deeper locations, the soil suction value is smaller. The results indicate the plant can speed up the increase of soil suction. This study has a limitation on the effect of vegetation and weather factors on soil suction, because less vegetation and weather parameters are considered in model development. It can be concluded that the proposed statistical model developed by a machine learning method, GP, possesses the ability and feasibility of predicting the spatiotemporal variations of soil suction in rooted soil.

This study encourages the application of machine learning approach in property assessment of green geotechnical infrastructure. The proposed GP model indicates a potential to be used as a helpful tool for an assessment of soil parameter variations in rooted soil. In future, more work should be done on how to make a more comprehensive and accurate analysis of vegetation and atmosphere factors on field-monitored soil suction variations using machine learning methods by considering more parameters that are related to vegetation/atmosphere and how to make a reasonable parameter optimization.

Contributors

Wan-huan ZHOU designed the research. Zhi-liang CHENG processed the corresponding data. Zhi-liang CHENG wrote the first draft of the manuscript. Wan-huan ZHOU, Zhi DING, and Yong-xing GUO helped to organize the manuscript. Wan-huan ZHOU and Zhi-liang CHENG revised and edited the final version.

Conflict of interest

Zhi-liang CHENG, Wan-huan ZHOU, Zhi DING, and Yong-xing GUO declare that they have no conflict of interest.

References

Ackora-Prah J, Oheneba-Osei FN, Andam PS, et al., 2015. A multigene genetic programming model for thyroid disorder detection. *Applied Mathematical Sciences*, 9(135): 6707-6722.
https://doi.org/10.12988/ams.2015.59563

Ahmad S, Kalra A, Stephen H, 2010. Estimating soil moisture using remote sensing data: a machine learning approach. *Advances in Water Resources*, 33(1):69-80.
https://doi.org/10.1016/j.advwatres.2009.10.008

Alavi AH, Gandomi AH, 2011. A robust data mining approach for formulation of geotechnical engineering systems. *Engineering Computations*, 28(3):242-274. https://doi.org/10.1108/02644401111118132

Alemdag S, Gurocak Z, Cevik A, et al., 2016. Modeling deformation modulus of a stratified sedimentary rock mass using neural network, fuzzy inference and genetic programming. *Engineering Geology*, 203:70-82. https://doi.org/10.1016/j.enggeo.2015.12.002

An N, Hemmati S, Cui YJ, 2017. Numerical analysis of soil volumetric water content and temperature variations in an embankment due to soil-atmosphere interaction. *Computers and Geotechnics*, 83:40-51. https://doi.org/10.1016/j.compgeo.2016.10.010

ASTM (American Society for Testing Material), 2011. Standard Practice for Classification of Soils for Engineering Purposes (Unified Soil Classification System), D2487-11. National Standards of USA.

Bishop CM, 2006. Pattern Recognition and Machine Learning. Springer, New York, USA.

Brungard CW, Boettinger JL, Duniway MC, et al., 2015. Machine learning for predicting soil classes in three semi-arid landscapes. *Geoderma*, 239-240:68-83. https://doi.org/10.1016/j.geoderma.2014.09.019

Cannavó F, 2012. Sensitivity analysis for volcanic source modeling quality assessment and model selection. *Computers & Geosciences*, 44:52-59. https://doi.org/10.1016/j.cageo.2012.03.008

Chan K, Tarantola S, Saltelli A, et al., 2000. Variance based methods. *In*: Saltelli A, Chan K, Scott EM (Eds.), Sensitivity Analysis. John Wiley & Sons, Chichester, UK, p.167-197.

Cheng ZL, Zhou WH, Garg A, 2020. Genetic programming model for estimating soil suction in shallow soil layers in the vicinity of a tree. *Engineering Geology*, 268:105506. https://doi.org/10.1016/j.enggeo.2020.105506

Cramer NL, 1985. A representation for the adaptive generation of simple sequential programs. Proceedings of the 1st International Conference on Genetic Algorithms and Their Applications, p.183-187.

Crawford MM, Bryson LS, Woolery EW, et al., 2019. Long-term landslide monitoring using soil-water relationships and electrical data to estimate suction stress. *Engineering Geology*, 251:146-157. https://doi.org/10.1016/j.enggeo.2019.02.015

Cui YJ, Gao YB, Ferber V, 2010. Simulating the water content and temperature changes in an experimental embankment using meteorological data. *Engineering Geology*, 114(3-4):456-471. https://doi.org/10.1016/j.enggeo.2010.06.006

Cukier RI, Fortuin CM, Shuler KE, 1973. Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. I theory. *The Journal of Chemical Physics*, 59(8):3873-3878. https://doi.org/10.1063/1.1680571

Dai H, Ye M, 2015. Variance-based global sensitivity analysis for multiple scenarios and models with implementation using sparse grid collocation. *Journal of Hydrology*, 528:286-300. https://doi.org/10.1016/j.jhydrol.2015.06.034

Feng S, Liu HW, Ng CWW, 2019. Analytical solutions for one-dimensional water flow in vegetated layered soil. *International Journal of Geomechanics*, 19(2):04018191. https://doi.org/10.1061/(ASCE)GM.1943-5622.0001343

Feng Y, Cui NB, Hao WP, et al., 2019. Estimation of soil temperature from meteorological data using different machine learning models. *Geoderma*, 338:67-77. https://doi.org/10.1016/j.geoderma.2018.11.044

Fredlund DG, Xing AQ, 1994. Equations for the soil-water characteristic curve. *Canadian Geotechnical Journal*, 31(4):521-532. https://doi.org/10.1139/t94-061

Fredlund DG, Sheng DC, Zhao JD, 2011. Estimation of soil suction from the soil-water characteristic curve. *Canadian Geotechnical Journal*, 48(2):186-198. https://doi.org/10.1139/T10-060

Fredlund MD, Wilson GW, Fredlund DG, 2002. Use of the grain-size distribution for estimation of the soil-water characteristic curve. *Canadian Geotechnical Journal*, 39(5):1103-1117. https://doi.org/10.1139/t02-049

Gadi V, Singh S, Singhariya M, 2018. Modeling soil-plant-water interaction: effects of canopy and root parameters on soil suction and stability of green infrastructure. *Engineering Computations*, 35(3):1543-1566. https://doi.org/10.1108/EC-07-2017-0280

Gadi VK, Hussain R, Bordoloi S, et al., 2019. Relating stomatal conductance and surface area with evapotranspiration induced suction in a heterogeneous grass cover. *Journal of Hydrology*, 568:867-876. https://doi.org/10.1016/j.jhydrol.2018.11.048

Gamse S, Zhou WH, Tan F, et al., 2018. Hydrostatic-season-time model updating using Bayesian model class selection. *Reliability Engineering & System Safety*, 169:40-50. https://doi.org/10.1016/j.ress.2017.07.018

Gandomi AH, Alavi AH, 2012. A new multi-gene genetic programming approach to non-linear system modeling. Part II: geotechnical and earthquake engineering problems. *Neural Computing and Applications*, 21(1):189-201. https://doi.org/10.1007/s00521-011-0735-y

Garg A, Garg A, Zhou WH, et al., 2015. A new simulation approach of genetic programming in modelling of soil water retention property of unsaturated soil. *Engineering Computations*, 32(3):914-930. https://doi.org/10.1108/EC-05-2014-0110

Garg A, Bordoloi S, Ni JJ, et al., 2019a. Influence of biochar addition on gas permeability in unsaturated soil. *Géotechnique Letters*, 9(1):66-71. https://doi.org/10.1680/jgele.18.00190

Garg A, Hazra B, Zhu H, et al., 2019b. A simplified probabilistic analysis of water content and wilting in soil vegetated with non-crop species. *CATENA*, 175:123-131. https://doi.org/10.1016/j.catena.2018.12.016

Gopal P, Bordoloi S, Ratnam R, et al., 2019. Investigation of infiltration rate for soil-biochar composites of water hyacinth. *Acta Geophysica*, 67(1):231-246. https://doi.org/10.1007/s11600-018-0237-8

Guo F, Ma JJ, Zheng LJ, et al., 2016. Estimating distribution of water uptake with depth of winter wheat by hydrogen and oxygen stable isotopes under different irrigation depths. *Journal of Integrative Agriculture*, 15(4):891-906. https://doi.org/10.1016/S2095-3119(15)61258-8

He LX, Liu Y, Bi SF, et al., 2019. Estimation of failure probability in braced excavation using Bayesian networks with integrated model updating. *Underground Space*, in press. https://doi.org/10.1016/j.undsp.2019.07.001

Hemmati S, Gatmiri B, Cui YJ, et al., 2012. Thermo-hydro-mechanical modelling of soil settlements induced by soil-vegetation-atmosphere interactions. *Engineering Geology*, 139-140:1-16. https://doi.org/10.1016/j.enggeo.2012.04.003

Hossain MA, Yin JH, 2010. Behavior of a compacted completely decomposed granite soil from suction controlled direct shear tests. *Journal of Geotechnical and Geoenvironmental Engineering*, 136(1):189-198. https://doi.org/10.1061/(ASCE)GT.1943-5606.0000189

Javadi AA, Rezania M, Nezhad MM, 2006. Evaluation of liquefaction induced lateral displacements using genetic programming. *Computers and Geotechnics*, 33(4-5):222-233. https://doi.org/10.1016/j.compgeo.2006.05.001

Jin YF, Yin ZY, Zhou WH, et al., 2019a. Multi-objective optimization-based updating of predictions during excavation. *Engineering Applications of Artificial Intelligence*, 78:102-123. https://doi.org/10.1016/j.engappai.2018.11.002

Jin YF, Yin ZY, Zhou WH, et al., 2019b. A single-objective EPR based model for creep index of soft clays considering L_2 regularization. *Engineering Geology*, 248:242-255. https://doi.org/10.1016/j.enggeo.2018.12.006

Johari A, Habibagahi G, Ghahramani A, 2006. Prediction of soil-water characteristic curve using genetic programming. *Journal of Geotechnical and Geoenvironmental Engineering*, 132(5):661-665. https://doi.org/10.1061/(asce)1090-0241(2006)132:5(661)

Kalnins A, 2018. Multicollinearity: how common factors cause Type 1 errors in multivariate regression. *Strategic Management*, 39(8):2362-2385. https://doi.org/10.1002/smj.2783

Karandish F, Šimůnek J, 2016. A comparison of numerical and machine-learning modeling of soil water content with limited input data. *Journal of Hydrology*, 543:892-909. https://doi.org/10.1016/j.jhydrol.2016.11.007

Kim J, Jeong S, Park S, et al., 2004. Influence of rainfall-induced wetting on the stability of slopes in weathered soils. *Engineering Geology*, 75(3-4):251-262. https://doi.org/10.1016/j.enggeo.2004.06.017

Kisi O, Dailr AH, Cimen M, et al., 2012. Suspended sediment modeling using genetic programming and soft computing techniques. *Journal of Hydrology*, 450-451:48-58. https://doi.org/10.1016/j.jhydrol.2012.05.031

Kucherenko S, Shah N, 2007. The importance of being global. Application of global sensitivity analysis in Monte Carlo option pricing. *Wilmott Magazine*, 2007: 2-10.

Landsberg J, 1999. The Ways Trees Use Water. Water and Salinity Issues in Agroforestry. Rural Industries Research and Development Corporation, Canberra, Australia, p.1-92.

Lary DJ, Alavi AH, Gandomi AH, et al., 2016. Machine learning in geosciences and remote sensing. *Geoscience Frontiers*, 7(1):3-10. https://doi.org/10.1016/j.gsf.2015.07.003

Lee LM, Gofar N, Rahardjo H, 2009. A simple model for preliminary evaluation of rainfall-induced slope instability. *Engineering Geology*, 108(3-4):272-285. https://doi.org/10.1016/j.enggeo.2009.06.011

Leung AK, Garg A, Ng CWW, 2015. Effects of plant roots on soil-water retention and induced suction in vegetated soil. *Engineering Geology*, 193:183-197. https://doi.org/10.1016/j.enggeo.2015.04.017

Makkeasorn A, Chang NB, Beaman M, et al., 2006. Soil moisture estimation in a semiarid watershed using RADARSAT-1 satellite imagery and genetic programming. *Water Resources Research*, 42(9):W09401. https://doi.org/10.1029/2005WR004033

Mehr AD, Nourani V, 2018. Season algorithm-multigene genetic programming: a new approach for rainfall-runoff modelling. *Water Resources Management*, 32(8):2665-2679. https://doi.org/10.1007/s11269-018-1951-3

Ni JJ, Leung AK, Ng CWW, et al., 2018. Modelling hydro-mechanical reinforcements of plants to slope stability. *Computers and Geotechnics*, 95:99-109. https://doi.org/10.1016/j.compgeo.2017.09.001

Nowamooz H, Jahangir E, Masrouri F, et al., 2016. Effective stress in swelling soils during wetting drying cycles. *Engineering Geology*, 210:33-44. https://doi.org/10.1016/j.enggeo.2016.05.021

Parasuraman K, Elshorbagy A, Si BC, 2007. Estimating saturated hydraulic conductivity using genetic programming. *Soil Science Society of America Journal*, 71(6):1676-1684. https://doi.org/10.2136/sssaj2006.0396

Pétrowski A, Ben-Hamida S, 2017. Genetic programming for machine learning. *In*: Evolutionary Algorithms, Volume 9. Wiley, USA, p.183-216. https://doi.org/10.1002/9781119136378.ch6

Pham BT, Son LH, Hoang TA, et al., 2018. Prediction of shear strength of soft soil using machine learning methods. *CATENA*, 166:181-191. https://doi.org/10.1016/j.catena.2018.04.004

Poli R, Langdon WB, McPhee NF, 2008. A Field Guide to Genetic Programming. Lulu, Raleigh, USA.

Prasad R, 1988. A linear root water uptake model. *Journal of Hydrology*, 99(3-4):297-306. https://doi.org/10.1016/0022-1694(88)90055-8

Qi XH, Zhou WH, 2017. An efficient probabilistic back-analysis method for braced excavations using wall deflection data at multiple points. *Computers and Geotechnics*, 85:186-198. https://doi.org/10.1016/j.compgeo.2016.12.032

Qi XH, Zhou WH, Yuen KV, 2017. Detection of stationary Markovian zones in a geologically heterogeneous area. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, 3(4): 04017026. https://doi.org/10.1061/AJRUA6.0000930

Rezania M, Javadi AA, 2007. A new genetic programming model for predicting settlement of shallow foundations. *Canadian Geotechnical Journal*, 44(12):1462-1473. https://doi.org/10.1139/T07-063

Roushangar K, Akhgar S, Salmasi F, et al., 2014. Modeling energy dissipation over stepped spillways using machine learning approaches. *Journal of Hydrology*, 508:254-265. https://doi.org/10.1016/j.jhydrol.2013.10.053

Saltelli A, Sobol' IM, 1995. About the use of rank transformation in sensitivity analysis of model output. *Reliability Engineering & System Safety*, 50(3):225-239. https://doi.org/10.1016/0951-8320(95)00099-2

Saltelli A, Annoni P, Azzini I, et al., 2010. Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer Physics Communications*, 181(2):259-270. https://doi.org/10.1016/j.cpc.2009.09.018

Samui P, Sitharam TG, 2011. Machine learning modelling for predicting soil liquefaction susceptibility. *Natural Hazards and Earth System Sciences*, 11(1):1-9. https://doi.org/10.5194/nhess-11-1-2011

Searson DP, Leahy DE, Willis MJ, 2010. GPTIPS: an open source genetic programming toolbox for multigene symbolic regression. Proceedings of the International MultiConference of Engineers and Computer Scientists, p.77-80.

Shahin MA, 2015. Genetic programming for modelling of geotechnical engineering systems. *In*: Gandomi AH, Alavi AH, Ryan C (Eds.), Handbook of Genetic Programming Applications. Springer, Cham, Germany. https://doi.org/10.1007/978-3-319-20883-1_2

Sobol' IM, 1990. On sensitivity estimation for nonlinear mathematical models. *Matematicheskoe Modelirovanie*, 2(1):112-118.

Sobol' IM, 2001. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation*, 55(1-3):271-280. https://doi.org/10.1016/S0378-4754(00)00270-6

Srivastava PK, Han DW, Ramirez MR, et al., 2013. Machine learning techniques for downscaling SMOS satellite soil moisture using MODIS land surface temperature for hydrological application. *Water Resources Management*, 27(8):3127-3144. https://doi.org/10.1007/s11269-013-0337-9

Sun SJ, Meng P, Zhang JS, et al., 2011. Variation in soil water uptake and its effect on plant water status in *Juglans regia* L. during dry and wet seasons. *Tree Physiology*, 31(12): 1378-1389. https://doi.org/10.1093/treephys/tpr116

Tan F, Zhou WH, Yuen KV, 2016. Modeling the soil water retention properties of same-textured soils with different initial void ratios. *Journal of Hydrology*, 542:731-743. https://doi.org/10.1016/j.jhydrol.2016.09.045

Tan F, Zhou WH, Yuen KV, 2018. Effect of loading duration on uncertainty in creep analysis of clay. *International Journal for Numerical and Analytical Methods in Geomechanics*, 42(11):1235-1254. https://doi.org/10.1002/nag.2788

Too VK, Omuto CT, Biamah EK, et al., 2014. Review of soil water retention characteristic (SWRC) models between saturation and oven dryness. *Open Journal of Modern Hydrology*, 4(4):173-182. https://doi.org/10.4236/ojmh.2014.44017

Totoev YZ, Kleeman PW, 1998. An infiltration model to predict suction changes in the soil profile. *Water Resources Research*, 34(7):1617-1622. https://doi.org/10.1029/98WR00825

Wang H, 2020. Finding patterns in subsurface using Bayesian machine learning approach. *Underground Space*, 5(1):84-92. https://doi.org/10.1016/j.undsp.2018.10.006

Weyl H, 1938. Mean motion. *American Journal of Mathematics*, 60(4):889-896. https://doi.org/10.2307/2371267

Whigham PA, Crapper PF, 2001. Modelling rainfall-runoff using genetic programming. *Mathematical and Computer Modelling*, 33(6-7):707-721. https://doi.org/10.1016/S0895-7177(00)00274-0

Wikipedia, 2019. Tree Crown Measurement. Wikipedia. https://wikimili.com/en/Tree_crown_measurement

Yang SR, Huang WH, Chung SH, 2015. Combined effects of temperature and moisture content on soil suction of compacted bentonite. *Journal of Marine Science and Technology*, 23(3):281-287. https://doi.org/10.6119/JMST-014-0326-2

Yin ZY, Jin YF, Shen SL, et al., 2017. An efficient optimization method for identifying parameters of soft structured clay by an enhanced genetic algorithm and elastic-viscoplastic model. *Acta Geotechnica*, 12(4):849-867.

https://doi.org/10.1007/s11440-016-0486-0
Yin ZY, Jin YF, Shen JS, et al., 2018. Optimization techniques for identifying soil parameters in geotechnical engineering: comparative study and enhancement. *International Journal for Numerical and Analytical Methods in Geomechanics*, 42(1):70-94.
https://doi.org/10.1002/nag.2714
Zhang P, Yin ZY, Jin YF, et al., 2020. A novel hybrid surrogate intelligent model for creep index prediction based on particle swarm optimization and random forest. *Engineering Geology*, 265:105328.
https://doi.org/10.1016/j.enggeo.2019.105328
Zhou WH, Qi XH, 2019. Root cohesion estimation of riparian trees based on model uncertainty characterization. *Journal of Materials in Civil Engineering*, 31(2):04018389.
https://doi.org/10.1061/(ASCE)MT.1943-5533.0002600
Zhou WH, Yuen KV, Tan F, 2013. Estimation of maximum pullout shear stress of grouted soil nails using Bayesian probabilistic approach. *International Journal of Geomechanics*, 13(5):659-664.
https://doi.org/10.1061/(ASCE)GM.1943-5622.0000259
Zhou WH, Garg A, Garg A, 2016. Study of the volumetric water content based on density, suction and initial water content. *Measurement*, 94:531-537.
https://doi.org/10.1016/j.measurement.2016.08.034
Zhou WH, Garg A, Garg A, 2017. Computation of coupled effects of root and shoot characteristics on transpiration based on optimization approach. *Engineering Computations*, 34(3):725-738.
https://doi.org/10.1108/EC-05-2016-0177
Zhou WH, Tan F, Yuen KV, 2018. Model updating and uncertainty analysis for creep behavior of soft soil. *Computers and Geotechnics*, 100:135-143.
https://doi.org/10.1016/j.compgeo.2018.04.006
Zhou WH, He SY, Garg A, et al., 2020. Field monitoring of suction in the vicinity of an urban tree: exploring termite infestation and the shading effects of tree canopy. *Acta Geotechnica*, 15(5):1341-1361.
https://doi.org/10.1007/s11440-019-00810-0
Zhu H, Zhang LM, Garg A, 2018. Investigating plant transpiration-induced soil suction affected by root morphology and root depth. *Computers and Geotechnics*, 103:26-31.
https://doi.org/10.1016/j.compgeo.2018.06.019

中文概要

题　目：基于机器学习算法估算根系土体特性的时空响应

目　的：在绿色岩土工程中，浅层土体特性通常受到当地气候和覆盖植被的影响。本文旨在探讨自然环境条件下不同植物和大气因素（与树的距离、空气湿度和距离地表的深度等）与土体基质吸力的关系，通过一种机器学习方法建立简化的统计模型，并对浅层根系土体中基质吸力的时空变化进行估算和预测。

创新点：1. 通过一种机器学习方法（即遗传编程算法）建立土体基质吸力和五个选定的影响因素之间的关系；2. 根据建立的统计模型，有效地预测了根系土体内基质吸力的时空变化。

方　法：1. 通过现场监测实验（图 3 和 4），量化土体基质吸力和不同影响参数随时间的变化（图 5 和 6）；2. 通过机器学习算法，构建土体基质吸力的时空变化与五个选定的影响参数之间的关系，得到一个简化的统计模型（公式(11)）；3. 通过误差分析，验证该简化统计模型在估算和预测土体基质吸力时空变化时的可靠性；4. 通过敏感性分析研究不同参数对土体基质吸力时空变化的影响（图 9）；5. 通过案例研究，验证利用该方法对根系土体基质吸力时空变化进行预测的可行性（图 11 和 12）。

结　论：1. 遗传编程算法可以有效地建立土体基质吸力和不同影响参数之间的关系，并能给出相应的数学公式以对土体基质吸力的时空变化进行可靠的估算和预测；2. 基于方差的全局敏感性分析方法发现干循环时间和初始基质吸力对土体基质吸力的时空变化有重要影响，而且其他的植物和大气相关参数对土体基质吸力的时空变化也有不可忽视的影响；3. 案例研究结果表明，本文所提方法可用于预测土体基质吸力的时空变化。

关键词：遗传编程；简化的统计模型；时空变化；土体基质吸力