PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

Soil-pipe interaction modeling for pipe behavior prediction with super learning based methods

Fang Shi, Xiang Peng, Huan Liu, Yafei Hu, Zheng Liu, et al.

Fang Shi, Xiang Peng, Huan Liu, Yafei Hu, Zheng Liu, Eric Li, "Soil-pipe interaction modeling for pipe behavior prediction with super learning based methods," Proc. SPIE 10602, Smart Structures and NDE for Industry 4.0, 1060207 (27 March 2018); doi: 10.1117/12.2300812

SPIE.

Event: SPIE Smart Structures and Materials + Nondestructive Evaluation and Health Monitoring, 2018, Denver, Colorado, United States

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 5/8/2018 Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

Soil-pipe interaction modeling for pipe behavior prediction with super learning based methods

Fang Shi[a], Xiang Peng[a], Huan Liu[b], Yafei Hu[c], Zheng Liu[a], and Eric Li[a]

[a]School of Engineering, University of British Columbia Okanagan, Kelowna BC, Canada
[b]School of Automation, China University of Geosciences, Wuhan, China
[c]City of Regina, Regina SK, Canada

ABSTRACT

Underground pipelines are subject to severe distress from the surrounding expansive soil. To investigate the structural response of water mains to varying soil movements, field data, including pipe wall strains in situ soil water content, soil pressure and temperature, was collected. The research on monitoring data analysis has been reported, but the relationship between soil properties and pipe deformation has not been well-interpreted. To characterize the relationship between soil property and pipe deformation, this paper presents a super learning based approach combining feature selection algorithms to predict the water mains structural behavior in different soil environments. Furthermore, automatic variable selection method, e.i. recursive feature elimination algorithm, were used to identify the critical predictors contributing to the pipe deformations. To investigate the adaptability of super learning to different predictive models, this research employed super learning based methods to three different datasets. The predictive performance was evaluated by R-squared, root-mean-square error and mean absolute error. Based on the prediction performance evaluation, the superiority of super learning was validated and demonstrated by predicting three types of pipe deformations accurately. In addition, a comprehensive understand of the water mains working environments becomes possible.

Keywords: Super learning, feature selection, sensor data, expansive soil, pipe deformation, prediction

1. INTRODUCTION

In North America, there are many water distribution networks locating in expansive soils, which has a significant impact on the buried pipe performance.[1] For instance, due to the soil moisture-suction variation, the expansive soil can undergo the extreme shrink-swell movement, which can cause pipe deformation and result in pipe breakage finally.[2] Especially in the case of small diameter pipes (i.e. < 200 mm),[3] pipe failure rates are directly affected by the shrink/swell behavior of soil.[1] Hence, it is critical to perform researches to understand the complex pipe-soil interactions,[4] which can be applied to predict the pipeline stability behavior that is vital in new design and the existing assets evaluation.[5]

Based on existing studies, the soil-pipe interaction was modeled using two methods primarily, the Winkler Spring Approach (WSA) and the Finite Element Analysis (FEA).[3] The WSA was proposed by Winkler in 1876,[6] simulating the pipe-soil interaction by nonlinear soil springs in the axial, lateral, and vertical directions.[7] Although WSA has been advanced to reflect the different physical aspects of the soil-structure interaction, it neglects the impact of rigid soil movements and interaction through soil from location to location.[8] Accordingly, Trickey and his colleagues[8] implemented FEA to observe the three-dimensional response of buried pipes. In 2014, a Gurson-Tvergaard-Needleman model-based FEA methodology was developed to predict the pipeline tensile strain capacity, which was validated against 4 full-scale tests.[9] However, through a parametric study based on the methodology, the developed equation for the strain prediction is very complex and not well understood. Additionally, the data required to model these physical mechanisms are always rarely available and prohibitively costly to acquire.[10] Although a relatively simplified modeling approach, e.i. a three dimensional finite difference continuum model, was used to analyze pipe behavior in the expansive soil,[11] the model was only validated by the

Further author information: (Send correspondence to Zheng Liu)
Zheng Liu: E-mail: zheng.liu@ubc.ca

Smart Structures and NDE for Industry 4.0, edited by Norbert G. Meyendorf, Dan J. Clingman, Proc. of SPIE Vol. 10602, 1060207 · © 2018 SPIE · CCC code: 0277-786X/18/$18 · doi: 10.1117/12.2300812

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 5/8/2018 Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

experiment results from the laboratory, which was still full of uncertainties when applied practically. Therefore, further studies are needed to better understand the field performance of buried pipes influenced by surrounding unsaturated soil conditions.[3]

For this purpose, this research investigated a novel super learning based framework to model the relationship between the expansive soil behavior and the pipe performance. Machine learning techniques with capability of exploring the hidden relationships in a dataset have been widely used in damage detection,[12] soil corrosivity analysis,[13] and pipe deterioration evaluation,[14,15] etc. Recently, field monitoring has been conducted to collect field data through a sensor network,[16] which makes field datasets including soil properties and pipe behavior become available. Feature selection approaches were first applied on the field data to identify the key predictors for the pipe deformation. The impacts of soil properties such as water contents, earth pressure and temperature were also analyzed through the feature selection procedure. The super learning based techniques then were used to predict the pipe performance based on the selected variables. To examine the adaptability and effectiveness of the proposed approach, the framework was applied to predict three types of pipe deformation, including hoop deformation, bending deformation and compression deformation along the longitudinal direction. The proposed methods provide a promising way to simplify soil-interaction modeling process and to accurately predict pipe performance.

The rest of the paper is constructed as follows. The methodology of the proposed super learning framework is presented in Sec. 2, where both the overview of the framework and the detail description of algorithm are provided. In Sec. 3, the analysis of collected field data along with the implementation of super learning based methods are described. The performance evaluation metrics and the experimental results are presented and discussed in Sec. 4. The conclusion based on the experimental results and findings are given in Sec. 5.

2. SUPER LEARNING BASED PREDICTION MODELING

Water mains installed in expansive clay soils often suffer from structural distress. Hence, it is important to identify and model the relationship between soil properties and pipe performance. Figure 1 describes the procedure of prediction modeling using super learning techniques. Three databases were constructed based on the data collected from a sensor network in a data acquisition system, which monitored the soil conditions and pipe deformations in a three-year period. Feature selection approaches were then used to identify the critical factors leading to pipe deformations. Not only did this process make the model simple and easily interpreted, it helped improve the prediction accuracy, as well. Using the selected variable, a group of base learners, e.i. machine learning algorithms, were trained and combined optimally to generate a super learner, which can be used to estimate the pipe performance based on soil properties.

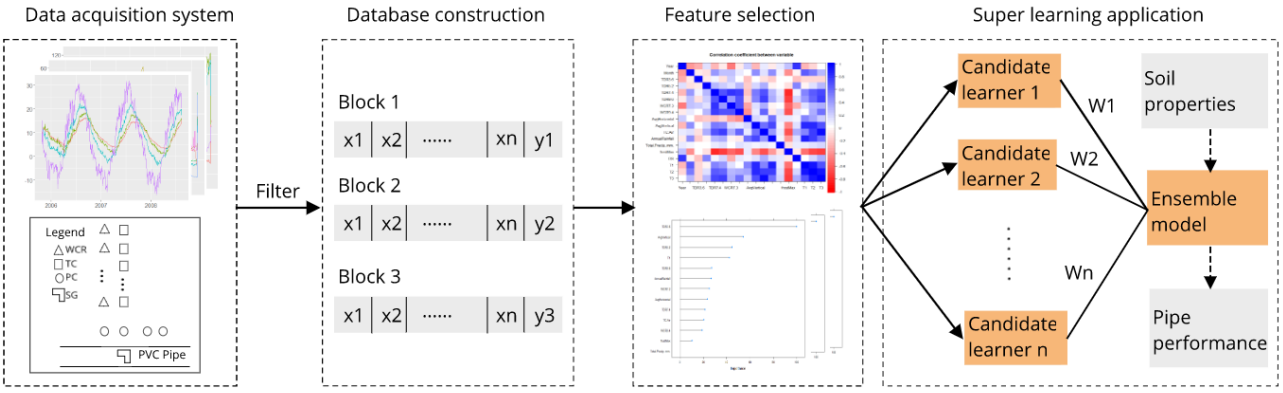


Figure 1. Work flow of the super learning based methodology.

2.1 Feature selection for sensor data

The advantages of variable selection mainly include improving the prediction performance of the predictors, reducing the algorithm training time, and providing a better understanding of the underlying process that

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 5/8/2018 Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

generated the data.[17] In the super learning system, there are two steps to achieve feature selection. The fist step of variable selection is independent from machine learning process, where highly correlated attributes were removed using Pearson correlation coefficient, a measurement for defining a linear relationship between two continuous variables. The formula of Pearson correlation coefficient was given in Eq. (1), where: cov is the covariance; σ_X is the standard deviation of X; σ_Y is the standard deviation of Y. The value of $\rho_{X,Y}$ is always between +1 and 1. The closer its absolute value is to 1, the stronger the relationship between the two variables and a correlation of 0 indicates an absence of the linear relationship.

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} \tag{1}$$

After removing the redundant features, an automatic feature selection method, called Random Forest-Recursive Feature Elimination (RF-RFE), was used to find the variable subset with the best performance. Beginning with using all predictors as inputs, RFE iteratively constructs models by excluding the worst performing feature, which are ranked according to its importance to the model. At each iteration, the variable rankings are recomputed by RF and the final ranking of the predictors are based on the order of their elimination.[18]

2.2 Super learner algorithm

Super learning, known as stacking ensemble, is a technique with ability of finding the optimum way to combine multiple, typically diverse, base learning algorithms to generate a powerful prediction function.[19] Table 1 describes the super learning algorithm. Prior to the construction of super learner, a variety of individual machine learning algorithms were evaluated by k-fold validation. Since super learning approach requires training a number of base learning algorithms, which is computationally intensive, algorithms with less training time were selected as the candidate learners.

Table 1. Super learning algorithm.

Algorithm: Super learning
1. Split data $D = \{x_i, y_i\}_{i=1}^n$ into K blocks for validation: $V_i, i = 1, 2...k$
2. **for** *Each candidate learner* $H_i, i = 1...n$ **do**
3. **for** *Each validation block* $V_i, i = 1, 2...k$ **do**
4. Train candidate learnerH_i on D excluding the validation block V_i
5. Predict the target value y in the validation block V_i based on H_i
6. **end**
7. **end**
8. Construct level-one data $D_1 = \{x_i^{'}, y_i\}_{i=1}^n$, where $x_i^{'} = \{V_1(y)...V_k(y)\}$
9. Train the meta learner M on D_1 to determined the weight of each candidate learner
10. Train each candidate learner $H_i, i = 1...n$ on full dataset D
11. Combine predictions from $H_i, i = 1...n$ with M
12. Return super learner S comprising of $H_i, i = 1...n$ and M models

After the selection of the candidate algorithms, the first step of super learning development is to evaluate the prediction performance of the candidate learners by k-fold cross-validation. For each candidate learner, the predicted results for each validation sets were constructed as variables for meta learner training, where the best weighted combination of the base learners were determined by cross-validated risk, i.e. the average error of cross validation. The trained meta learner is then used to combine the base learners trained with the full data set. The super learner is the combination of these base learners with determined weights.[20]

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 5/8/2018 Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

3. IMPLEMENTATION OF PIPE BEHAVIOR PREDICTION

3.1 Data collection in a field monitoring system

In the City of Regina, the volume change of the native clay deposit is always a problem for the underground water distribution pipes. Hence, a field instrumentation program was conducted to monitor the performance of a PVC pipe with 150mm diameter and its surrounding soil condition in south Regina, Saskatchewan, where a high pipe breakage rate was reported. The monitoring system consists of a group of high quality sensors, measuring pipe wall strain, in situ soil temperature, soil pressure, and soil water contents. The data used in the exploratory research was collected from this system in the period between May 5, 2005 to April 21, 2008.[16] This data set contains 973 records with soil data measured by 23 different sensors as well as pipe deformation information obtained from 8 strain gauge. Table 2 describes the details of the cleaned sensor data used in this study, where unreliable records were removed and then grouped into categories.

Table 2. Summary of the sensor data used in this study.[16]

Sensor data	Sensor type	#Sensors	Description
Soil temperature	Thermocouples (TC)	13	Soil temperature measured in different depth.
Soil water content	Water content reflectometer(WCR)	6	Water contents measured in different depth.
Horizontal soil pressure	Pressure cell (PC)	2	Horizontal stress measured near the pipe.
Vertical soil pressure	Pressure cell (PC)	2	Vertical stress measured near the pipe.
Circumferential strain	Strain gauge (SG)	4	Pipe hoop deformation measured inside.
Longitudinal strain	Strain gauge (SG)	4	Pipe bending deformation and tensile deformation in the longitudinal direction.

3.2 Data analysis to remove redundant variables

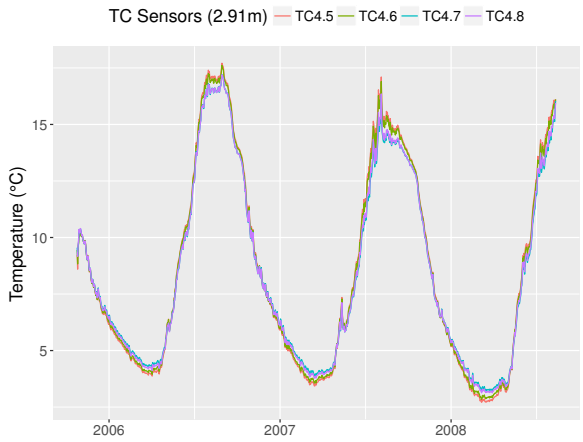
To observe the changes in soil properties vertically and horizontally, multiple sensors were installed to measure the soil temperature and water contents. This resulted in a large number of features, and more importantly, a correlation between these attributes collected by the same type of sensors. Therefore, prior to super learning, data was analyzed and processed to filter the redundant variables and to extract features from the sensor nodes.

3.2.1 Soil properties

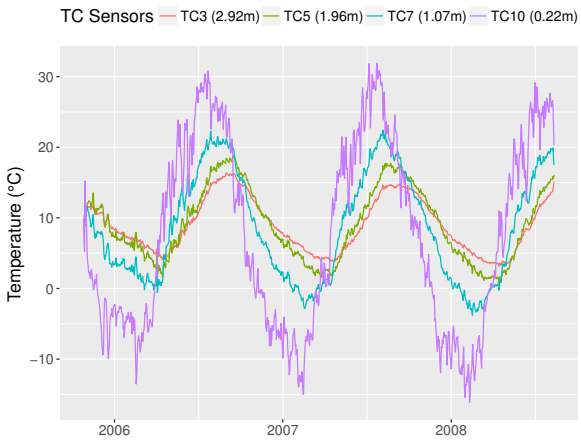
1) Soil temperature: Figures 2a and 2b present the measured temperature varying horizontally and vertically, where a periodic variation was observed. As shown in Fig. 2a, there is almost no difference between the temperature measured horizontally while Figure 2b indicates that the closer to the surface, the larger fluctuation the temperature has. Since Figures 2a and 2b demonstrate a strong relationship between the temperature variables, which can cause the multicollinearity problem, we remove the temperature variables whose Pearson correlation coefficient is greater than 0.9. Finally, TC5 (represented the temperature near the pipe) and TC10 (represented the temperature near the ground) ware selected as the inputs for RF-REF algorithm.

2) Soil water content: Water content is a critical property for the expansive soil, as the high variation in water content will cause soil swell-shrink, which could result in soil deformation. Based on analysis by Yafei Hu and his colleagues,[16] it was found that water content is primarily influenced by climate factors, e.g. snow melt, rainfall, freezing, and frost etc. Figure 2c presents the variations of typical volumetric water contents in the trench backfill (TDR) and in the native clay surrounding the trench (WCR). As shown in Fig. 2c, with an increase in depth, the magnitudes of the variations generally decreased. Furthermore, variation patterns in volumetric water content at 1.07m depth and 0.45m depth with time were similar while those at 1.96m depth and 1.07m depth were different. Hence, Pearson correlation coefficient was used to tackle multicollinearity issues as well. Based on the Equation (1), correlation coefficient between every two variables were calculated and variables correlated with each other more than 0.7 were removed. Finally, TDR3-6, TDR5-2, TDR7-4 and TDR9-6 were considered for the next step analysis.

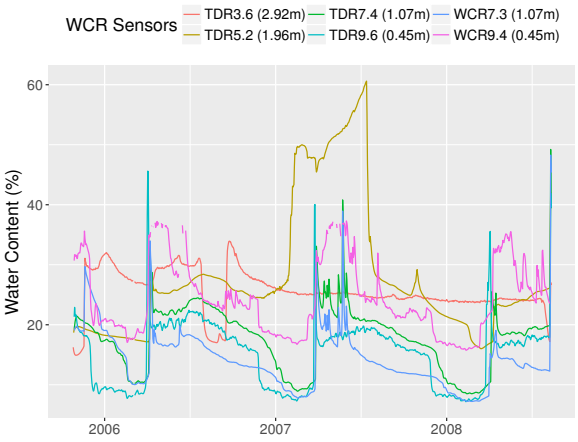
Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 5/8/2018 Terms of Use: https://www.spiedigitallibrary.org/terms-of-use



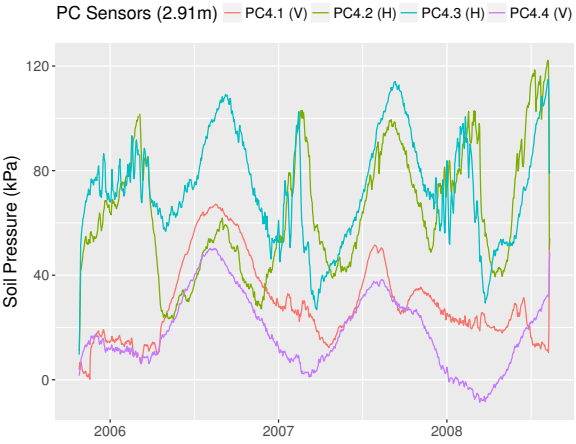
(a) Temperature measured horizontally.



(b) Temperature measured vertically.



(c) Water content measured vertically.



(d) Vertical and horizontal soil pressure.

Figure 2. Daily time series of soil properties.

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 5/8/2018 Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

3) Soil pressure: Soil pressure depends mainly on the nature of the soil, its natural saturation degree and the variation in soil moisture.[3] Figure 2d shows the variation of earth pressures on four pressure cells, where two distinct peaks can be observed annually in the vertical soil pressure, while the horizontal soil pressure only has one peak in the summer. Since the earth pressures measured by different sensors have similar variations with little difference in the magnitude, the average value of soil pressure was used to replace two highly related explanatory variables.

3.2.2 Pipe Deformation

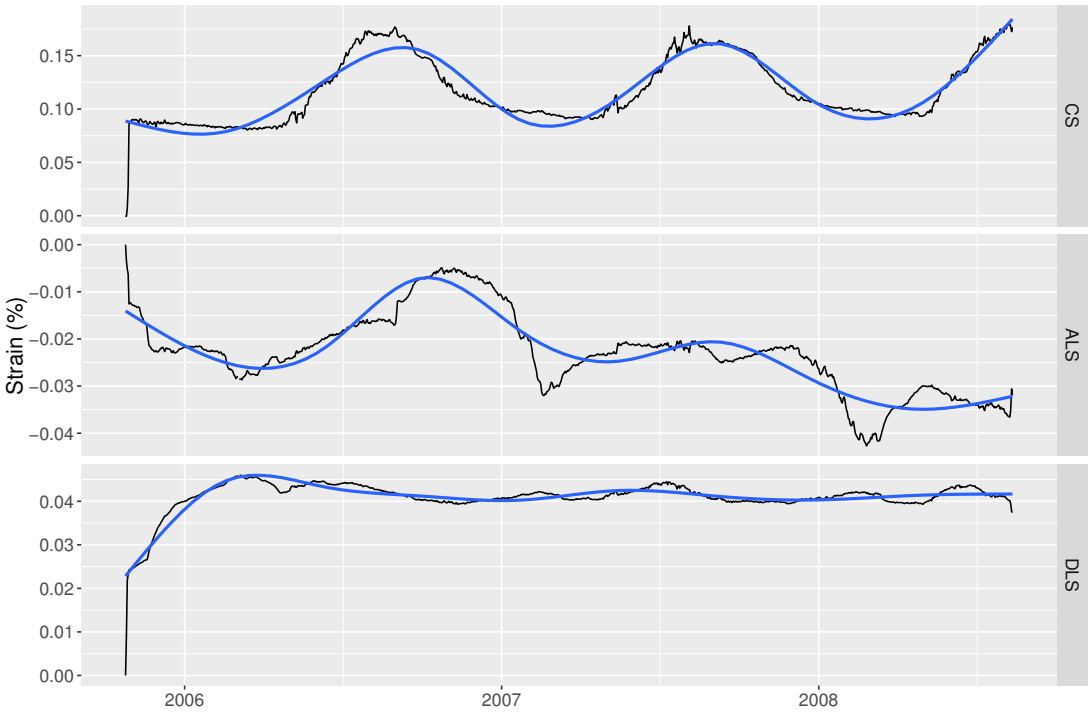


Figure 3. Daily time series of strains (The blue smooth lines describe the curve trend).

1) Circumferential strain (CS): Figure 3 shows the daily time series of measured CS where a seasonal variation can be observed. Generally, the strains are mainly a consequence of the seasonal pipe temperature variations,[21] which is demonstrated in Fig. 4 as well. Figure 4 depicts the pairwise relationships, where the lower left of the matrix consists of the correlation coefficient between variables while the upper right of the matrix highlights the strong correlation with a larger circle. Based on the observation of Fig. 4, we found that a close relationship among temperature, CS, and the soil pressure in vertical direction (AvgVPC), as the correlation coefficient calculated in pair is larger than 0.8. Since the correlation between TC5 and CS is larger than that between TC10 and CS, the temperature around pipe has a more significant impact on the CS.

2) Average longitudinal strains (ALS): The ALS is the mean of the longitudinal strain values measured at four points (top, bottom, and two springlines), which represents the tensile deformation in the longitudinal direction.[16] In Fig. 3, the time variation of the ALS is described, where a seasonal pattern can be found as well. However, different from CS, the variation magnitude of ALS decreases and the measurements tend to increase over time. Additionally, although there are correlations between ALS and temperature as well as soil pressure in vertical direction, they are not as obvious as those for CS.

3) Differential longitudinal strains (DLS): Figure 3 describes the variations of DLS, which reflects the pipe bending deformation. As shown in Fig. 3, the DLS curve tends to be smooth after April 2006, during which the value is relatively constant. Based on the description of Fig. 4, it appears that the DLS is not sensitive to the soil temperature and there is no strong correlation between the DLS and other variables.

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 5/8/2018 Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

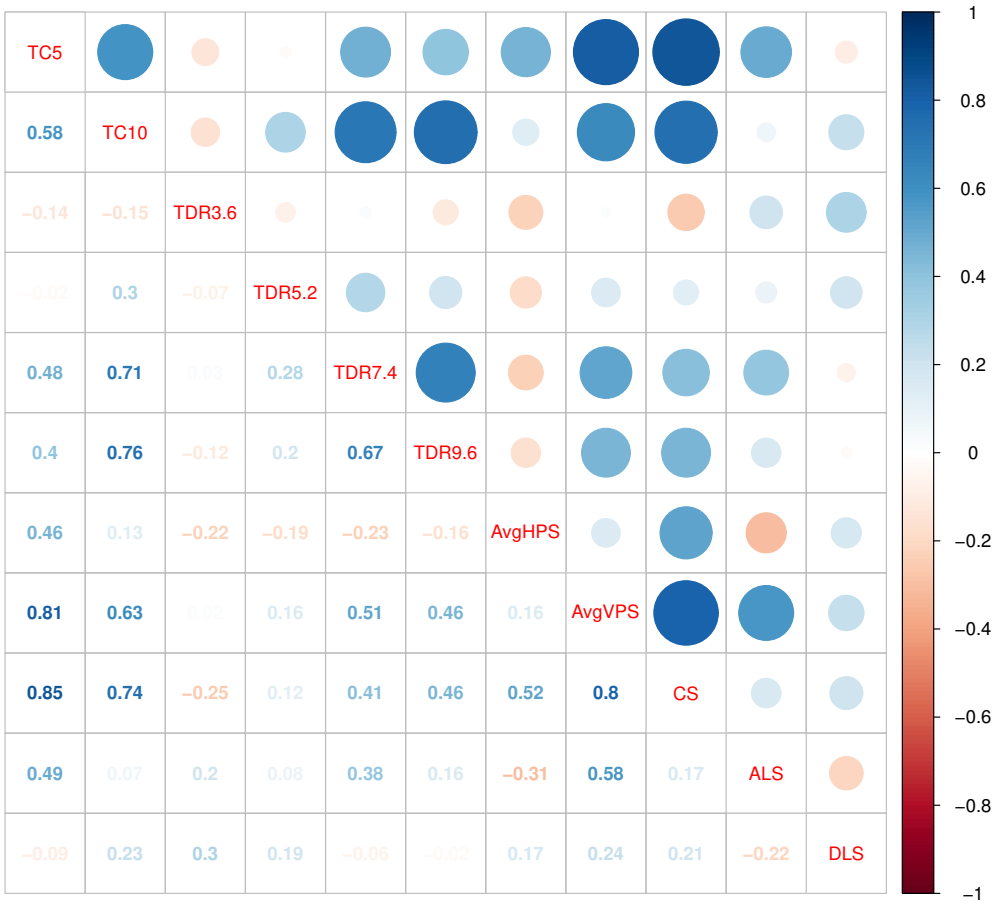


Figure 4. Correlation matrix for soil properties and pipe deformation performance.

3.3 Feature selection with random forest-recursive feature elimination

Although a number of redundant features have been removed during data analysis, the importance of each variables was unknown. Hence, RF-RFE was used to further identify confounding variables within the selected sensor data and, hence, to determine the optimum subset of variables. Constructing three datasets (CS, ALS, and DLS), along with soil property dataset, the key factors contributing to different type of deformations were ranked respectively by RF-RFE .

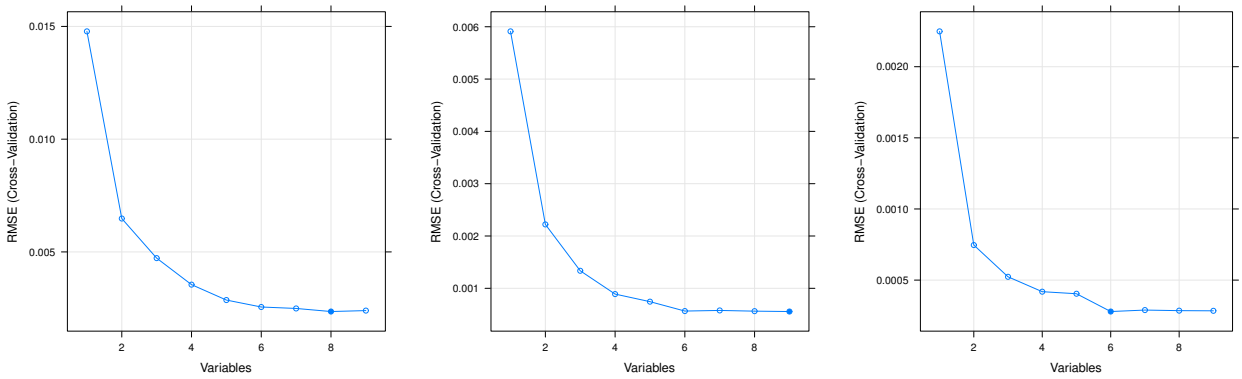
In this study, the caret package (version 6.0.77) in R (version 3.4.2) were used to implement RF-RFE with the full study data. Figure 5 demonstrates the feature selection results using RF-RFE for three different prediction outcomes, which describes the model accuracy when using different number of variables. As shown in Fig. 5, the optimum number of variables were between 6 to 9, which were represented as a blue solid point. However, the model error tends to be fairly small and has no much difference when the number of variables increased to 6. Therefore, considering both computational constrains and accuracy, the 6 most important predictors were selected for super learning construction. The predictor importance rankings were reported in Tab. 4.

3.4 Super learning framework implementation

3.4.1 Candidate learner selection

After variable selection, highly correlated soil attributes were removed and the optimum number of variables for modeling was determined by automatic feature selection methods, which also identified the most important factors contributing different type of pipe deformation. Utilizing the filtered data, multiple diverse regression

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 5/8/2018 Terms of Use: https://www.spiedigitallibrary.org/terms-of-use



(a) Data set with CS outcome. (b) Data set with ALS outcome. (c) Data set with DLS outcome.

Figure 5. RF-RFE results for feature selection (Blue solid points indicate the optimum number of variables used for the prediction).

Table 3. Details of the machine learning algorithms with training time on each data set.

Model	Model Value	Training Time (Sec)		
		CS	ALS	DLS
Linear Regression	lm	0.53	0.50	0.48
k-Nearest Neighbors	knn	0.61	0.54	0.51
Regression Tree	rpart	0.62	0.66	0.61
Conditional Inference Tree	ctree	1.06	1.03	0.96
Multi-Layer Perceptron	mlp	4.47	3.98	3.83
Support Vector Machines with Radial Basis Function Kernel	svmRadial	4.50	3.75	4.31
Random Forest	rf	50.64	45.67	47.85
Gaussian Process with Polynomial Kernel	gaussprPoly	62.69	63.52	65.58

algorithms available in Caret packaged (version 6.0.77) were investigated to determine the candidate learners for super learning. As shown in Tab. 3, the training time for each algorithm fitting the full dataset was computed, where rf and gaussprPoly required much more time to compute. Considering diversity, training time efficiency, and prediction accuracy, 6 models, e. i. lm, knn, rpart, ctree, mlp, and svmRadial, were selected as the base learners.

3.4.2 Super learner implementation details

Super learning is an ensemble technique for prediction modeling that combines a variety of single algorithms optimally. To implement pipe behavior prediction modeling using super learning, caretEnsemble package (version 2.0.0) was applied in this study. Each dataset was divided into two parts, where 75% of the data was used for model training and the rest for test purpose. 10-fold cross-validation was applied to evaluate each candidate learner in the training process, and then the weights of these fitted model were determined by their prediction performance across the validation sets.

To examine the prediction ability of super learning and the effectiveness of RF-RFE, 6 super learning models were built in this study. As given in Tab. 4, both SL1.1 and SL1.2 were built to predict CS, the only difference between these two was all attributes after correlation analysis were used by SL1.1 while only the 6 most important

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 5/8/2018 Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

Table 4. Details of the six super learning models using all selected base learners (Variables highlighted in boldface are features excluded by RE-RFE).

Model	Target	Covariates (Order by RF-RFE Importance Ranking)
SL1.1	CS	TC5, TDR3.6, TC10, AvgHPS, AvgVPS, TDR5.2, **TDR7.4, TDR9.6, Month**
SL1.2	CS	TC5, TDR3.6, TC10, AvgHPS, AvgVPS, TDR5.2
SL2.1	ALS	TDR3.6, Month, AvgVPS, TDR7.4, TDR5.2, AvgHPS, **TDR9.6, TC5, TC10**
SL2.2	ALS	TDR3.6, Month, AvgVPS, TDR7.4, TDR5.2, AvgHPS
SL3.1	DLS	TDR3.6, Month, TDR9.6, TDR5.2, AvgVPS, TDR7.4, **AvgHPS, TC5, TC10**
SL3.2	DLS	TDR3.6, Month, TDR9.6, TDR5.2, AvgVPS, TDR7.4

features were considered in SL1.2. Similarly, SL2.1 and SL2.2 were constructed to predict ALS in pair but using different number of covariates, and the pair of SL3.1 and SL3.2 was for DLS prediction.

4. EXPERIMENT RESULTS

4.1 Performance evaluation metrics

To evaluate the super learning predication performance, three criteria were used in this experiment: the root mean square error ($RMSE$), R-squared (R^2), and mean absolute (MAE). In statistics, $RMSE$ and MAE are two metrics frequently used to evaluate the prediction modeling accuracy. $RMSE$ is used to measure the differences between predicted values and the observed values while MAE measures the average magnitude of the errors without considering their direction. Since $RMSE$ gives a relatively high weight to large errors, it is more useful when large errors are particularly undesirable. The equations of $RMSE$ and MAE are given in Eqs. (2) and (4), respectively. Equation (3) is the formula of R^2, which describes the fitting degree of models in a $[0, 1]$ range, where the larger the R^2 value is, the better the model fits the data.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y_i})^2} \tag{2}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \widehat{y_i}| \tag{3}$$

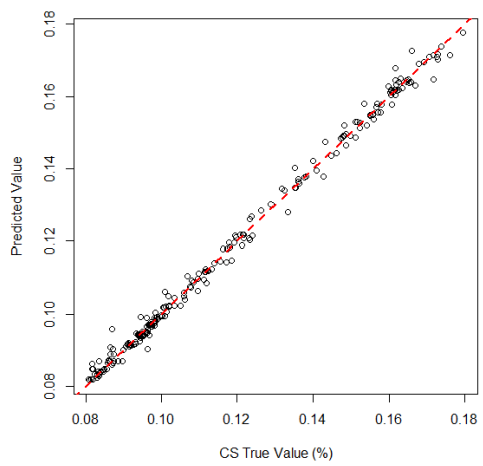
$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \widehat{y_i})^2}{\sum_{i=1}^{n}(y_i - \overline{y_i})^2} \tag{4}$$

where n is the number of observations in the data set, and y_i and $\widehat{y_i}$ represent the observed outcomes and predicted target values, respectively.

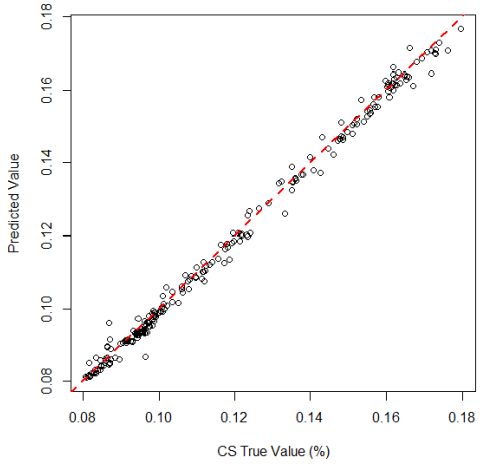
4.2 Experiment results and discussion

Table 5 provides the $RMSE$, MAE, and R^2 values for both training sets and test sets for all super learners and Figure 6 shows the prediction results obtained by different super learning models. From the experimental results given in the Tab. 5 and Fig. 6, the following conclusions were drawn: super learning algorithm is powerful and adaptive to different data sets, which has the capability of predicting the CS, ALS, and DLS accurately; benefiting from RF-RFE algorithm, the super learner prediction performance is still excellent even only using 6 variables; finally, the relationship between the soil properties and pipe deformation was successfully modeled using the proposed super learning based approaches.

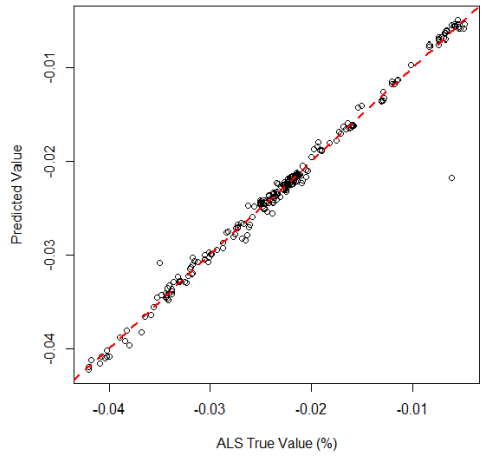
Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 5/8/2018 Terms of Use: https://www.spiedigitallibrary.org/terms-of-use



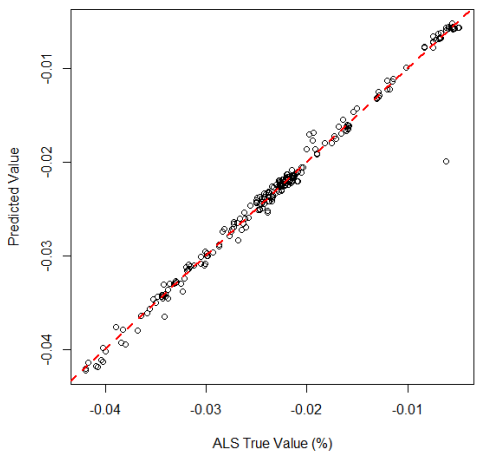
(a) SL1.1.



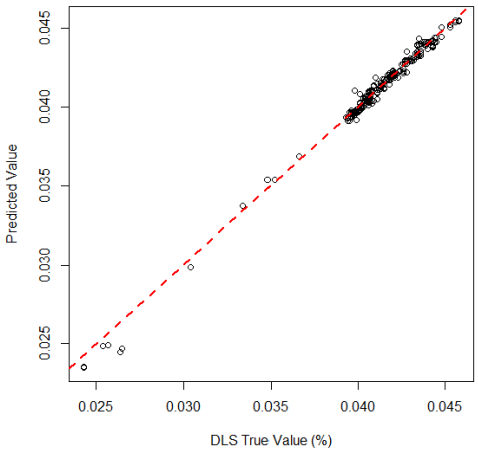
(b) SL1.2.



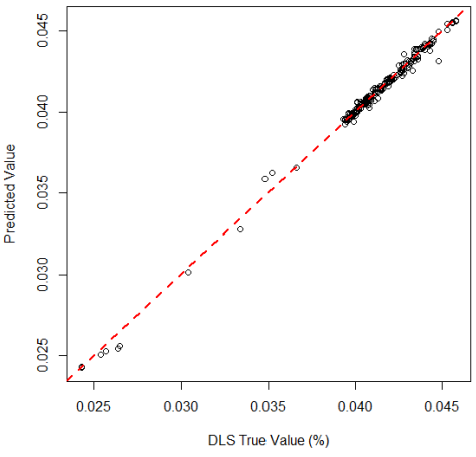
(c) SL2.1.



(d) SL2.2.



(e) SL3.1.



(f) SL3.2.

Figure 6. Prediction results of six different super learning models.

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 5/8/2018 Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

Table 5. Model assessment with $RMSE$, MAE, and R^2.

Model	Training			Testing		
	RMSE	MAE	R^2	RMSE	MAE	R^2
SL1.1	0.00304	0.00169	0.98677	0.00194	0.00132	0.99553
SL1.2	0.00321	0.00187	0.98576	0.00218	0.00162	0.99506
SL2.1	0.00073	0.00046	0.99186	0.00116	0.00049	0.98214
SL2.2	0.00073	0.00049	0.99217	0.00107	0.00051	0.98488
SL3.1	0.00059	0.00030	0.96367	0.00034	0.00024	0.99119
SL3.2	0.00052	0.00025	0.96772	0.00025	0.00016	0.99367

5. CONCLUSION

In this study, the relationship between pipe deformations and soil properties was analyzed and characterized. Based on the data collected from the sensor monitoring system, we employed super learning techniques to predict the pipe deformation behavior. Combined with automatic feature selection approaches, the proposed methods are accurate, scalable and simple, of which performance were thoroughly investigated for the pipe performance prediction. Using three historical sensor data sets, we validated the adaptability of super learning for combining information from a set of algorithms to improve the prediction results. The modeling approach provided precise results for the pipe deformation prediction, considering soil temperature, water contents, and earth pressure. Additionally, RF-RFE algorithm helped to achieve the accurate prediction with only 6 factors, which also reported the importance of each factors contributing to the pipe deformation. Based on the data analysis results, we also found that temperature around the pipe has a closed linear relationship with the vertical soil pressure, hence, resulting in pipe hoop deformation. However, another two types of deformation (e.g. tensile deformation in the longitudinal direction and bending deformation) are not linearly correlated with soil properties. The proposed super learning methods in this study can facilitate the evaluation of pipe distress in the expansive soil environment.

ACKNOWLEDGEMENT

This work is partially supported through the NSERC Discovery Grant.

REFERENCES

[1] Hu, Y., Vu, H. Q., and Lotfian, K., "Instrumentation of a section of ac pipe in expansive soil," in [*Pipelines 2008: Pipeline Asset Management: Maximizing Performance of our Pipeline Infrastructure*], 1–10 (2008).

[2] Rajeev, P., Chan, D., and Kodikara, J., "Ground–atmosphere interaction modelling for long-term prediction of soil moisture and temperature," *Canadian Geotechnical Journal* **49**(9), 1059–1073 (2012).

[3] Saadeldin, R., Hu, Y., and Henni, A., "Numerical analysis of buried pipes under field geo-environmental conditions," *International Journal of Geo-Engineering* **6**(1), 1–22 (2015).

[4] Wijewickreme, D., "Role of geotechnical engineering in assuring the integrity of buried pipeline systems," in [*International Conference on Sustainable Built Environments*], (2012).

[5] Tian, Y. and Cassidy, M. J., "Modeling of pipe–soil interaction and its application in numerical simulation," *International Journal of Geomechanics* **8**(4), 213–229 (2008).

[6] Winkler, E., "Theory of elasticity and strength," *Czechoslovakia: Dominicus, Prague (In German)* (1867).

[7] Liu, X., Zhang, H., Han, Y., Xia, M., and Zheng, W., "A semi-empirical model for peak strain prediction of buried x80 steel pipelines under compression and bending at strike-slip fault crossings," *Journal of Natural Gas Science and Engineering* **32**, 465–475 (2016).

[8] Trickey, S. A. and Moore, I. D., "Three-dimensional response of buried pipes under circular surface loading," *Journal of Geotechnical and Geoenvironmental Engineering* **133**(2), 219–223 (2007).

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 5/8/2018 Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

[9] Tang, H., Fairchild, D., Panico, M., Crapps, J., and Cheng, W., "Strain capacity prediction of strain-based pipelines," in [*Proceedings of the 10th international pipeline conference, IPC*], (2014).

[10] Kleiner, Y., Nafi, A., and Rajani, B., "Planning renewal of water mains while considering deterioration, economies of scale and adjacent infrastructure," *Water Science and Technology: Water Supply* **10**(6), 897–906 (2010).

[11] Rajeev, P. and Kodikara, J., "Numerical analysis of an experimental pipe buried in swelling soil," *Computers and Geotechnics* **38**(7), 897–904 (2011).

[12] Salehi, H., Das, S., Chakrabartty, S., Biswas, S., and Burgueño, R., "A machine-learning approach for damage detection in aircraft structures using self-powered sensor data," in [*Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2017*], **10168**, 101680X, International Society for Optics and Photonics (2017).

[13] Yajima, A., Wang, H., Liang, R. Y., and Castaneda, H., "A clustering based method to evaluate soil corrosivity for pipeline external integrity management," *International Journal of Pressure Vessels and Piping* **126**, 37–47 (2015).

[14] Liu, Z., Sadiq, R., Rajani, B., and Najjaran, H., "Exploring the relationship between soil properties and deterioration of metallic pipes using predictive data mining methods," *Journal of Computing in Civil Engineering* **24**(3), 289–301 (2009).

[15] Liu, Z., Hu, Y., and Wu, W., "Pipe performance analysis with nonparametric regression," in [*Nondestructive Characterization for Composite Materials, Aerospace Engineering, Civil Infrastructure, and Homeland Security*], **7983**, 798337–1–798337–9, International Society for Optics and Photonics (2011).

[16] Hu, Y. and Vu, H. Q., "Analysis of soil conditions and pipe behaviour at a field site," *Canadian Geotechnical Journal* **48**(6), 847–866 (2011).

[17] Guyon, I. and Elisseeff, A., "An introduction to variable and feature selection," *Journal of machine learning research* **3**(Mar), 1157–1182 (2003).

[18] Granitto, P. M., Furlanello, C., Biasioli, F., and Gasperi, F., "Recursive feature elimination with random forest for ptr-ms analysis of agroindustrial products," *Chemometrics and Intelligent Laboratory Systems* **83**(2), 83–90 (2006).

[19] LeDell, E. E., [*Scalable Ensemble Learning and Computationally Efficient Variance Estimation*], University of California, Berkeley (2015).

[20] Ju, C., Combs, M., Lendle, S. D., Franklin, J. M., Wyss, R., Schneeweiss, S., and van der Laan, M. J., "Propensity score prediction for electronic healthcare databases using super learner and high-dimensional propensity score methods," *arXiv preprint arXiv:1703.02236* (2017).

[21] Kuraoka, S., Rajani, B., and Zhan, C., "Pipe-soil interaction analysis of field tests of buried pvc pipe," *Journal of infrastructure systems* **2**(4), 162–170 (1996).

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 5/8/2018 Terms of Use: https://www.spiedigitallibrary.org/terms-of-use