Proc. of the 8th International Symposium on Geotechnical Safety and Risk (ISGSR)
Edited by Jinsong Huang, D.V. Griffiths, Shui-Hua Jiang, Anna Giacomini and Richard Kelly

doi: 10.3850/978-981-18-5182-7_08-003-cd

ISGSR 2022

Prediction of Parallel Desiccation Cracks of Clays Using a Classification and Regression Tree (CART) Technique

Abolfazl Baghbani[1], Susanga Costa[2], Tanveer Choundhury[3] and Roohollah Shirani Faradonbeh[4]

[1]School of Engineering, Deakin University, 3086 VIC, Australia.
E-mail: abaghbani@deakin.edu.au
[2]School of Engineering, Deakin University, 3086 VIC, Australia.
E-mail: susanga.costa@deakin.edu.au
[3]School of Engineering, IT and Physical Sciences, Federation University Australia, 3842 VIC, Australia.
E-mail: t.choudhury@federation.edu.au
[4]WA School of Mines: Minerals, Energy and Chemical Engineering, Curtin University, Kalgoorlie, WA, 6430, Australia.
E-mail: roohollah.shiranifaradonbeh@curtin.edu.au

Abstract: One of the most effective and common phenomena, especially in arid and semi-arid regions, is cracking caused by soil drying. Cracking changes soil properties such as permeability. As a result, recognizing the amount and nature of cracks in clay and predicting them can greatly contribute to infrastructure design. This paper seeks to predict the number of parallel clay cracks using a classification and regression tree (CART) technique under several variables including initial water content, soil thickness and sample width and length as input parameters. A database consisting of 31 datasets was used in the study. To evaluate the accuracy of the model, two statistical indices, namely the coefficient of determination (R^2), root mean square error (RMSE) have been used. The results are compared with the Multiple linear regression (MLR) method. Results show that for testing database, R^2 and RMSE based on the classification and regression tree (CART) model are 0.989 and 1.285, respectively, while the R^2 and RMSE in the multiple linear regression method were equal to 0.777 and 41.696, respectively. As a result, the classification and regression tree (CART) has performed well in predicting the number of desiccation cracks.

Keywords: Parallel desiccation cracking; Clay; Artificial Intelligence; Classification and Regression Tree.

1 Introduction

Desiccation cracking is a natural phenomenon in soils that occurs due to loss of soil moisture. The importance of cracking is that it can change the effective parameters of soils. Cracks in nature can occur in various dimensions, from less than a few millimeters to several hundred meters. The investigation of cracks has been studied by researchers in various fields such as geology, geomechanics, and other related fields (Peron et al. 2009; Sahebzadeh et al. 2017; Costa et al. 2018; Xie et al. 2020; Cordero et al. 2021; Zhuo et al. 2022, Baghbani et al., 2022a).

In the last two decades, number of efforts have been made by researchers to develop numerical and analytical models to predict desiccation cracks in clay soils. (Kodikara and Costa 2013; Tang et al. 2021). Researchers used numerical methods, the finite difference method (FDM) and the finite element method (FEM) to model desiccation cracking. The finite element method (FEM), despite its limitations for evolving discontinuities in porous media, has been used as a common method for investigating soil cracking (Amarasiri et al. 2011; Tang et al. 2021).

One of the methods that geotechnical engineers have used to predict various parameters with high complexity and high number of factors, especially in the last two decades, is artificial intelligence (AI) methods (Baghbani et al. 2022b). To date, only one study (Choudhury and Costa 2018) has been published on applying AI methods to soil desiccation crack prediction. In the study conducted by Choudhury and Costa (2018), the artificial neural network (ANN) was successfully used to predict soil desiccation cracking though they only used a database containing 16 datasets available in literature. A small dataset can reduce and limit the distribution of effective parameters. One of the most popular AI methods that has not yet been used in the field of desiccation cracking is the decision tree method. The decision tree method is a white box method that means a tree as a model result, can be easily used by readers to predict the output parameter based on new inputs. Comparing the results obtained from the decision tree method and ANN can provide a proper view of both methods, namely the black box (like ANN) and the white box (like the decision tree).

The current study aims to develop a new decision tree algorithm which is the classification and regression tree (CART) to predict the number of cracks in thin parallel (1D) desiccation cracks. For this purpose, a database of parallel cracks in thin long clay layers was compiled using published literature (16 datasets) as well as new tests (15 datasets). Image analysis techniques were used to collect the data on number of cracks. Different CART-based models were developed, and the best CART model was selected. The database of these models included four inputs, namely, soil thickness, initial water content, width and length of the sample, and the output was the number of cracks.

2 Laboratory cracking tests

2.1 Materials and experimental program

Two types of clays were used, namely kaolin and bentonite. These two were mixed in different percentages to create several clay mixtures with varying characteristics. As it can be seen in Table 1, three different clay mixtures were prepared. In this way, a range of different clay characteristics were created from low to high liquid limit (LL). Atterberg limits of the mixtures are also shown in Table 1.

Table 1. The different types of mixture by mixing kaolin and bentonite.

Mixture	Kaolin (%)	Bentonite (%)	Liquid Limit (%)	Plastic Limit (%)
1	100	0	74	34
2	50	50	155	35
3	0	100	223	36

2.2 Samples preparation and testing procedures

Desiccation cracking tests were conducted in long, rectangular moulds with different thicknesses, at 5, 10, 15, 20 and 25 mm. Figure 1a shows the 5 acrylic moulds. They all have the same width (25 mm) and length (600 mm) but varying thicknesses.

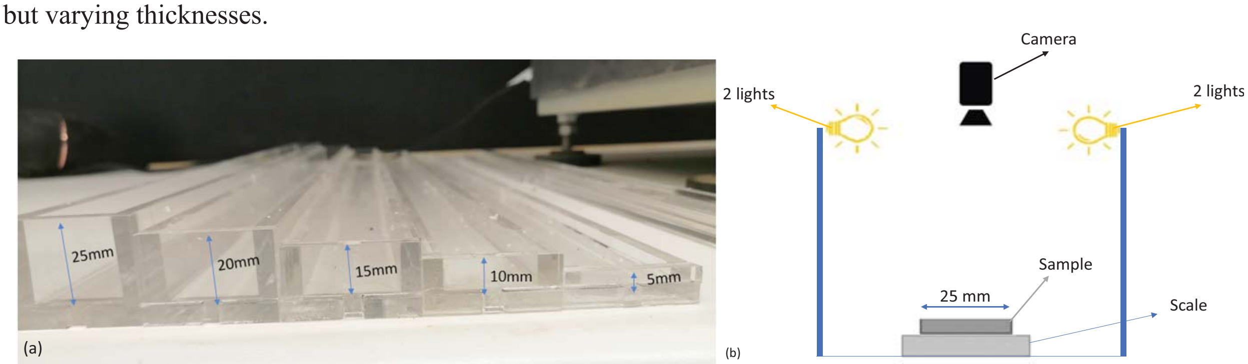

Figure 1. (a) Moulds and (b) test set up for the parallel (1D) cracks.

Samples with different thicknesses and the initial water content equal to the liquid limit were prepared. To minimize the adhesion of the mould wall and clay samples, a thin layer of grease was applied on the mould walls. Prepared samples were dried under flood lights to accelerate the desiccation (see Figure 1b). Mass of the samples was measured at every 30 minutes to calculate the drying rate and the moisture content. The test duration for each sample was 270 minutes (4.5 hrs). At the end of this period, a photo was taken to count the number of cracks which was the test output.

2.3 Classification and regression tree (CART) modelling

Classification and regression tree (CART), introduced by Breiman et al. (2017), is a machine learning algorithm that can be used for prediction and classification tasks, depending on the output parameter type (i.e., quantitative, or qualitative). This method can determine the latent relationship between input and output parameters without prior knowledge. On the other hand, this algorithm can deal with noisy data by automatically identifying the outliers and categorizing them into separate clusters. The output of this method is a binary tree structure that can be used easily to predict parameters (Hasanipanah et al. 2017; Khandelwal et al. 2017).

Figure 2 shows a typical decision tree. A tree contains several nodes and branches. As shown in Figure 2, a set of nodes and branches creates a leaf. Each node is divided into left and right nodes based on the determined rule for each branch (e.g., if Node A<=a, then Node C). Finally, in the last node of each leaf, the predicted output is presented (e.g., nodes D-G in Figure 2). In contrast to other common machine learning algorithms (e.g., ANN), CART benefits from simplicity, explicability and low computational costs. This, on the other hand, makes it more practical for geotechnical engineering applications (Shirani Faradonbeh et al. 2022).

2.3.1 Database

In addition to performing 15 soil desiccation cracking tests on Werribee clay, a material from Werribee Victoria in Australia, 16 more datasets were also compiled from the literature (Costa 2009) to create a bigger database (a total of 31 datasets). As the input and output parameters have different ranges of values, and to prevent over-fitting problem, the parameters were normalized using Equation 1.

$$X_{normal} = (X - x_{min}) / (x_{max} - x_{min}) \tag{1}$$

where X_{normal}, X, x_{min} and x_{max} are the normalized, actual, minimum, and maximum parameter value, respectively.

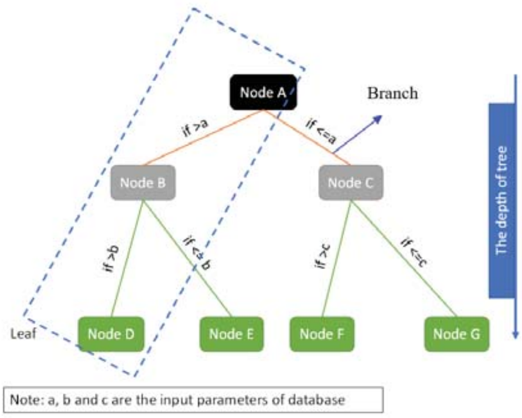

Figure 2. A typical decision tree for CART.

3 Test results

Figure 3 shows the database containing 31 datasets. The results show the effect of different factors on the number of cracks.

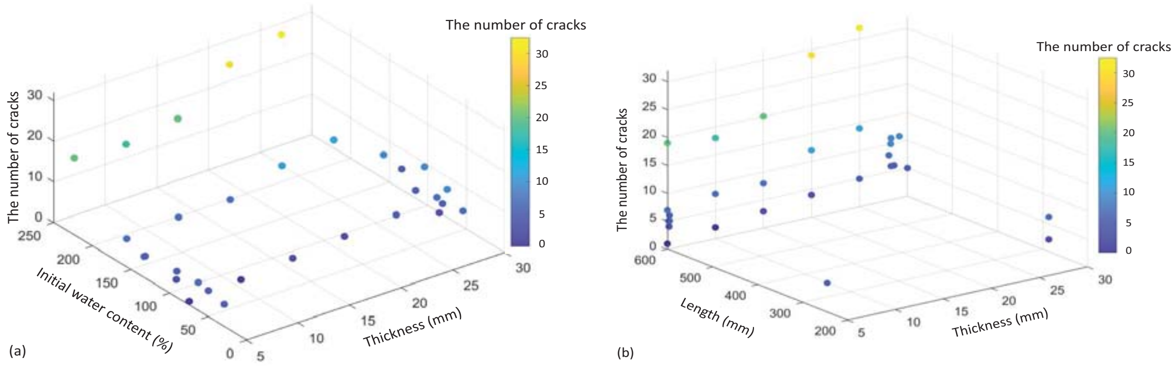

Figure 3. The used database (including 31 datasets) of desiccation cracking for CART modelling.

3.1 Effect of soil thickness and initial water content on the number of cracks

Figure 4a shows the effect of sample thickness on the number of clay cracks. As can be seen, the increase in thickness has relatively increased the number of cracks in the clay. This can be related to the increase in soil sample size and consequently the increase in loose and failure plates. As it is shown in Figure 4a, the correlation between the number of cracks and soil thickness is nonlinear. It can also be concluded that increasing the thickness of the clay soil sample can change the state of stresses between particles during soil drying. This can affect the state of moisture (water) transfer and thermal energy in sample. Increasing the thickness can also increase the number of interparticle colonies, which is due to electrochemical reasons. Increasing the number of clay colonies can increase soil cracking. Figure 4b shows the effect of the initial water content of the clay sample on the number of desiccation cracks. The results show that with increasing initial water content, the number of clay desiccation cracks increased. The reason for this can be the decrease in clay suction due to the increase in soil water content and degree of saturation. Also, reducing the suction of clay can decrease the tensile strength of clay, which can also increase the number of soil cracks.

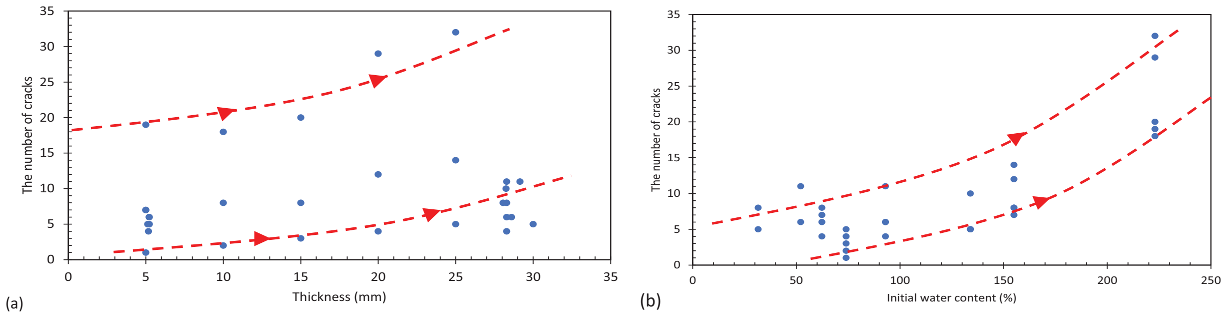

Figure 4. The effect of (a) soil thickness and (b) initial water content on the number of desiccation cracks.

4 Classification and regression tree (CART) modelling approach

The database consisting of 31 datasets, including four inputs and one output, was used. The input parameters include thickness, initial soil water content, sample length and width, and the output parameter is the number of cracks. To examine the AI model, the database was randomly divided into two categories, including a training database with 80% of the total database (i.e., 25 datasets) and a testing database with 20% of the total database (i.e., 6 datasets). Tables 2 and 3 show the statistical information of the training and testing databases, respectively. According to these tables, it can be observed that both the training and testing databases have an almost similar distribution of data, including minimum, maximum, mean and standard deviation were close to each other.

Table 2. Statistical information of training database.

Variable	Observations	Minimum	Maximum	Mean	Std. deviation
Number of cracks	25	1.00	12.00	6.28	2.78
Thickness	25	5.00	30.00	17.79	9.95
Initial MC	25	31.60	223.00	114.15	61.66
Length	25	250.00	600.00	558.00	116.08
Width	25	25.00	50.00	26.00	5.00

Table 3. Statistical information of testing database.

Variable	Observations	Minimum	Maximum	Mean	Std. deviation
Number of cracks	6	14.00	32.00	22.00	6.96
Thickness	6	5.00	28.29	12.29	9.73
Initial MC	6	52.08	223.00	121.85	62.39
Length	6	600.00	600.00	600.00	0.00
Width	6	25.00	25.00	25.00	0.00

To evaluate the accuracy of different models, two parameters of coefficient of determination (R^2) and root mean square error (RMSE), which are calculated using the following equations, were used.

$$R^2 = \left[\frac{\sum_N (X_m - \bar{X}_m)(X_p - \bar{X}_p)}{\left[\sum_N (X_m - \bar{X}_m)^2 \sum_N (X_p - \bar{X}_p)^2\right]}\right]^2 \tag{2}$$

$$RMSE = \sqrt{\sum_N \frac{(X_m - X_p)^2}{N}} \tag{3}$$

where X_m, X_p, $\overline{X_m}$, $\overline{X_p}$ and N are actual and predicted values, the average of actual and predicted values, and the number of datasets, respectively. The best model is a model that has the R^2 of 1 and RMSE equal 0.
To evaluate the relationship between the input parameters and the corresponding output parameter, the Pearson correlation (r) was calculated, and the results were listed as the correlation matrix in Table 4. This table shows no considerable correlation between the inputs, which reveals that the inputs will not suffer from the multicollinearity problem. On the other hand, the correlation between the inputs and the output parameter is lower than 40%, indicating that the number of cracks cannot be easily predicted only based on a single input parameter. In other words, a more complex model is required to detect the interrelationship of the parameters.

To evaluate the possible linear relationship between parameters, the multiple linear regression (MLR) analysis was carried out using MATLAB software. The results of MLR shows that the coefficient of determination (R^2) for training and testing sections was 0.169 and 0.777, respectively.

Also, the root mean square error (RMSE) of the obtained MLR model for the training and testing sections is 2.773 and 41.696, respectively. The results show that the relationship between parameters is not linear, and simple statistical methods, such as MLR, cannot accurately predict the number of cracks. Therefore, it is necessary to use other artificial intelligence (AI) methods, such as CART, that have appropriate capabilities to find the non-linear relationship between input and output parameters.

Table 4. Statistical information of testing database.

	Thickness	Initial water content	Length	Width	Number of cracks
Thickness	1	-0.171	-0.107	0.220	0.012
Initial MC	-0.171	1	0.442	-0.210	0.047
Length	-0.107	0.442	1	0.075	-0.052
Width	0.220	-0.210	0.075	1	0.354
Number of cracks	0.012	0.047	-0.052	0.354	1

4.1 Classification and regression tree (CART) results

Figure 5 represents the best CART model for predicting the number of desiccation cracks based on four input parameters. The prediction of output parameter in this tree shows similar results compared to the history of studies (Tang et al. 2021). The interpretation of the obtained CART model is quite simple, and the output parameter (i.e., the number of cracks) can be straightforwardly estimated. For example, in Figure 5, there are two rules in node 1:

Left branch: If the initial water content is less than or equal to 189%.

Right branch: If the initial water content is greater than 189%.

Then, assuming the right branch is selected, in node 3, a new rule, which is based on the soil thickness, for going to different branches is presented again. This process continues until the predicted value at the end of each leaf is reached.

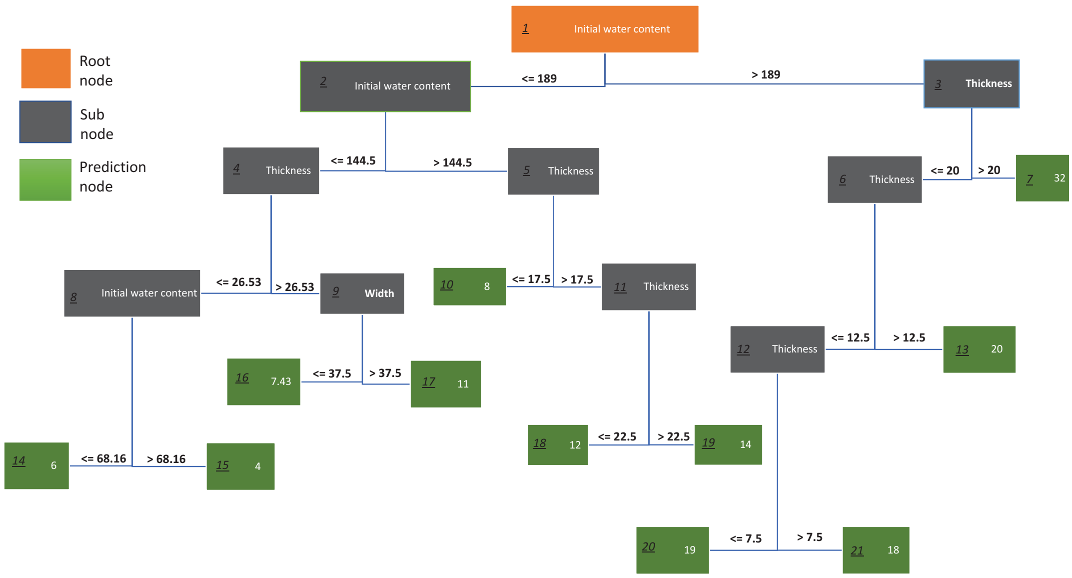

Figure 5. The best tree of CART.

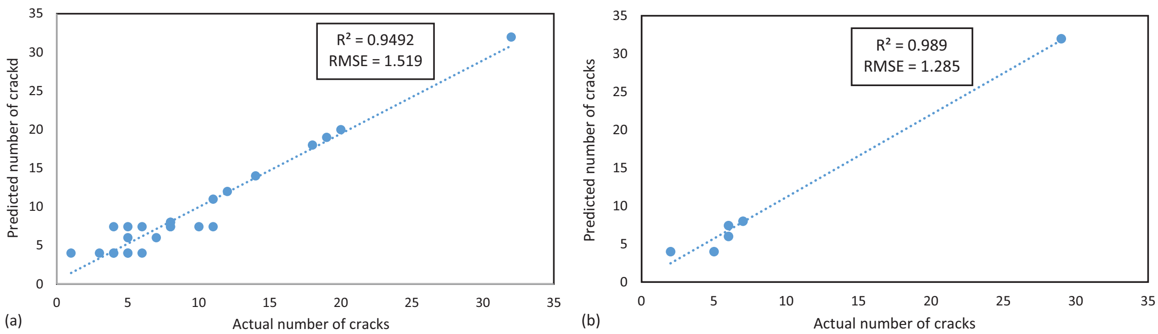

Figure 6. The predicted number of cracks from the best CART model versus actual number of cracks, for (a) training and (b) testing databases.

The relationship between the measured and predicted output parameter for training and testing stages is shown in Figure 6. As can be observed from this figure, the prediction performance indices of R^2 and RMSE for the trained model are 0.949 and 1.519, respectively. The values of the forgoing indices for the testing stage are 0.989

and 1.285. These results show the high capability of the CART model compared to the MLR technique in predicting the number of cracks.

5 Conclusions

One of the most influential phenomena on soil behaviour is desiccation cracking. Classification and regression tree model was used for the first time to predict the number of parallel (1D) desiccation cracks in soil. To prepare this database, 15 desiccation tests with five different thicknesses and three clay mixes were prepared. Further 16 datasets were collected from literature.

The generated database was then used to model different CARTs. This database consisted of four inputs, namely soil thickness, initial water content, length and width, and number of soil cracks, the only output of this database. The results show that the best CART has correctly predicted the effect of various input parameters. The proposed CART also had a R^2 of 0.949 and 0.989 for the training and testing database, respectively. Also, the CART error (RMSE) was 1.519 and 1.685 for the training and testing database, respectively. In summary, the results indicate that CART can provide acceptable results for predicting soil desiccation cracking. It should also be noted that this study focused on the most simplified type of desiccation cracking (ID) with a simple output of crack number. Future research should investigate more complex crack patterns and use robust outputs that can describe crack patterns more accurately.

Acknowledgments

This research is supported by an Australian Government Research Training Program (RTP) Scholarship.

References

Amarasiri, A., Kodikara, J. and Costa, S. (2011). Numerical Modelling of Desiccation Cracking, *International Journal for Numerical and Analytical Methods in Geomechanics*, 35, 82 – 96.

Baghbani, A., Baghbani H., Shalchiyan MM. and Kiany K. (2022a). Utilizing artificial intelligence and finite element method to simulate the effects of new tunnels on existing tunnel deformation. *Journal of computational and Cognitive Engineering 2022*, vol 1(3), 152-162.

Baghbani, A., Choudhury, T., Costa, S. and Reiner, J. (2022b). Application of artificial intelligence in geotechnical engineering: A state-of-the-art review. *Earth-Science Reviews*, 103991.

Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J. (2017). Classification and regression trees. *Routledge*.

Choudhury, T. and Costa, S. (2018). Prediction of parallel clay cracks using neural networks–a feasibility study. *in International Congress and Exhibition" Sustainable Civil Infrastructures: Innovative Infrastructure Geotechnology"*, 214-24.

Costa, S. (2009). Study of desiccation cracking and fracture properties of clay soils. *Doctor of Philosophy thesis, Monash University*.

Costa, S., Kodikara, J., Barbour, S. and Fredlund, D. (2018). Theoretical analysis of desiccation crack spacing of a thin, long soil layer. *Acta Geotechnica*, 13 (1), 39-49.

Hasanipanah, M., Faradonbeh, R.S., Amnieh, H.B., Armaghani, D.J. and Monjezi, M. (2017). Forecasting blast-induced ground vibration developing a CART model. *Engineering with Computers*, 33 (2), 307-16.

Khandelwal, M., Armaghani, D.J., Faradonbeh, R.S., Yellishetty, M., Majid, M.Z.A. and Monjezi, M. (2017). Classification and regression tree technique in estimating peak particle velocity caused by blasting. *Engineering with Computers*, 33 (1), 45-53.

Kodikara, J. and Costa, S. (2013). Desiccation cracking in clayey soils: mechanisms and modelling. in Multiphysical testing of soils and shales, *Springer*, 21-32.

Peron, H., Hueckel, T., Laloui, L. and Hu, L.B. (2009) Fundamentals of desiccation cracking of fine-grained soils: experimental characterization and mechanisms identification. *Can Geotech J* 46:1177–1201

Cordero, J.A., Prat, P.C. and Ledesma, A. (2021) Experimental analysis of desiccation cracks on a clayey silt from a large-scale test in natural conditions. *Engineering Geology* 292. https://doi.org/10.1016/j.enggeo.2021.106256

Sahebzadeh, S., Heidari, A., Kamelnia, H. and Baghbani, A. (2017). Sustainability features of Iran's vernacular architecture: a comparative study between the architecture of hot–arid and hot–arid–windy regions. *Sustainability*, 9 (5), 749.

Shirani Faradonbeh, R., Taheri, A. and Karakus, M. (2022). Fatigue Failure Characteristics of Sandstone Under Different Confining Pressures. *Rock Mechanics and Rock Engineering*, 1-26.

Tang, C.S., Zhu, C., Cheng, Q., Zeng, H., Xu, J.J., Tian, B.G. and Shi, B. (2021). Desiccation cracking of soils: A review of investigation approaches, underlying mechanisms, and influencing factors. *Earth-Science Reviews*, 216, 103586.

Tang, T., Hededal, O. and Cardiff, P. (2015). On finite volume method implementation of poro-elasto-plasticity soil model. *International journal for numerical and analytical methods in geomechanics*, 39(13), 1410-1430.

Xie, Y., Costa, S., Zhou, L. and Kandra, H. (2020) Mitigation of desiccation cracks in clay using fibre and enzyme, *Bulletin of Engineering Geology and the Environment*, 79:4429-4440.

Zhuo, Z., Zhu, C., Tang, C.S., Xu, H., Shi, X. and Mark, V. (2022) 3D characterization of desiccation cracking in clayey soils using s structured light scanner. *Engineering Geology*. 299 https://doi.org/10.1016/j.enggeo.2022.106566.