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
 Published by Research Publishing, Singapore.
doi: 10.3850/978-981-18-5182-7_10-005-cd

ISGSR 2022

3D Underground Stratification Using GLASSO

Kota Kageie[1] and Takayuki Shuku[2]

[1]Graduate School of Environmental and Life Science, Okayama University, 3-1-1 Tsushimanaka, Kita-ku, Okayama 700-8530, Japan.E-mail: pms67xgq@s.okayama-u.ac.jp
[2]Graduate School of Environmental and Life Science, Okayama University, 3-1-1 Tsushima naka, Kita-ku, Okayama 700-8530, Japan. E-mail: shuku@cc.okayama-u.ac.jp

Abstract: This paper presents application examples of Geotechnical lasso (Glasso) in three-dimensional underground stratification of actual sites,Hollywood (South Carolina, USA) (Stuedlein et al. 2016) and Baytown (Texas, USA). Glasso is a data-driven site characterization method for estimating trends and detecting layer boundaries consistently and is based on a sparse machine learning method called least absolute shrinkage and selection operator (lasso). Cone penetration test (CPT) data are available in both sites, and they were converted to the soil behavior type (SBT) index for the underground stratification. The performance of Glasso in underground stratification was evaluated with validation set based on two performance metrics, root-mean-square error (RMSE) of I_c and identification ratio (IR) of SBT. Glasso is capable of detecting layer boundaries without the need to choose basis functions, and this is a notable advantage of Glassoin underground stratification.

Keywords: Soil stratification;Geotechnical lasso;Cone penetration test.

1 Introduction

In several countries, the underground space is an integral component in urban planning. This requires to implement undergrounds tratification that can be used for the design of geotechnical structures, such as foundations of buildings, bridges, and tunnels. In particular, the requirement for three-dimensional (3D) spatial information is rapidly increasing because it provides significant spatial insights, precise and objective representations of real-world phenomena, and accurate interpretation of spatial relations. Notwithstanding the high demand for 3D underground stratification, the characteristics of geotechnical data is considered as Multivariate, Uncertain and unique, Sparse, and In Complete (MUSIC) (Phoon et al. 2016), and the “sparsity” particularly leads to difficulty in implementing the accurate/reliable underground stratification. In this paper, Geotechnical lasso (Glasso, Shuku and Phoon 2021) is used as a main method to address this difficulty.

Glasso is a data-driven site characterization method for estimating trends and detecting layer boundaries of soil properties consistently, which is based on a machine learning method called least absolute shrinkage and selection operator (lasso). With Glasso in underground stratification, no discussion on how to select the basis functions to estimate.This is one of the great advantages compared to cases taking regression-based approaches.

Actual sites for the underground stratification are Baytown (Texas, USA) (Stuedlein et al. 2012) and Hollywood(South Carolina, USA) (Stuedlein et al. 2016). Cone penetration test (CPT) data are available in both sites, and they were converted to the soil behavior type (SBT) index for the underground stratification. The practical performance of Glasso in underground stratification was evaluated with the root-mean-square error (RMSE) of I_c and identification ratio (IR) of soil types.

The objective of this study is to implement the underground stratification to two real case histories, Hollywood and Baytown in the US. The straightforward computational implementation of this geotechnical lasso (Glasso) is not feasible in 3D problems because significant runtime and memory exceedance. In this paper, Glasso and the efficient ADMM (eADMM)(Shuku et al. 2021) are applied to address this difficulty.

2 Glasso for 3D underground stratification

This section outlines geotechnical lasso (Glasso)(Shukuand Phoon 2021), which is a lasso-based method to solve underdetermined problems in geotechnical engineering, especially for underground stratification. Underground stratification problems are usually underdetermined because the number of available data is much smaller than that of unknown parameters.In a lasso-based method, this underdetermined problem is solved by minimizing the following objective function:

$$J_{\ell 1} = \frac{1}{2}(\mathbf{y}-\mathbf{A}\mathbf{x})^{\mathrm{T}}(\mathbf{y}-\mathbf{A}\mathbf{x}) + \lambda_{\mathrm{v}}\left|\mathbf{B}_{\mathrm{v}}\mathbf{x}\right| + \lambda_{\mathrm{h}}\left|\mathbf{B}_{\mathrm{h}}\mathbf{x}\right| \tag{1}$$

where **y** is an m-dimensional observation vector, **x** is an n-dimensional (unknown) parameter vector, and **A** is an $m \times n$ matrix whose entries are 0 or 1,and $|\cdot|$ denotes absolute value or ℓ_1 norm, λ is regularization parameter that adjusts the significance of the regularization term, and **B** is a linear operator (matrix) to impose some "sparsity" in **x**. A sparsity between adjacent pixels/cells called structured sparsity can also be imposed.Lasso with the$\mathbf{D}_1$ matrix has also been applied in detecting soil layer boundaries in depth-dependent profile data (Shuku 2019; Shuku et al. 2020). A useful form of the matrix**B** includes a matrix that penalizes the discrete second derivative, $\mathbf{D}_2$ to yield a piecewise linear fit (Tibshirani 2014). These two types of structured sparsity are widely used in trend estimation and image processing.

Spatial variability is, in general, markedly anisotropic, with a large degree of homogeneity (and, hence, a stronger correlation structure) in the horizontal direction. This is because most depositional processes result in stratigraphic geometries that are not significantly distant from a set of horizontal layers with vertical thickness significantly less than the horizontal extension. Therefore, it is suitable to use different λ values for the vertical and horizontal directions, and where the subscripts "v" and "h" depict vertical and horizontal directions respectively. Minimizing Eq. (1) is subtle because of the multiple nondifferential ℓ_1 term. Therefore, we reformulate Eq. (1) as follows for ease of computation.

$$J_{\ell 1} = \frac{1}{2}(\mathbf{y}-\mathbf{Ax})^{\mathrm{T}}(\mathbf{y}-\mathbf{Ax}) + \lambda_{\mathrm{v}} \left| \begin{bmatrix} \mathbf{B}_{\mathrm{v}} \\ \lambda_{\mathrm{h}} / \lambda_{\mathrm{v}} \mathbf{B}_{\mathrm{h}} \end{bmatrix} \mathbf{x} \right| = \frac{1}{2}(\mathbf{y}-\mathbf{Ax})^{\mathrm{T}}(\mathbf{y}-\mathbf{Ax}) + \lambda_{\mathrm{v}} |\mathbf{Bx}| \tag{2}$$

where $|\cdot|$ denotes absolute value, and the dimensions of **B** matrix and **x** vector are $3m$-by-n *and* n respectively.This is a generalized lasso specialized for geotechnical problems and applicable to underground stratification and geophysical tomography (Shuku 2019).If matrix **A** is designed based on spatial interpolation methods such as a polynomial regression and kriging, the Glasso can manage irregular datasets with incomplete data, for example, soundings with unequal length and several missing values owing to equipment fault or other reasons. This is a notable advantage of the Glasso because data incompleteness is a common problem in geotechnical practice.

3 Applications to real cases

3.1 Baytown, Texas

Glasso was applied to underground stratification at an actual site in Baytown, Texas, located approximately 50 km east of Houston (Stuedlein 2012).Cone penetration tests (CPTs), nine in total, three borings (CPT-1 to 3)drilled to an approximate depth of 15 m and six borings (CPT-F1 to 6) drilled to an approximate depth of 5.5 m, were performed at this site. The layout of the CPTs is shown in Figure 2The filled square indicates CPTs for training, and the white circle indicates CPTs for validation. Thedetails of the testing program in Baytown site are given by Stuedlein (2012).

We estimated the 3D spatial distribution of soil behavior type index (I_c) and soil behavior type (SBT) using 6 CPT data. We adopted the definitions of $I_c$and SBT proposed by Robertson (2016) and Robertson and Wride (1998). The ground was modeled with 1.0m ×0.1m × 1.0m cuboids, and the total number of elements that corresponded to the number of unknown parameters n, is 50,700. We considered two types of matrices,$\mathbf{D}_1$ and $\mathbf{D}_2$, for **B** in Glasso. $\mathbf{D}_1$ and $\mathbf{D}_2$ achieved a constant and linear trend in the estimates, respectively. We write "$\mathbf{D}_1$ model" and "$\mathbf{D}_2$ model" for simplicity. We determined two regularization parameters,λ_v and λ_h using L-curve method.

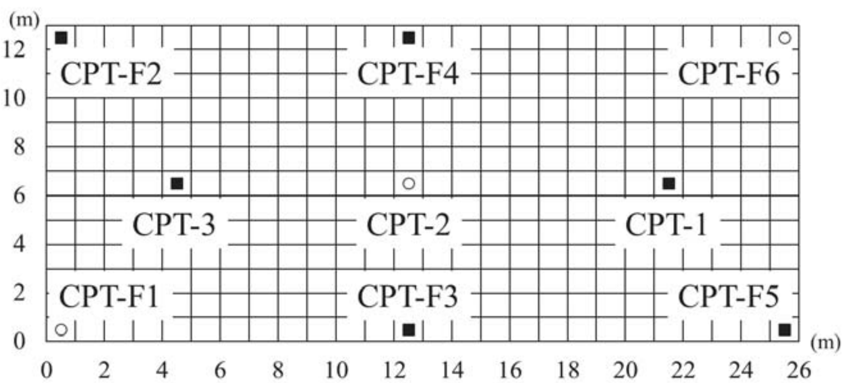


Figure 2. Layout of CPTs in Baytown site (Modified from Stuedlein 2012)

Figure 3shows a comparison of the I_c trends estimated by the Glasso to the corresponding real I_c data. The regularization parameters are set to$(\lambda_v, \lambda_h) = (0.1, 0.5)$for the D_1 and D_2 models. The red thick lines indicate estimated trends, and cross marks indicate the real $I_c$data. The profiles estimated by the $\mathbf{D}_1$ and $\mathbf{D}_2$ models record the actual data trends almost accurately. Only CPT-F1 boring estimation of $I_c$might be overfitting with other data because the estimation line was far from the actual data at the depth of 1-4m.Figure 4 shows a comparison of the SBT profiles. In the figure, the red thick lines and cross marks indicate the SBT classified by estimated I_c and actual I_c, respectively. Only the estimated SBT in CPT-F1 is a little different from the actual one at the depth of 1.5 – 3.5m.

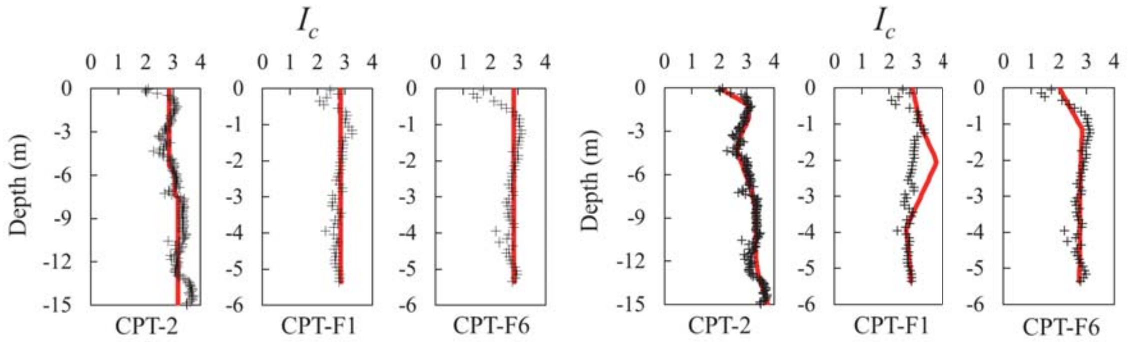


Figure 3.I_c profiles estimated by Glasso in Baytown site (left; $\mathbf{D}_1$ model, right; $\mathbf{D}_2$ model)

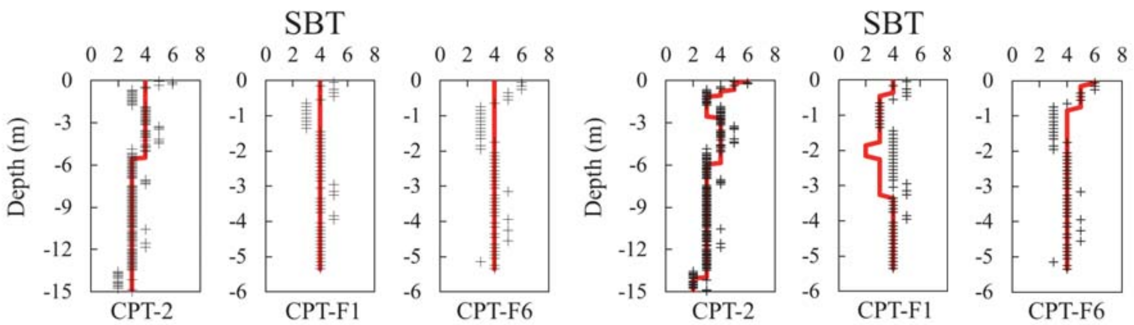


Figure 4. SBT profiles estimated by Glasso in Baytown site (left; $\mathbf{D}_1$ model, right; $\mathbf{D}_2$ model)

To quantitatively evaluate the performance of the Glasso, the root-mean-square error (RMSE) the I_c and identification ratio (IR) of SBT in Baytown site analysis are summarized in Table 1. The RMSE of the $I_c$and IR of the SBT are defined by

$$\mathrm{RMSE} = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i - x_i)^2} \tag{3}$$

$$\mathrm{IR} = \frac{1}{m}\sum_{i=1}^{m} I_i\text{,}\quad I_i = \begin{cases} 1 & \mathrm{SBT}(y_i) = \mathrm{SBT}(x_i) \\ 0 & \mathrm{SBT}(y_i) \neq \mathrm{SBT}(x_i) \end{cases} \tag{4}$$

where y_i is the actual I_c data at ith depth, x_i is the estimated I_c value at ith depth, and SBT(y_i) and SBT(x_i) are the SBTs corresponding to y_i and x_i. There are no significant differences between the results of the$\mathbf{D}_1$ and $\mathbf{D}_2$ models. The models show similar performance in terms of RMSE and IR.

Table 1.RMSE and IR of underground stratification in Baytown site

	Boring	RMSE	IR
D_1 model	CPT-2	0.250	0.693
	CPT-F1	0.245	0.685
	CPT-F6	0.383	0.574
D_2 model	CPT-2	0.175	0.680
	CPT-F1	0.426	0.519
	CPT-F6	0.236	0.630

3.2 Hollywood, South Carolina

Another actual site is located in Hollywood South Carolina, approximately 20 km west of Charleston and 20 km north of the Atlantic coast line (Stuedlein 2016). Cone penetration tests (CPTs), 25 in total, all borings drilled to an approximately depth of 13 m, were performed at this site. The layout of the CPTs is shown in Figure 5 Details of the testing program in Hollywood site are given by Stuidlein (2016).

We estimated the 3D spatial distribution I_c and SBT using 17 CPT data. The ground was modeled with 0.5m ×0.1m × 0.5m cuboids, and the total number of elements that corresponded to the number of unknown parameters n, is 60,480. Four validation sets are prepared as shown in Table 2. In this paper, we mainly show the results of validation set 1 only for simplicity.

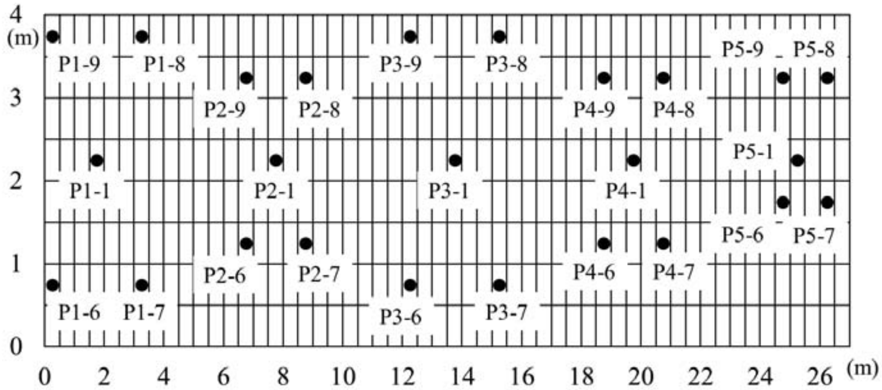


Figure 5. Layout of CPTs in Hollywood site (Modified from Stuedlein 2016)

Table 2. Validation sets in Hollywood site

Validation 1	Validation 2	Validation 3	Validation 4
P1-1	P1-1	P1-1	P1-1
P2-1	P2-1	P2-1	P2-1
P5-1	P5-1	P5-1	P5-1
P1-7	P1-6	P1-9	P1-8
P2-7	P2-6	P2-9	P2-8
P3-7	P3-6	P3-9	P3-8
P4-7	P4-6	P4-9	P4-8
P5-7	P5-6	P5-9	P5-8

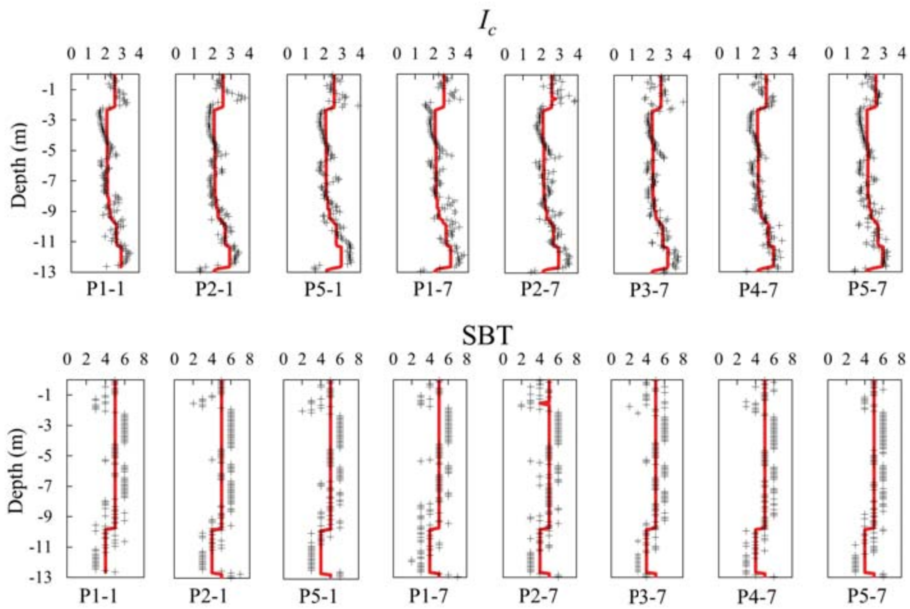


Figure 5.$I_c$and SBT profiles estimated by Glasso with $\mathbf{D}_1$ model

Figure 5shows a comparison of the I_c trends estimated by the Glasso to the corresponding real I_c data in the case of validation 1. The regularization parameters are also set to$(\lambda_v, \lambda_h) = (0.1, 0.5)$ for the $\mathbf{D}_1$ model and $\mathbf{D}_2$ model. The red thick lines indicate estimated trends, and cross marks indicate the real $I_c$data. The profiles estimated by the D_1 and D_2 models record the actual data trends accurately. Figure 6 shows a comparison of the SBT profiles. In the figure, the red thick lines and cross marks indicate the SBT classified by estimated I_c and actual I_c, respectively. The SBT estimated by the Glasso is consistent with the actual data.

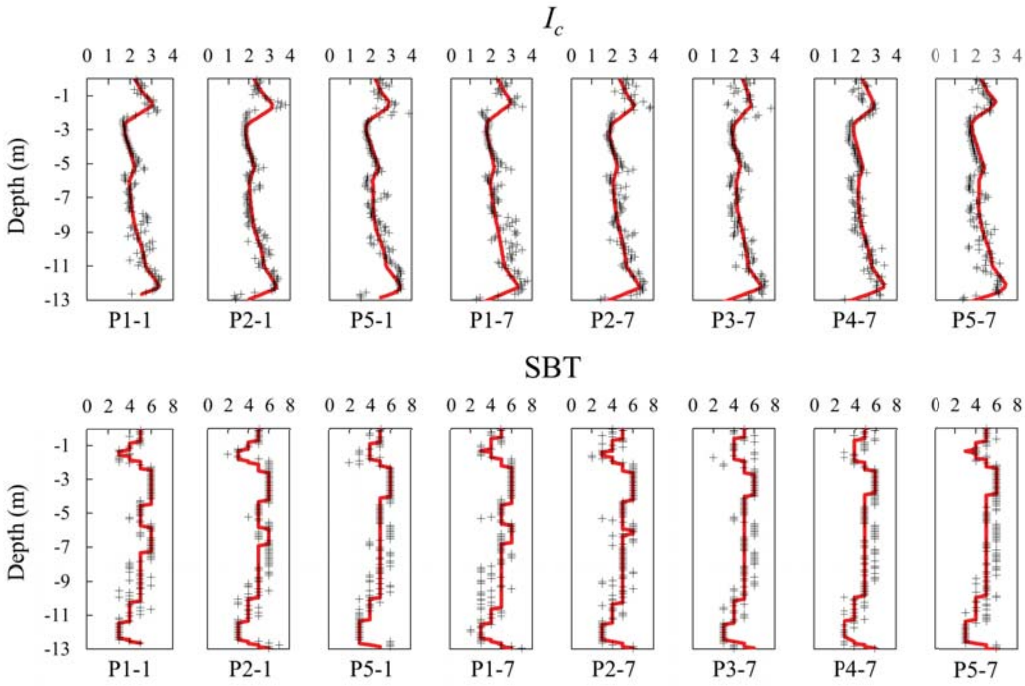
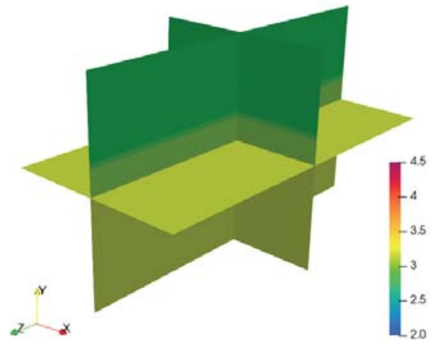


Figure 6.$I_c$and SBT profiles estimated by Glasso with $\mathbf{D}_s$model

The RMSE of the I_c and IR of SBT in Hollywood site are summarized in Table 3.The performance of the $\mathbf{D}_2$ model is better than that of the $\mathbf{D}_1$ model in terms of both RMSE and IR.**Figures 7** and **8** show the color map of I_c and SBT distribution estimated by the Glasso with the $\mathbf{D}_1$ and $\mathbf{D}_2$ models.The $\mathbf{D}_1$ model estimates simpler models compared to the$\mathbf{D}_2$ model because the matrix$\mathbf{D}_1$ assumes that$\mathbf{x}$ is constant in each layer whereas $\mathbf{D}_2$ assumes that$\mathbf{x}$ has a linear trend. Although the matrix$\mathbf{B}$ significantly affects the GLasso results, it is difficult to select an appropriate$\mathbf{B}$ matrix in advance. The selection of the $\mathbf{B}$ matrix is a typical model selection problem and can be solved using well-known information criteria, such as AIC and BIC (Bishop 2006; Murphy 2012). We can consider the matrices$\mathbf{D}_1$ and $\mathbf{D}_2$ in GLasso by reformulating Eq. (1) and (2): This can be a promising approach to manage the selection of the matrix$\mathbf{B}$.

Table 3.RMSE and IR of underground stratification in Hollywood site

		Boring	RMSE	IR		Boring	RMSE	IR
Validation 1	D_1 model	P1-1	0.315	0.346	D_2 model	P1-1	0.253	0.606
		P2-1	0.368	0.308		P2-1	0.310	0.554
		P5-1	0.384	0.357		P5-1	0.373	0.566
		P1-7	0.445	0.341		P1-7	0.391	0.462
		P2-7	0.364	0.376		P2-7	0.307	0.617
		P3-7	0.345	0.356		P3-7	0.363	0.556
		P4-7	0.301	0.405		P4-7	0.291	0.570
		P5-7	0.349	0.326		P5-7	0.338	0.530
Mean	D_1 model	P1-1	0.326	0.346	D_2 model	P1-1	0.269	0.579
		P2-1	0.379	0.308		P2-1	0.319	0.552
		P5-1	0.394	0.357		P5-1	0.363	0.552



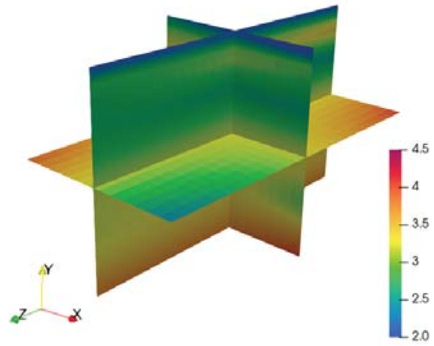
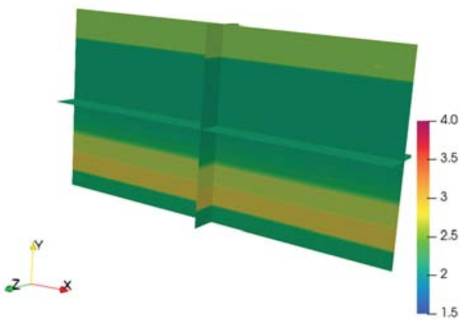


Figure 7. Colormap of I_c distributionin Baytown site (left; D_1 model, right; D_2 model)



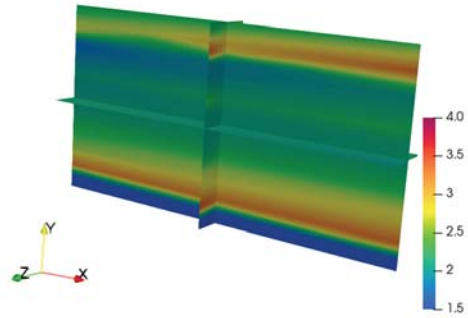


Figure 8. Colormap of$I_c$distributionof validation 1 in Hollywood site (left; D_1 model, right; D_2 model)

4 Conclusion

This study presented application examples of Glasso for two real case histories of 3D underground stratification.The applicability of Glasso was investigated by comparing the estimated results with validation dataset.The comparisons showed that Glasso can be a promising method for underground stratification in geotechnical engineering.

References

Bishop C.M. Pattern recognition and machine learning, *Springer 2006*.

Hu, Y. and Wang. Y. (2020). Probabilistic soil classification and stratification in a vertical cross-section from limited cone penetration tests using random field and Monte Carlo simulation, *Computers and Geotechnics*, 124, 103634.

Murphy K P. Machine learning: A probabilistic perspective. *MIT Press 2012*; 1067p.

Phoon, K. K., W. A. Prakoso, Y. Wang, and J. Ching. Uncertainty representation of geotechnical design parameters. In Reliability of geotechnical structures in ISO2394 2016; 49-87. *London: CRC Press*.

Robertson P K. Cone penetration test (CPT)-based soil behaviour type (SBT) classification system - an update. *Canadian Geotechnical Journal 2016*; 53(12): 1910-1927. https://doi.org/10.1139/cgj-2016-0044.

Robertson P K and Wride C E. Evaluating cyclic liquefaction potential using the cone penetration test. *Canadian Geotechnical Journal 1998*; 35: 442-459. https://doi.org/10.1139/t98-017.

Shuku, T. Sparse modeling in geotechnical engineering. *Proceedings of the 7th International Symposium on Geotechnical Safety and Risk (ISGSR) 2019*; ISBN: 978-981-11-2725-0; doi:10.3850/978-981-11-2725-0 bs3-cd.

Shuku, T., Phoon K. K. and Yoshida, I. Trend estimation and layer boundary detection in depth- dependent soil data using sparse Bayesian lasso. *Computers and Geotechnics 2020*; 128: 103845. https://doi.org/10.1016/j.compgeo.2020.103845.

Shuku, T. and Phoon, K.K. (2021). Three-dimensional subsurface modeling using geotechnical lasso, *Computers and Geotechnics*, 133(5), 104068. https://doi.org/10.1016/j.compgeo.2021.104068.

Stuedlein, A.W., Gianella, T.N., and Canivan, G.J. (2016). Densification of Granular Soils using Conventional and Drained Timber Displacement Piles. *Journal of Geotechnical and Geoenvironmental Engineering*, 142(12), 04016075. https://doi.org/10.1061/(ASCE)GT.1943-5606.0001554.

Stuedlein, A.W., Kramer, S.L., Arduino, P., and Holtz, R.D. (2012). Reliability of Spread Footing Performance in Desiccated Clay. *Journal of Geotechnical and Geoenvironmental Engineering, ASCE*, 138(11), 1314-1325. https://doi.org/10.1061/(ASCE)GT.1943-5606.0000706.

Tibshirani, R. J. Adaptive piecewise polynomial estimation via trend filtering. *Annals of Statistics 2014*; 42(1): 285-323. DOI: 10.1214/13-AOS1189.

Wang, Y., Hu, Y., and Zhao, T. (2020). CPT-based subsurface soil classification and zonation in a 2D vertical cross-section using Bayesian compressive sampling, *Canadian Geotechnical Journal*, 57(7), 947-958.