Proc. of the 9th International Symposium on Geotechnical Safety and Risk (ISGSR)
Edited by Zhongqiang Liu, Jian Dai and Kate Robinson
Published by Research Publishing, Singapore. ©2025 The Author(s), CC-BY 4.0 licence.
doi: 10.3850/GRF-25280825_isgsr-163-P304-cd

DATA-DRIVEN DEVELOPMENT OF THREE-DIMENSIONAL SUBSURFACE GEOLOGICAL MODEL FROM LIMITED BOREHOLES AND PRIOR GEOLOGICAL KNOWLEDGE FOR SITE CHARACTERIZATION

Borui Lyu

Department of Civil and Environmental Engineering, the Hong Kong University of Science and Technology, Hong Kong.
E-mail: lyuborui@ust.hk

YuWang

Department of Civil and Environmental Engineering, the Hong Kong University of Science and Technology, Hong Kong.
E-mail: wang.yu@ust.hk

The delineation of subsurface stratigraphy is an indispensable task in geotechnical site characterization, which benefits subsequent geotechnical design and construction. In recent years, three-dimensional (3D) subsurface geological modelling has received increasing attention because it can realistically represent subsurface stratigraphic variations at a specific site. Geotechnical site investigation data (e.g., boreholes) are necessary for developing 3D subsurface geological models. Due to time, budget, or technical constraints, measurements are commonly limited and sparse in engineering practice, leading to a long-lasting challenge in properly developing 3D subsurface geological models. To tackle this challenge, a generative and data-driven machine learning method called multi-scale generative adversarial network (MS-GAN) is adopted in this study. This method can properly learn 3D stratigraphic information from a 3D training image representing prior geological knowledge for automatically generating 3D subsurface geological models conditioned on limited boreholes in a data-driven manner. The performance of this method is illustrated using a simulated data example. The results indicate that MS-GAN not only accurately and effectively delineates stratigraphic variations in 3D space, but also quantifies stratigraphic uncertainty from multiple 3D realizations.

Keywords: Data-driven; 3D modelling; subsurface stratigraphy; generative adversarial network; sparse measurements; site characterization.

1. Introduction

A three-dimensional (3D) geological model is a virtual representation of subsurface stratigraphy at a specific site. It allows geotechnical engineers to easily understand complex site conditions, such as stratigraphic variations and connectivity in 3D space, and benefits geotechnical design and subsequent construction (e.g., Phoon et al. 2022). The construction of 3D geological models highly relies on geological site investigation data, such as boreholes. Due to time and budget constraints, boreholes available from a specific site or project are commonly limited and sparse. As a result, it is particularly difficult to develop a 3D geological model solely based on limited boreholes. In engineering practice, prior geological knowledge at a nearby site with similar geological settings may be available from previous projects. This leads to a question of how to effectively leverage prior geological knowledgefor 3D geological modelling?

Recently, many studies have adopted multiple point statistics (MPS) for geological modelling from boreholes(e.g., Shi and Wang 2021a). For MPS-based methods, spatial geological prior knowledge is represented as training images for integration with site-specific measurements and conditional stochastic simulation using a set of MPS parameters tuned before the simulation (e.g., Tahmasebi, 2018). In addition to MPS, emerging machine learning methods offer a promising way to effectively integrate measurements from various sources due to their capability to identify intrinsic patterns and complex associations within the training data. Of these methods, Lyu et al. (2024) developed multi-scale generative adversarial networks (MS-GAN) to build 3D geological models with quantified uncertainty conditioned on limited boreholes and a 3D training image representing geological prior knowledge in a data-driven manner. In this study, a framework based on MS-GAN is proposed for data-driven development of 3D geological models from limited boreholes and prior geological knowledge.

The remainder of this study is organized as follows. Section 2 presents the framework for iterative generation of 3D subsurface geological models. In Sections 3, a simulated data example is used for illustrating the proposed framework, and the results are evaluated in terms of accuracy, uncertainty, and computational efficiency. Finally, some concluding remarks are provided in Section 4.

2. Proposed Data-driven Framework for 3D Subsurface Geological Modelling

This section proposes a data-driven framework for integrating site-specific boreholes with prior geological knowledge at a nearby site to delineate subsurface stratigraphy. A recently developed data-driven method, MS-GAN (Lyu et al. 2024), is adopted for the integration of boreholes and prior geological knowledge. As shown in

Fig. 1, Site A and Site B are two nearby sites sharing similar geological settings. Based on MS-GAN, 3D subsurface geological models at Site B are generated with the aid of prior geological knowledge at Site A, which is represented as a 3D training image, and conditioned on limited borehole logs observed from Site B in a data-driven manner. In the following two subsections, basic idea of MS-GAN for 3D geological modelling and uncertainty quantification of MS-GAN results are presented.

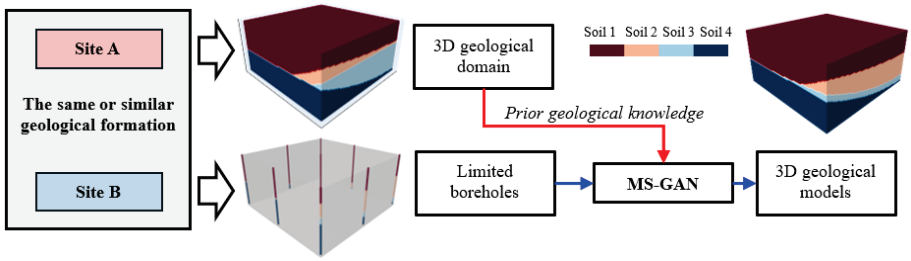


Fig. 1 Proposed framework for integrating limited boreholes with prior geological knowledge for data-driven development of 3D geological model.

2.1. Multi-scale generative adversarial network (MS-GAN)

Fig. 2 illustrates the basic architecture of MS-GAN, which consists of multiple GAN models. GAN (Goodfellow et al. 2014) is an emerging machine learning method for generating realistic samples (e.g., images) from training data. With limited boreholes as input, it is extremely difficult to directly generate a complete 3D model. In the context of MS-GAN, a 3D geological model is iteratively generated using multiple GAN models from different scales. Specifically, the role of the GAN model at each scale is to generate stratigraphic information in depth direction at only a limited number of prescribed locations, rather than the stratigraphic information at entire 3D space. For convenience, the prediction at each scale is referred to as predicted boreholes.

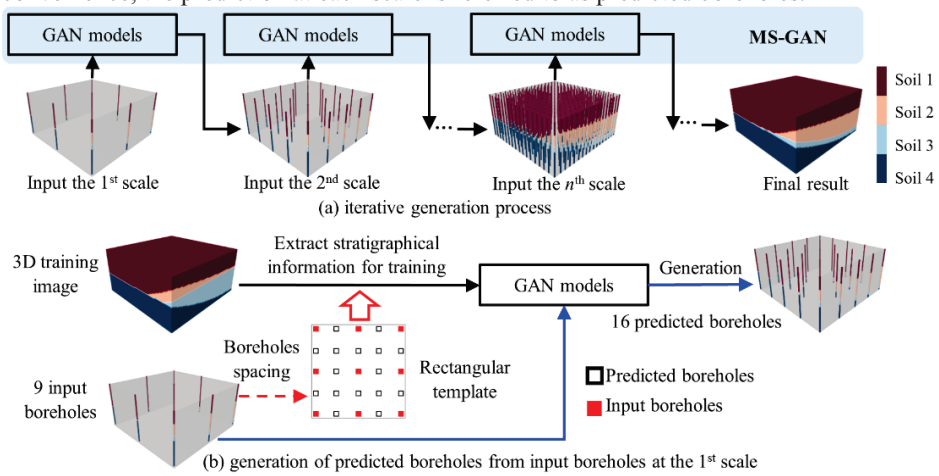


Fig. 2 MS-GAN for generating 3D subsurface geological models (after Lyu et al. 2024): (a) iterative generation process; and (b) generation of predicted boreholes from input boreholes at the 1st scale.

As shown in Fig. 2a, the available boreholes from a site of interest are adopted as the input boreholes at the 1st scale, and corresponding GAN models at the 1st scale are trained to generate predicted boreholes. In MS-GAN, the number and location of predicted boreholes is determined by a 5×5 rectangular template. As shown in Fig. 2b, the rectangular template comprises 9 input boreholes (red solid squares) and 16 predicted boreholes (empty squares). The horizontal distribution of input boreholes is completely dependent on horizontal spacing between input boreholes, while each predicted borehole is located at the midpoint between any two neighboring input boreholes. Based on horizontallayout of input boreholes and predicted boreholes, corresponding stratigraphic information is extracted from the 3D training image to train GAN models at the 1st scale, for generating 16 predicted boreholes from 9 input boreholes, as shown in Fig. 2b. After generation process at the 1st scale, the predicted boreholes together with input boreholes at the 1st scale are adopted as the input boreholes at the next scale. This results in a decrease in the horizontal distance between adjacent input boreholes. For generation process at the next scale, a 5×5 rectangular template is constructed based on horizontal spacing of corresponding input boreholes. Following the same process as the 1st scale, GAN models at the 2nd scale are trained to generate 16 predicted boreholes from 9 input boreholes. As shown in Fig. 2a, the iterative process results in a continuous decrease in the horizonal spacing of the boreholesfor developing a complete 3D geological model. It is worth

noting that the input boreholes at each scale should be amenable to the layout of defined by the 5×5 rectangular template.

When available boreholes are irregularly distributed, a series of sub-templates are developed to construct regularly distributed boreholes from available boreholes. Unlike the 5×5 rectangular template shown in Fig. 2b, each sub-template for developing regularly distributed boreholes is used to generate one predicted borehole from its nearby available boreholes. The establishment of sub-templates is determined by the layout of site-specific boreholes. For example, Fig. 3 shows the horizontal layout of 7 irregularly distributed boreholes. Two sub-templates (i.e., sub-template A & B) are established based on the borehole layout in this site for constructing regularly distributed boreholes from available boreholes. As shown in Fig. 3, each sub-template is used to generate one borehole from neighbouring two boreholes. Based on these sub-templates, MS-GAN can generate a complete 3D subsurface geological model from input boreholes in any horizontal layout.

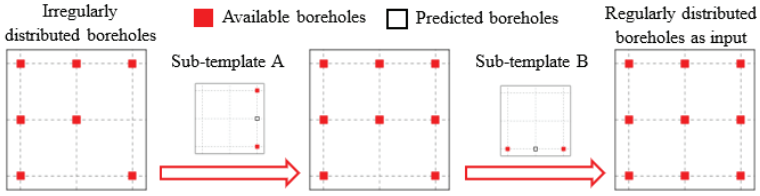


Fig. 3 Examples of sub-templates for development of regularly distributed boreholes from irregularly distributed boreholes.

2.2. Stratigraphic uncertainty modelling by MS-GAN

After training process at each scale, MS-GAN can generate multiple 3D realizations of subsurface stratigraphy conditioned on limited boreholes. The most probable prediction (MPP) of 3D realization is considered as the final result of MS-GAN. It is developed by assigning each voxel $\boldsymbol{x}$ with soil type of the highest frequency among all 3D realizations. Based on the MPP and N_r 3D realizations, the stratigraphic uncertainty at each voxel $\boldsymbol{x}$ can be measured by dispersion, Dp($\boldsymbol{x}$):

$$\mathrm{Dp}(\boldsymbol{x}) = \frac{\sum_{k=1}^{N_r} I_k}{N_r} \quad I_k = \begin{cases} 1 & Z_k(\boldsymbol{x}) \neq Z_{\mathrm{MPP}}(\boldsymbol{x}) \\ 0 & Z_k(\boldsymbol{x}) = Z_{\mathrm{MPP}}(\boldsymbol{x}) \end{cases} \tag{1}$$

where $Z_k(\boldsymbol{x})$ and $Z_{\mathrm{MPP}}(\boldsymbol{x})$ refer to the soil type at each voxel $\boldsymbol{x}$ of the k^{th} 3D realization and the MPP, respectively. When a 3D ground truth subsurface geological domain is available (e.g., the simulated data example to be shown in Section 3), accuracy of the MPP, Acc, can be computed as follows:

$$\mathrm{Acc} = \frac{\sum_{i}^{N_X \times N_Y \times N_Z} I_i}{N_X \times N_Y \times N_Z} \quad I_i = \begin{cases} 1 & Z_T(\boldsymbol{x}_i) = Z_{\mathrm{MPP}}(\boldsymbol{x}_i) \\ 0 & Z_T(\boldsymbol{x}_i) \neq Z_{\mathrm{MPP}}(\boldsymbol{x}_i) \end{cases} \tag{2}$$

where $N_X \times N_Y \times N_Z$ refers to the total number of voxels; $Z_T(\boldsymbol{x}_i)$ and $Z_{\mathrm{MPP}}(\boldsymbol{x}_i)$ denote the soil type at voxel $\boldsymbol{x}_i$ of the 3D ground true geological domain and the MPP, respectively.

3. Illustrative Example

In this section, the proposed framework is illustrated using two 3D geological domains provided by Shi and Wang (2022). As shown in Fig. 4, both 3D geological domains contain four soil types and have a volume of 100m×100m×50m. Using a resolution of 1m along three directions, each 3D geological domain is discretized into 101×101×51=520251 voxels, each storing corresponding soil types. In this example, the 3D geological domain in Fig. 4a is used for testing, and 20 irregularly distributed boreholes are extracted from the 3D testing geological domain as input boreholes at the 1st scale (see Fig. 4c). Using the same resolution, the 20 input boreholes comprise 20×51=1020 voxels, which only occupy 1020/520251 = 0.2% of the 3D testing geological domain. The remaining 3D geological domain shown in Fig. 4b is adopted as the 3D training image for training MS-GAN models. Conditioned on 20 input boreholes, 100 3D realizations were generated based on trained GAN models, and the MPP of these realizations was developed accordingly. The simulation time for each 3D realization is about 10 minutes.

Fig. 5a shows three outer surfaces of the MPP for comparison with corresponding outer surfaces of the 3D testing geological domain shown in Fig. 4a. It is observed that the three outer surfaces of the MPP and the 3D testing geological domain share similar stratigraphic connectivity. Using Eq. (2), the accuracy of the MPP is computed as 88%. In addition, the stratigraphic uncertainty of the MS-GAN results is quantified using Eq. (1). Fig. 5b shows three outer surfaces of the 3D dispersion. It can be observed that boundaries between different soil types tend to have large dispersion values. This is observation is consistent with previous studies (e.g., Shi and Wang 2021b). In addition to three outer surfaces, Fig. 5c and Fig. 5d respectively show a cross-section at Y = 70m extracted from the MPP and 3D dispersion, together with red lines representing the true stratigraphic

boundaries. The accuracy of the cross-section in Fig 5c is 92%. It is also observed in Fig. 5d that large dispersion values accumulate around the true stratigraphic boundaries.

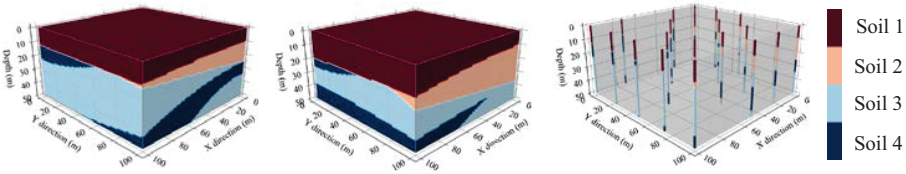


Fig. 4 Simulated 3D geological models (after Shi and Wang 2022): (a) 3D geological model for testing; (b) 3D geological model for training; (c) 20 boreholes extracted from (a).

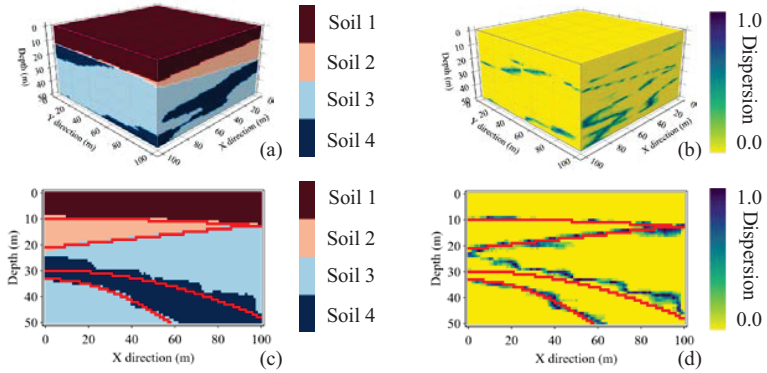


Fig. 5 MS-GAN results conditioned on 20 input boreholes: (a) the most probable prediction (MPP); (b) 3D dispersion; (c) cross section of MPP at Y = 70m; and (d) cross section of dispersion at Y = 70m.

4. Conclusion

This study presented a framework for data-driven development of 3D geological models by integrating limited boreholes and prior geological knowledge using a data-driven machine learning method, MS-GAN. This method can effectively extract 3D stratigraphic connectivity from a 3D training image representing prior geological knowledge to supplement limited boreholes for 3D geological modelling. The proposed framework was illustrated by a simulated data example. The results indicate that MS-GAN accurately and efficiently generates 3D subsurface geological models with quantified stratigraphic uncertainty from sparse boreholes.

Acknowledgement

The work described in this paper was supported by a grant from the Research Grant Council of Hong Kong Special Administrative Region (Project no. 11202121).

References

Goodfellow, I., J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems*, 27.

Lyu, B., Y. Wang, and C. Shi. (2024). Multi-scale generative adversarial networks (GAN) for generation of three-dimensional subsurface geological models from limited boreholes and prior geological knowledge. *Computers and Geotechnics*, 170, 106336.

Phoon, K. K., J. Ching, and T. Shuku (2022). Challenges in data-driven site characterization. *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, 16(1), 114–126.

Shi, C. and Y. Wang (2021a). Nonparametric and data-driven interpolation of subsurface soil stratigraphy from limited data using multiple point statistics. *Canadian Geotechnical Journal*, 58(2), 261-280.

Shi, C. and Y. Wang (2021b). Development of subsurface geological cross-section from limited site-specific boreholes and prior geological knowledge using iterative convolution XGBoost. *Journal of Geotechnical and Geoenvironmental Engineering*, 147(9), 04021082.

Shi, C. and Y. Wang (2022). Data-driven construction of Three-dimensional subsurface geological models from limited Site-specific boreholes and prior geological knowledge for underground digital twin. *Tunnelling and Underground Space Technology*, 126, 104493.

Tahmasebi, P. (2018). Multiple Point Statistics: A Review, In B. Daya Sagar, Q. Cheng, and F. Agterberg (Eds), *Handbook of Mathematical Geosciences*, pp. 613-643. Springer.