EG-ICE 2025
GLASGOW

Leveraging Probabilistic Machine Learning for Subsoil Modelling to Estimate Excavated Material Volumes

DAFYDD COTOARBĂ[1], DANIEL STRAUB[2,3], IAN FC SMITH[1]

[1]TUM Georg Nemetschek Institute, Technical University of Munich, Munich, Germany
[2] Engineering Risk Analysis Group, Technical University of Munich, Munich, Germany
[3] Munich Data Science Institute, Technical University of Munich, Garching, Germany

ABSTRACT
The construction of buildings and infrastructure often involves the excavation and transportation of large volumes of soils, which contributes significantly to project costs and environmental impacts. A major challenge in planning the management of excavated materials is the significant uncertainty in quantifying soil volumes before excavation. This uncertainty arises from the limited availability of data, such as borehole soundings and cone penetration tests, as well as the reliance on deterministic modelling approaches. Recent advancements in probabilistic machine learning have enabled the training of models to create probabilistic 3D subsoil models, which explicitly account for uncertainties. In this study we investigate the impact of model selection, such as choice of kernel and hyperparameters, on the resulting 3D geological models. Additionally, we explore how these probabilistic models can improve the management of excavation materials. In a study based on the design of an excavation pit for a metro station in Munich, Germany we show first results on how probabilistic models can inform early decisions regarding machine fleet composition and time estimations.

KEYWORDS
Probabilistic digital twin, machine learning, geological modelling, management of excavated materials

1. INTRODUCTION

Construction projects often require management of significant volumes of excavated materials and can represent a significant portion of overall project costs. For example, in infrastructure projects, managing excavated materials and soil can account for up to 30% of total project costs while substantial CO_2 emissions are generated (Magnusson *et al.* 2015). Consequently, optimizing the management of excavated materials offers substantial potential for reducing costs and environmental impacts (Kenley and Harfield, 2011).

Material reuse is the main strategy to minimize the transportation of excavated soils to distant landfills and to reduce CO_2 emissions. Reuse can occur either on-site or in other projects, potentially after pretreatment (Magnusson et al., 2015). Decisions regarding material reuse require detailed planning and depend on the geotechnical properties (e.g., particle size, density, and deformation characteristics) and the chemical composition (e.g., contamination) of the excavated materials.

A major challenge in managing excavated materials is the significant uncertainty in quantifying soil volumes before excavation. This uncertainty results from the limited availability of subsurface data, such as borehole soundings and cone penetration tests, as well as the reliance on deterministic modeling approaches. These traditional methods fail to capture the inherent

variability in subsurface conditions, potentially leading to inefficiencies in excavation planning.

Recent advancements in probabilistic machine learning (ML) have enabled the development of probabilistic 3D subsoil models, which explicitly account for uncertainties (e.g., Ching et al., 2020, 2023; Shuku and and Phoon, 2023; Qian and Shi, 2024; Zinas et al., 2025). These models are particularly valuable within the framework of probabilistic digital twins, where traditional digital twin approaches are extended to incorporate uncertainty quantification (Cotoarbă et al., 2024).

Despite these advancements, implementations of probabilistic modeling in geotechnical design and construction remain limited. Additionally, research is required to investigate how such models can be used to improve excavation material management. For instance, quantifying uncertainty can support risk-based optimization strategies, leading to more informed decision-making in excavation planning.

The objective of this study is to explore how 3D probabilistic geological models can enhance excavation material management by improving volume estimation and supporting data-driven planning decisions. Additionally, we investigate the impact of ML model parameters on prediction outcomes.

The study is based on the design of an excavation pit for a metro station in Munich, Germany (see sketch in Figure 1). Site-specific soil information is available in the form of borehole soundings.

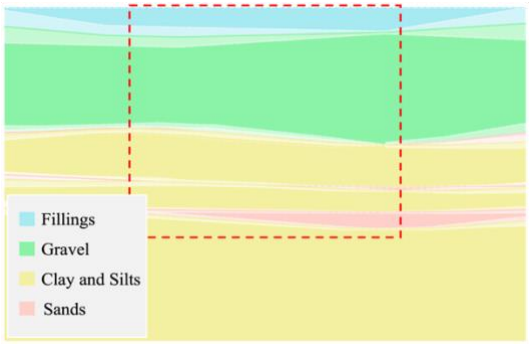

Figure 1. A probabilistic soil model for a case study in Munich, illustrating the uncertainty in boundary locations and highlighting the dimensions of the excavation pit.

2. MACHINE LEARNING METHODS FOR 3D GEOLOGICAL MODELING

For planning the excavation of an excavation pit, an investigation is required to estimate the expected soil types and volumes for a given area of interest. Investigations are performed through sparse measurements from, e.g., borehole soundings or cone penetration tests. ML methods for 3D geological modeling take such data as input to train a model to predict soil types at unknown locations. Due to data sparsity, this task is challenging, resulting in many locations where confident classification is impossible. In such cases, it is recommended to use models that output a probability distribution over possible classes rather than a single prediction (Murphy, 2022). Probabilistic approaches to machine learning include Bayesian neural networks, Gaussian Processes, and Bayesian deep learning (Phoon et al., 2019). Subsoil modeling can be categorized into two main approaches, each leveraging machine learning techniques differently:

a) Voxel-based approach in which the area of interest is discretized into a grid of voxels, and probabilistic *machine learning* techniques *for classification* are used to predict soil types at unknown locations. Approaches include Markov Random Fields (Shuku and and Phoon, 2023; Qian and Shi, 2024) or Gaussian Process Regression (Zinas et al., 2025).

b) Layer-based approach where probabilistic machine learning for regression is used to infer the parameters that describe the boundaries between soil layers. This approach to subsoil modeling is intuitive and computationally efficient as it uses a volume representation of layers. However, it may deliver lower accuracy for complex soil conditions, as the simplifying assumptions about layer boundaries may fail to capture the true variability of the subsurface. Approaches differ in their degree of of automation and types of functions used to describe surfaces (e.g., Liu et al., 2023, 2021; Lyu et al., 2023, 2021).

In this study, we use the surface-based approach for geological modeling. The input data are borehole soundings, which provide categorical data by observing soil types over depth (see *Figure 2b*). The goal is to learn a set of functions that describe the boundaries of each layer from the data (see *Figure 2a*). To obtain a probabilistic prediction of the surface functions, we use Gaussian process regression (GPR).

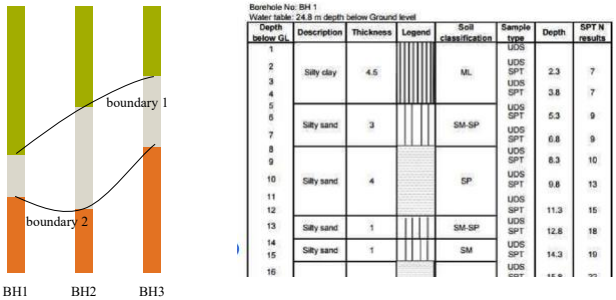

Figure 2.a) sketch of the surface-based approach; b) Example snippet from a borehole-sounding report

Gaussian Process Regression

Gaussian Process Regression (GPR) is a non-parametric probabilistic modeling approach, which is also known as Kriging and is widely used for spatial interpolation tasks. A GP is fully described by its mean function $m(\boldsymbol{x})$ and covariance function (or kernel function) $k(\boldsymbol{x}, \boldsymbol{x}')$ (Rasmussen and Williams, 2005):

$$m(\boldsymbol{x}) = E[f(\boldsymbol{x})],$$
$$k(\boldsymbol{x}, \boldsymbol{x}') = E[\big(f(\boldsymbol{x}) - m(\boldsymbol{x})\big)\big(f(\boldsymbol{x}') - m(\boldsymbol{x}')\big)]$$

$k(\boldsymbol{x}, \boldsymbol{x}')$ determines the smoothness and variability of the function. The joint distribution of the function value is multinormal.

For the proposed problem of geological modeling from borehole soundings, we define $\boldsymbol{x} \equiv (x, y)$ and $f(\boldsymbol{x}) \equiv z$ for each layer boundary.

Posterior

The training dataset is defined as $\{\boldsymbol{x_i}, z_i\}_{i=1}^{N}$. For the measurement error it is common to assume a zero-mean Gaussian mean with unknown variance. Thus, the covariance of the observation is given as

$$k_z(z_i, z_j) = k(\boldsymbol{x_i}, \boldsymbol{x_j}) + \sigma_n^2 \delta_{ij}$$

Where δ_{ij} is the Kronecker delta for which $\delta_{ij} = 1$ if $z_i = z_j$ and 0 else.

The joint Gaussian distribution over observed outputs z at locations X and predicted values f^* at discrete locations X_* can be written as:

$$\begin{bmatrix} z \\ f^* \end{bmatrix} \sim N\left(0, \begin{bmatrix} K(X,X) + \sigma_n^2 I & K(X,X_*) \\ K(X_*,X) & K(X_*,X_*) \end{bmatrix}\right)$$

where $K(X_*, X)$ is the covariance matrix between locations X_* and X, $K(X_*, X_*)$ is the covariance matrix of X_*, and σ_n^2 is the measurement noise variance.

Applying Bayesian inference, the conditional posterior Gaussian distribution at locations X_* has the following mean and convariance matrix:

$$f^* = K(X_*, X)[K(X,X) + \sigma_n^2 I]^{-1} z,$$

$$\Sigma_* = K(X_*, X_*) - K(X_*, X)[K(X,X) + \sigma_n^2 I]^{-1} K(X, X_*)$$

Covariance Function and Hyperparameters

The choice of the covariance function can influence model performance. One commonly used option is the Matérn covariance function describes the correlation between two points x and x' as follows

$$k(x, x') = \frac{1}{\Gamma(\nu) 2^{\nu-1}} \left(\frac{\sqrt{2\nu}}{l} d(x, x') \right)^{\nu} K_\nu \left(\frac{\sqrt{2\nu}}{l} d(x, x') \right).$$

Where $\Gamma(\nu)$ is the gamma function, K_ν is the modified Bessel function of second kind, $d(x, x')$ is the Euclidean distance between points. ν is a a parameter controlling the smoothness of the kernel and is fixed beforehand, and l is the length-scale and is a hyperparameter. The smoothness of the function decreseases as ν becomes smaller. Special cases of the matern include:

- $\nu = \frac{1}{2}$: the Matérn becomes the equivalent of to the absolute exponentation kernel
- $\nu \to \infty$: the kernel converges to the Radial Basis Function (RBF) kernel.

The kernel hyperparameters were learned by maximizing the log-likelihood using the Limited-Memory BFGS optimization algorithm, as implemented in the python GPy library (GPy, 2012). In this process, the hyperparameters are adjusted to maximize the likelihood of the observed data given the model. For each identified geological layer in the borehole data, a separate Gaussian Process model is trained. Each layer is defined by a set of (x, y, z)-coordinates. The training data consist of known z-values at borehole locations, and the model is trained to predict z at previously unobserved locations (see *Figure 2a*). An example 3D subsoil model from three boreholes and using an RBF kernel is shown in *Figure 3*.

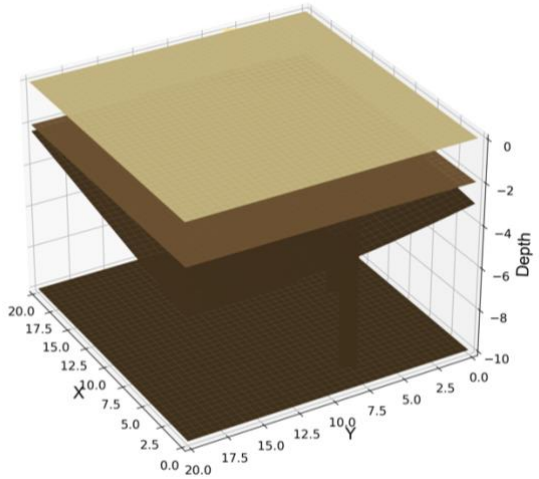

Figure 3. One sample from the model generated from three borehole soundings and the RBF kernel

3. EXCAVATION SIMULATION MODEL

Traditionally, machine fleet requirements have been estimated using deterministic models and approaches. However, due to significant uncertainties, the accuracy of these methods is limited, often leading to differences between initial estimates and actual excavation duration and costs (Zhou et al., 2019; Heravi et al., 2021). Typically, parameters for excavation-related calculations are derived from empirical data and include factors such as machine fuel consumption, available equipment, haul distance, road quality, and operator experience. The interdependencies among these factors and their effects on key quantities of interest

(duration, cost, and emissions) are conceptually illustrated in *Figure 5*. We classify input parameters into three main categories:

1) Site Parameters

Key site parameters that influence excavation productivity include soil type and excavation volume. The soil type dictates excavation difficulty, material weight, and the required attachments for the excavator (e.g., rock breaker required to break rocks).

Additionally, the location of the construction site affects transportation distance, haul time, and road conditions and directly impacts soil transport efficiency. Lastly, the experience level of available machine operators will affect the overall efficiency related to machine operations.

2) Excavators

Parameters influencing excavator performance include the number of available machines, bucket size, and the time required to complete a full operational cycle. The operational cycle consists of moving the bucket from its starting position to excavation, unloading into the truck, and returning to the starting position.

The excavation productivity for a given soil type and operator is calculated as

$$\begin{aligned} &Excavation\ Productivity \\ &\quad = Cycles\ per\ Hour\ \times Bucket\ Capacity \\ &\quad \times Cycle\ Efficiency \\ &\quad \times Operator\ Efficiency \\ &\quad \times Availability[\frac{m^3}{hour}] \end{aligned}$$

Additionally, factors such as hourly costs, fuel type, and consumption directly impact overall costs and emissions.

3) Trucks

For trucks, relevant parameters also include fleet size and individual truck performance. Key time metrics include the duration required for loading and unloading operations, as well as the truck loading capacity, which is constrained either by weight or volume capacity. Truck productivity is determined as:

$$\begin{aligned} &Truck\ Productivity \\ &\quad = Cycles\ per\ Hour\ \times Actual\ Capacity \\ &\quad \times Operator\ Efficiency \\ &\quad \times Availability[\frac{m^3}{hour}] \end{aligned}$$

Costs and emissions are additionally influenced by fuel consumption and hourly usage costs.

Time

The total excavation hours required per machine can be determined by allocating a portion of the excavation volume to each machine and dividing it by its productivity. Machine productivity is influenced by the soil type present on-site, and the required excavation hours can be calculated as follows:

$$Hours\ Needed\ (Machine) = \frac{Volume}{Machine\ Productivity}$$

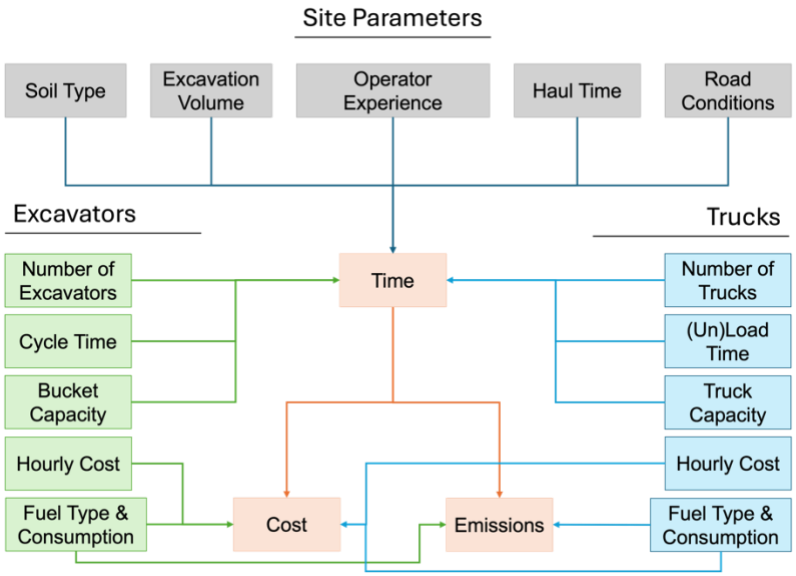

Figure 5. Dependencies of the key quantities of interest, time, cost and emissions, on parameters for an excavation project.

Cost Estimation

The cost associated with one machine for either excavation or hauling operations is calculated as follows:

$$Cost_{machines} = time * machine\ hourly\ cost$$

where the hourly cost includes direct usage time, operational costs, and fuel consumption. The total cost is given by sum of individual costs.

Emissions Estimation

Environmental impact assessments can be performed by estimating CO2 emissions based on fuel consumption:

$$EmissionsMachine = hours_{needed} * hourly\ fuel\ consumption * co2\ emissions\ per\ liter$$

4. METHODOLOGY

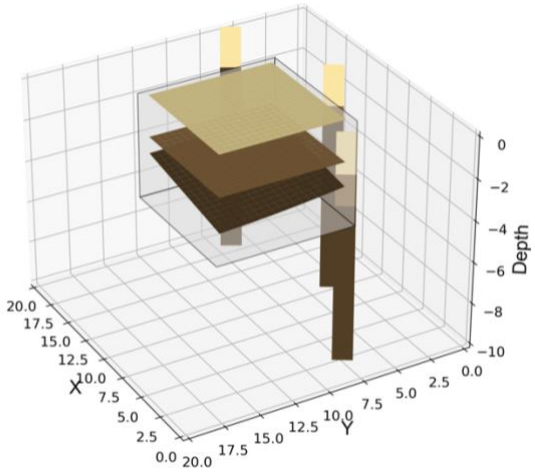

Figure 6. Example output from the Monte Carlo simulation, showing borehole data as vertical bars and the corresponding inferred soil layers within the area of interest.

Using the GPR approach for geological modeling introduced in *Section 2*, a distribution of expected soil types within the excavation area can be obtained.

A straightforward method for conducting a probabilistic analysis of excavation parameters is to perform Monte Carlo simulations, generating n_s samples of 3D subsoil models and calculating excavation time, costs, and emissions for each sample, as described in *Section 3*. An example of one such sample is illustrated in *Figure 6*. By repeating this process for n_s samples, a probability distribution of the quantities of interest can be obtained, capturing the inherent uncertainty within the geological model.

5. ILLUSTRATIVE CASE STUDY

The case study is inspired by the construction of a subway station in Munich, Germany, introduced in the work of Pelz (2010). The station was built using the cut-and-cover method and features a total length of 202 meters, an excavation depth of 15.9 meters, and an embedment depth of 7.6 meters

We limit the excavation width to 90 meters for Illustration purposes. The spatial dimensions of the modeled area and the excavation pit are summarized in *Table 1* and illustrated in *Figure 7.*

Table 1. Dimensions of the modeled soil area and excavation pit

	X [m]	Y [m]	Z [m]
Model	100	50	30
Excavation Pit	90	20	20

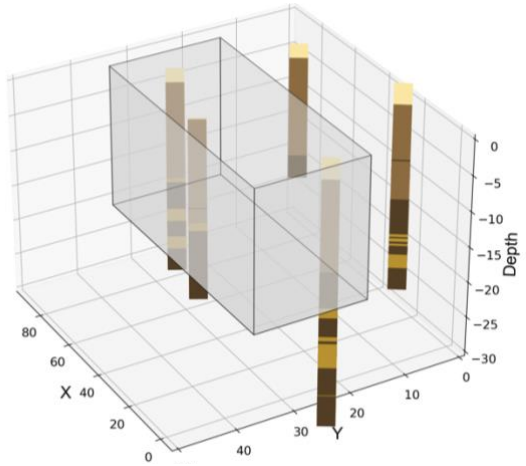

Figure 7. Visualization of the excavation pit and the available boreholes

Geotechnical Data

Soil Type	USCS Class	Excavation Factor	Weights $\gamma/\gamma'[kN/m^3]$	Required Attachment
Filling	SM	1.0	20/11	Bucket
Gravel	GW	1.2	23/14	
Clays and Silts	CL	1.0	20/11	
Sands	SW	0.9	21/12	

Table 2. Geotechnical parameters for the four soil types encountered in the Munich case study.

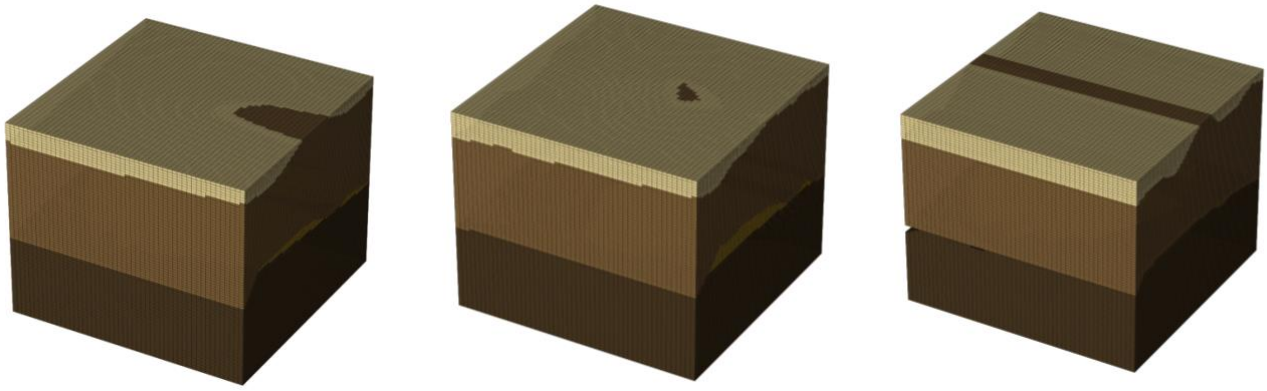

Figure 8. Comparison of 3D subsoil layer models generated using Gaussian process regression with different kernel functions: (a) Radial Basis Function kernel and (b) Matérn kernel with $\boldsymbol{\nu = 3/2}$ and (c) Matérn Kernel with $\nu = 5/2$.

The geological model of the site is obtained based on five borehole soundings from locations near the construction site (see *Figure 7)*. They are extracted through the Bavarian State Office for the Environment and is publicly available in the UmweltAtlas database (Bayerisches Landesamt für Umwelt, Augsburg, Germany).
From these borehole soundings, four primary soil types were identified, which are characteristic of the subsurface conditions of Munich: fillings, quaternary gravel, tertiary clays and silts, and tertiary sands. The relevant geotechnical parameters, including excavation factors, weights, and required attachments, are summarized in *Table 2*.

3D Subsoil Model

As described in *Section 2*, a 3D geological model is generated using the surface-based approach and Gaussian process regression. To compare how different kernels apply, we apply three different kernel functions for geological modeling: Radial Basis Function (RBF), Matérn Kernel with ($\nu = 3/2$), and Matérn Kernel ($\nu = 5/2$). For each, we generate $n_s = 1000$ samples of soil profiles. One example for each can be seen in *Figure 8*.

Table 1. Excavator parameters are taken to resemble a Liebherr R920 Compact excavator

Parameter	Unit	Value
Bucket Capacity	$[m^3]$	0.95
Hourly Cost	$[EUR/h]$	125
Fuel Consumption	$[L/h]$	15
CO_2 Emissions	$[kg/L]$	2.54
Attachments	-	Bucket

Fleet Scenario

A hypothetical fleet scenario is defined to analyze excavation and hauling efficiency. In this study, we consider the following configuration of available excavators (N $\in$ {2,3,4}) and trucks ($M \in \{1 * n_{trucks}, 2 * n_{trucks}, 3 * n_{trucks}\}$). A scenario analysis is conducted for the different fleet configurations, resulting in a total of 9 scenarios.
The characteristics of the excavators used in the scenario analysis are detailed in *Table 3*. The specifications of the truck are outlined in *Table 4*. The operator efficiency and cost are modeled using skill-based cycle factors and hourly wage rates, as shown in *Table 5*. A more complex model could include a factor connected to the years of experience.

Table 2. Operator parameters to calculate productivity and costs.

Experience	Efficiency	Hourly Rate [EUR]
Beginner	0.6	50
Intermediate	0.8	60
Expert	0.95	70

Table 3. Truck parameters to calculate productivity, costs and emissions

Parameter	Unit	Value
Truck Capacity	$[m^3]$	13
Hourly Cost	$[EUR/h]$	100
Fuel Consumption	$[L/km]$	34
CO_2 Emissions	$[kg/L]$	2.54
Max. Payload	$[tonnes]$	25
Max. Speed Empty/Loaded	$[km/h]$	40/50

Although real-world excavation operations involve additional complexities, such as variations in machine productivity, mechanical failures, and operator efficiency, each scenario in this study assumes an equal distribution of the total excavation volume among the available excavators and trucks.

6. RESULTS

For this analysis, the key metrics evaluated are excavation time, costs, and emissions, computed using the methodology outlined in Section 3. Calculations are conducted for $n_s = 1000$ samples per kernel function.

Comparison of Kernels

To provide first insights into how model selection influences excavation planning, we assess the impact of different kernel functions on estimated costs. In *Figure 9,* the distribution of soil types for each kernel function is illustrated. Although the RBF and Matérn kernels yield different shape approximations (*Figure 8*), it can be seen that the overall volume distribution of soil types remains largely consistent.

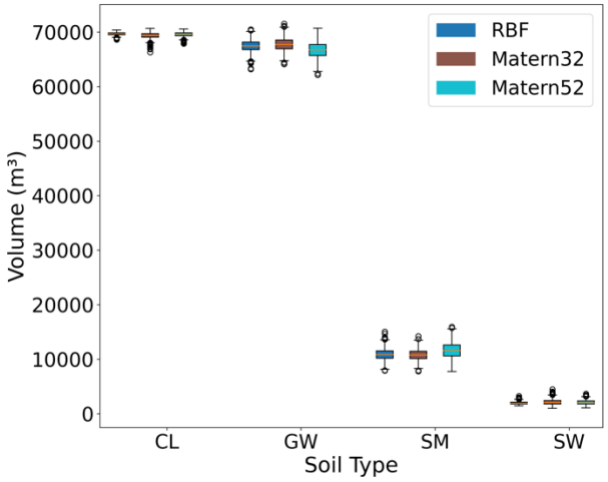

Figure 9. Distribution of volume of soil types for different kernels

Figures 10, 11, and 12 illustrate the estimated excavation and transportation time, costs, and emissions for a fleet configuration of three excavators, six trucks, and intermediate-level operators. As expected, given the similarity in soil volume distributions, no significant differences in the expected duration, costs, or emissions are observed. The results also indicate low uncertainty in the final estimates, which can be attributed to two factors: 1) the low variability in soil type volumes, and 2) the similarity of excavation properties of the encountered soil layers.

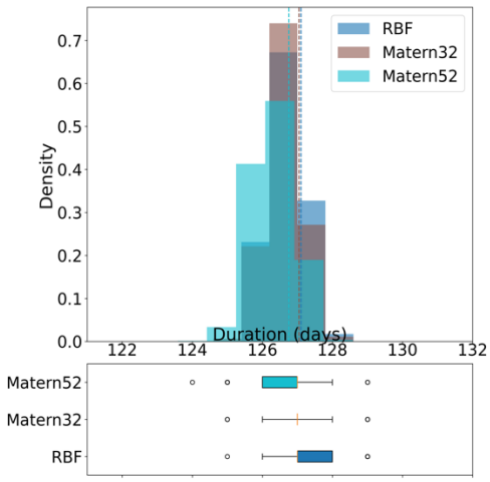

Figure 10. Expected duration of excavation works for various kernels and 1000 samples each

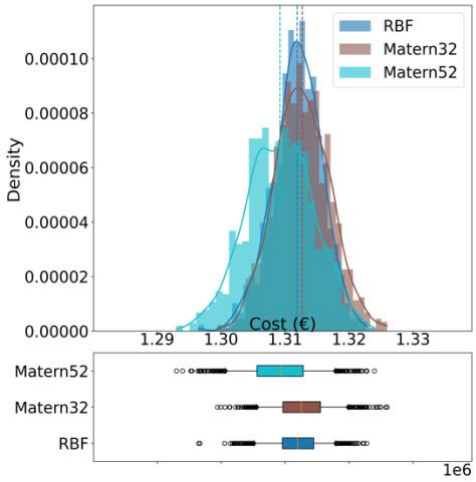

Figure 11. Expected costs of the excavation works for various kernels and 1000 sample each

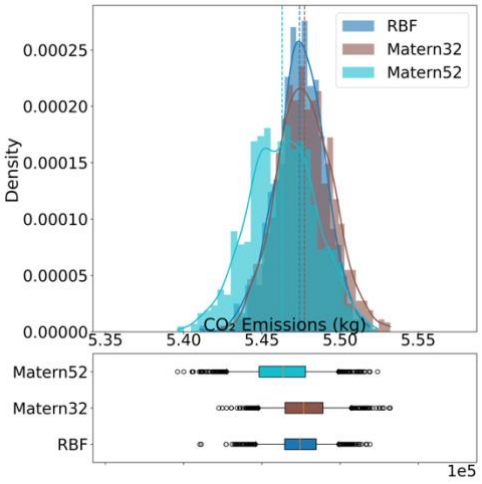

Figure 12. Expected CO2 emissions for various kernels and 1000 samples each

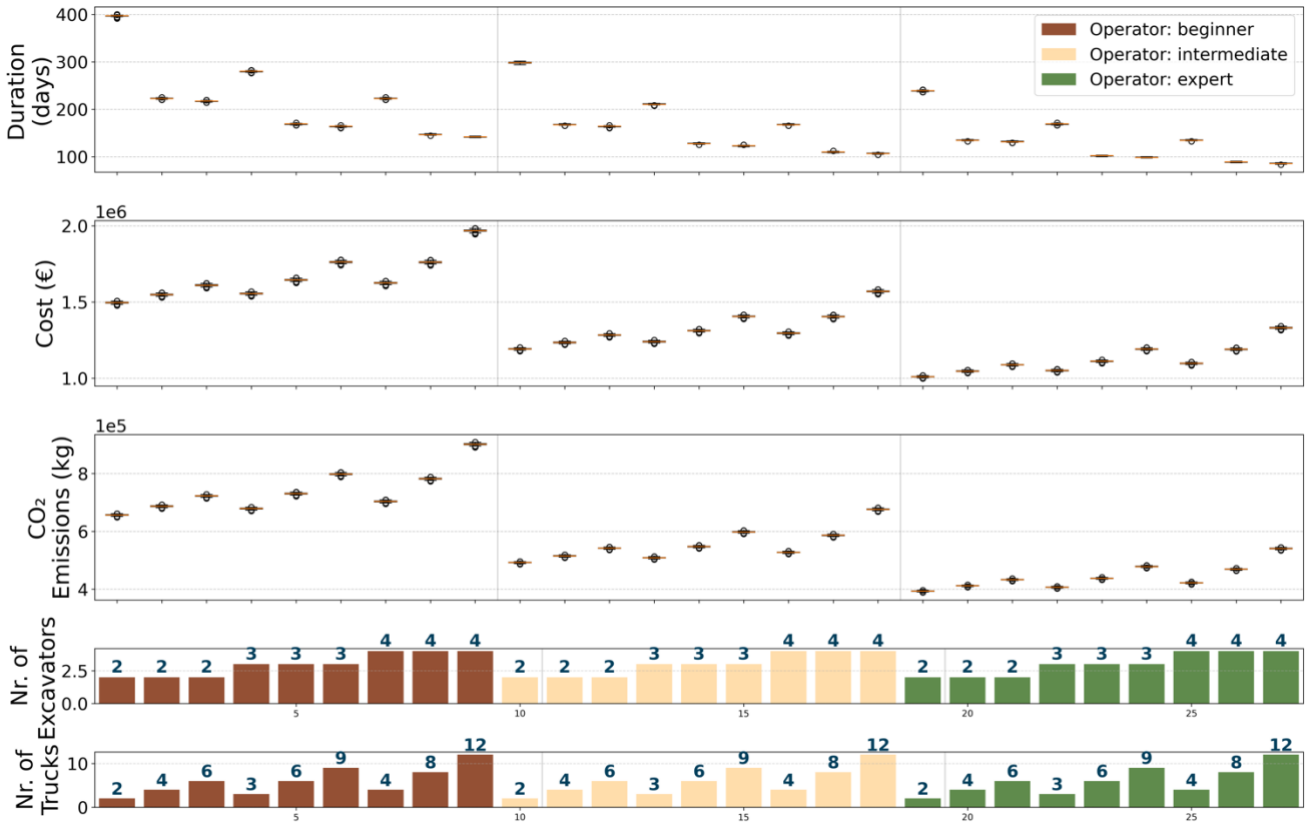

Figure 13. Results for the scenario analysis for different types of operators, number of excavators and number of trucks.

Comparison of Fleet Scenarios

Figure 13 shows the results of a fleet scenario analysis using samples obtained for the RBF kernel. This analysis examines how variations in the number of excavators and trucks (bottom bar plots) and differences in operator experience levels (color-coded) affect excavation performance metrics. The analysis investigates the effect of varying the number of trucks assigned per excavator. We consider three configurations: a) one truck per excavator, b) two trucks per excavator and c) three trucks per excavator.

Increasing the number of excavators from two to three leads to a notable reduction in excavation time. However, this effect becomes less pronounced when the number increases from three to four.

Similarly, assigning two trucks per excavator significantly improves excavation times, but the addition of a third truck offers minimal further benefit. While excavation time decreases slightly, expected costs also continue to increase. This indicates that excavator productivity, rather than truck availability, becomes the limiting factor for this case.

Additionally, the value of operator experience is positive, as the productivity gains for higher-skilled operators lead to cost savings that exceed their higher hourly wages. However, the above conclusions are limited to the current configuration, as a lower productivity gain and higher wage could lead to different results. Despite these limitations, the results demonstrate the potential of such methods to enhance fleet planning and decision-making processes.

7. CONCLUSION & OUTLOOK

In this work, we introduced a novel approach for integrating 3D probabilistic geological models into the planning of excavation works. To this end, we developed a simulation framework to predict excavation times, costs, and emissions under specific site conditions, using samples generated from probabilistic subsoil models.

A preliminary investigation into model selection was performed by varying the configurations for the surface-based probabilistic machine learning method used to generate geological models. This initial analysis did not reveal substantial differences in model volume outcomes for different kernels. Howevern, further research involving additional case studies and more complex kernel functions is needed to draw definitive conclusions.

Future work should also explore alternative geological modeling methods to improve the understanding of the relationship between

systematic uncertainties of such methods and expected excavation durations. Enhancing this understanding is essential for building trust in the reliability of digital planning tools.

The developed simulation model serves as a first step toward improving planning tools for excavation management. While more complex simulation models could increase the predictive accuracy, they would also involve larger computational costs for optimization. Validation with real-data is required to identify an appropriate balance between both aspects.

Overall, this study demonstrates the potential of integrating probabilistic machine learning models with a simulation framework to enhance the management of excavation projects. By enabling more accurate volume estimation and supporting data-driven decision-making, this integrated approach aims to enable more efficient, cost-effective, and environmentally conscious construction practices.

ACKNOWLEDGEMENT

This research is supported by the TUM Georg Nemetschek Institute Artificial Intelligence for the Built World.

REFERENCES

Ching, J., Huang, W.-H., Phoon, K.-K., 2020. 3D Probabilistic Site Characterization by Sparse Bayesian Learning. J. Eng. Mech. 146, 04020134. https://doi.org/10.1061/(ASCE)EM.1943-7889.0001859

Ching, J., Yoshida, I., Phoon, K.-K., 2023. Comparison of trend models for geotechnical spatial variability: Sparse Bayesian Learning vs. Gaussian Process Regression. Gondwana Res., Data driven models 123, 174–183. https://doi.org/10.1016/j.gr.2022.07.011

Cotoarbă, D., Straub, D., Smith, I.F., 2024. Probabilistic digital twins for geotechnical design and construction. https://doi.org/10.48550/arXiv.2412.09432

GPy, 2012. GPy: A Gaussian process framework in python.

Heravi, G., Taherkhani, A.H., Sobhkhiz, S., Mashhadi, A.H., Zahiri-Hashemi, R., 2021. Integrating risk management's best practices to estimate deep excavation projects' time and cost. Built Environ. Proj. Asset Manag. 12, 180–204. https://doi.org/10.1108/BEPAM-11-2020-0180

Kenley, R., Harfield, T., 2011. Greening procurement of infrastructure construction: optimising mass-haul operations to reduce greenhouse gas emissions. https://doi.org/10.25916/sut.26225807.v1

Liu, H., Li, W., Gu, S., Cheng, L., Wang, Y., Xu, J., 2023. Three-dimensional modeling of fault geological structure using generalized triangular prism element reconstruction. Bull. Eng. Geol. Environ. 82, 118. https://doi.org/10.1007/s10064-023-03166-8

Liu, Z., Zhang, Z., Zhou, C., Ming, W., Du, Z., 2021. An Adaptive Inverse-Distance Weighting Interpolation Method Considering Spatial Differentiation in 3D Geological Modeling. Geosciences 11, 51. https://doi.org/10.3390/geosciences11020051

Lyu, M., Ren, B., Wang, X., Wang, J., Yu, J., Han, S., 2023. Neural spline flow multi-constraint NURBS method for three-dimensional automatic geological modeling with multiple constraints. Comput. Geosci. 27, 407–424. https://doi.org/10.1007/s10596-023-10202-9

Lyu, M., Ren, B., Wu, B., Tong, D., Ge, S., Han, S., 2021. A parametric 3D geological modeling method considering stratigraphic interface topology optimization and coding expert knowledge. Eng. Geol. 293, 106300. https://doi.org/10.1016/j.enggeo.2021.106300

Magnusson, S., Lundberg, K., Svedberg, B., Knutsson, S., 2015. Sustainable management of excavated soil and rock in urban areas – A literature review. J. Clean. Prod. 93, 18–25. https://doi.org/10.1016/j.jclepro.2015.01.010

Murphy, K.P., 2022. Probabilistic Machine Learning: An introduction. MIT Press.

Phoon, K.-K., Ching, J., Wang, Y., 2019. Managing Risk in Geotechnical Engineering – From Data to Digitalization, in: Proceedings of the 7th International Symposium on Geotechnical Safety and Risk (ISGSR 2019). Presented at the Proceedings of the 7th International Symposium on Geotechnical Safety and Risk (ISGSR 2019), Research Publishing Services, pp. 13–34. https://doi.org/10.3850/978-981-11-2725-0-SL-cd

Qian, Z., Shi, C., 2024. Prior geological knowledge enhanced Markov random field for development of geological cross-sections from sparse data. Comput. Geotech. 173, 106587. https://doi.org/10.1016/j.compgeo.2024.106587

Rasmussen, C.E., Williams, C.K.I., 2005. Gaussian Processes for Machine Learning. The MIT Press. https://doi.org/10.7551/mitpress/3206.001.0001

Shuku, T., and Phoon, K.-K., 2023. Data-driven subsurface modelling using a Markov random field model. Georisk Assess. Manag. Risk Eng. Syst. Geohazards 17, 41–63. https://doi.org/10.1080/17499518.2023.2181973

Zhou, Y., Li, S., Zhou, C., Luo, H., 2019. Intelligent Approach Based on Random Forest for Safety Risk Prediction of Deep Foundation Pit in Subway Stations. J. Comput. Civ. Eng. 33, 05018004. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000796

Zinas, O., Papaioannou, I., Schneider, R., Cuéllar, P., 2025. Multivariate Gaussian Process Regression for 3D site characterization from CPT and categorical borehole data. Eng. Geol. 352, 108052. https://doi.org/10.1016/j.enggeo.2025.108052