공학석사 학위논문

Application of the 3-D Integrated Analysis Method on Subsurface Stratification in Guri Site.

구리시 현장의 지반층상정보에 대한 3차원 통합분석 기법의 적용

2022년 8월

서울대학교 대학원

건설환경공학부

한 정 우

Application of the 3-D Integrated Analysis Method on Subsurface Stratification in Guri Site.

지도 교수 정 충 기

이 논문을 공학석사 학위논문으로 제출함

2022년 6월

서울대학교 대학원

건설환경공학부

한 정 우

한정우 석사 학위논문을 인준함

2022년 8월

위 원 장 박 준 범 (인)

부위원장 정 충 기 (인)

위 원 김 성 렬 (인)

Abstract

Application of the 3-D Integrated Analysis Method on Subsurface Stratification in Guri Site

Han, Joung Woo
Department of Civil and Environmental Engineering
The Graduate School
Seoul National University

In geotechnical engineering, spatial uncertainty has always been one of the primary concerns. In order to minimize unforeseen risks, extensive geotechnical surveys are necessary prior to the construction of important structures. However, the number of surveys is often limited due to spatial, temporal, and financial issues. Therefore, geostatistical interpolation methods have been widely adopted in geotechnical engineering to overcome spatial variability of soil and deficiency of subsurface information. One of the most common applications of geostatistical approaches is estimating subsurface stratification conditions. Subsurface stratification is one of the most basic geotechnical information; however, its significance cannot be overlooked as it provides valuable insights to engineers. Thus, it is crucial to perform reliable spatial analysis to estimate the subsurface stratification condition accurately. In previous studies, integrated analysis has been developed to provide site-specifically optimized estimation of subsurface stratification.

The primary focus of this study is to propose reliable geostatistical interpolation methods and site-specifically optimized subsurface stratification models. In this study, the integrated analysis methods were adopted along with the conventional borehole-based interpolation method to present a reliable subsurface stratification model of a subway line construction site that has recently experienced a ground subsidence incident during excavation. Unlike the conventional method, the integrated analysis utilizes an integrated set of borehole and geophysical survey data. In this study, kriging and simulation based integrated analysis methods were applied.

The geotechnical survey data were digitized and standardized before the spatial analysis. The three geostatistical subsurface stratification estimation methods were adopted, and their estimation performances were evaluated through the leave-one-out cross-validation method. Moreover, the effect of outlining geotechnical survey data was also assessed. Finally, site-specifically optimized soil profile models were proposed for all three estimation methods.

Keywords: Site-specific Optimization, Subsurface Stratification, Geostatistics, Integrated Analysis, Outlier Analysis, Integration of Geotechnical Survey Data

Student Number: 2020-28616

Contents

List of Tables

List of Figures

Chapter 1 Introduction

1.1 Background

"Historically, geostatistics are as old as mining itself. As soon as mining men concerned themselves with foreseeing results of future works and, in particular as soon as they started to pick and to analyze samples, and compute mean grade values, weighted by corresponding thickness and influence-zones, one may consider that geostatistics were born" (Matheron, 1963). Despite its origin, geostatistics has been extensively used not only in mining but in many different fields, including geology, hydrology, petroleum engineering, and geotechnical engineering, due to its strong advantage in handling spatially temporally correlated data.

In general, one of the popular usages of geostatistical spatial interpolation methods is an estimation of ground surface level. However, in geotechnical engineering, the soil profile below the ground level is a more serious concern. Due to the spatial variability of soil, it is often hard to grasp the subsurface conditions only with a limited amount of geotechnical survey data. On some occasions, such uncertainty of the ground can lead to catastrophic events. Therefore, it is crucial to provide a reliable estimation of geotechnical information such as soil profile and material properties.

The flow of the study is shown in Figure 1.1. First of all, borehole survey data and geophysical survey data were standardized. Then, the datasets were outlined to observe the effects of outlier analysis. The original and

outlined datasets are used for estimating the subsurface stratification of the target area with three geostatistical interpolation methods. The first is a conventional spatial interpolation method using ordinary kriging, which only uses borehole survey data. The other two methods are branches of integrated analysis; one is a kriging-based method, and the other is a simulation-based method. As it can be inferred from its name, integrated analysis requires the integration of borehole survey data and geophysical survey data based on the threshold value before estimation. Through a cross-validation based optimization process, integrated analysis proposes site-specific layer boundary classification criteria. After determining the site-specific criteria, the optimal estimation result can be obtained using the criteria to select the dataset.

This study evaluates the estimation performance of three geostatistics-based layer boundary interpolation methods through leave-one-out cross-validation and proposes a site-specifically optimized subsurface stratification model. For the analysis, a 26m × 81m sized area of the subway line construction site near the ground subsidence zone was selected as a target area. Vertical and horizontal boring data and seismic wave tomography data collected from the target area were used for the analysis.

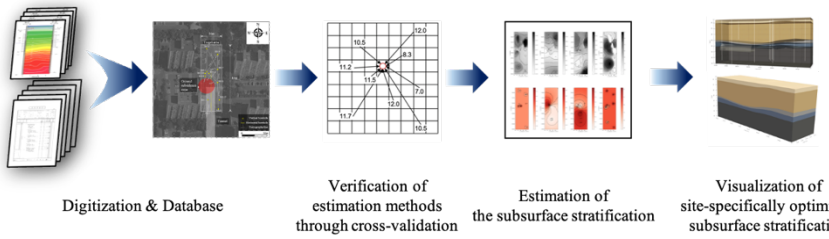


Figure 1.1 Research flow diagram

1.2 Objectives

The aim of this research is to propose reliable site-specifically optimized subsurface stratification estimation methods and provide optimal soil profile models of the subway construction site that experienced ground subsidence. This dissertation deals with the effects of data outlining and the spatial interpolation method on the estimation of subsurface stratification.

1.3 Dissertation Organization

This dissertation documents a proposal of the site-specifically optimized subsurface stratification estimation method for a subway line construction site that experienced ground subsidence. In CHAPTER 1, the introduction presents the research background, research objectives, and dissertation organization. In CHAPTER 2, literature reviews for integration of geotechnical survey data, database management system, and general explanations about geostatistical interpolation methods are presented. In CHAPTER 3, digitization, standardization, integration, and outlining process of both borehole survey data and geophysical survey data are explained. CHAPTER 4 describes the site and datasets, explains the site-specifically optimized spatial interpolation method, evaluates their performances, and proposes a site-specifically optimized soil profile model. CHAPTER 5 summarizes the result of this study.

Chapter 2 Literature Review

2.1 Integration of Geotechnical Survey Results

Interpolation of a soil profile

Many studies have tested and assessed mathematical, empirical, and geostatistical interpolation methods to obtain reliable maps of geographical information, such as ground surface. One of the most common applications is interpolating ground surface elevation to develop Digital Elevation Model (DEM) derived from aerial images (Arun, 2013; Ajvazi, 2019), and some studies expanded their scopes to below the ground surface to estimate subsurface stratification conditions with the aid of boring data (Chun, 2005; Kim, 2016).

Borehole survey data

A borehole survey, one of the most fundamental forms of geotechnical investigation, allows operators to check ground conditions visually and provides reliable stratification information. Thus, if possible, stratigraphy information obtained from boreholes, such as ground elevation, layer thickness, depth to layer-boundary, and depth to bedrock, are considered hard data. Although boring data gives reliable information about t

he target grounds, the number of borings made in practice at constructi on sites is often limited due to spatial, temporal, and financial problems at the cost of safety. To minimize risks, "*the number of investigation points should be extended if it is deemed necessary to obtain an accur ate insight into the complexity and the variability of the ground at the site.*" (Eurocode 7, 2007) Nevertheless, the borehole survey has an inhe rent limitation that cannot be overcome because it is conducted in a 1-dimensional form. Therefore, it is hard to grasp the overall geotechnical properties of the target ground, which varies over 3-dimensional space.

Integration of borehole survey data and geophysical survey data

Compared to borehole survey data, geophysical survey data has the advantage that it provides a continuous 2-dimensional profile of the engineering properties of the soil. However, geophysical survey data is considered soft data, which cannot be a solid representation of the actual ground condition. Some studies have integrated soft data with hard data to analyze and reproduce original ground conditions (Kim, 2012a; Kim, 2015; Oh, Chung, and Lee, 2004; Gallerini and Donatis, 2009). Notably, Kim (2014) proposed the integration of seismic refraction wave velocity data with stratification profile data from the borehole survey. Also, Kim (2015) suggested integrating electric resistivity data with the stratification profile.

2.2 Database Management System

Database and Database Management System

A Database is a collection of data. A Database Management System (DBMS) is a software manager that controls the database structure and access to the stored data. In general, a DBMS acts as a mediator between the user and the storage and includes the following features (Rigaux, 2001).

- Defining of database (data type, structures, and constraints)
- Constructing of the database (data storage)
- Manipulating the database
- Querying of the database (data retrieval)
- Updating the database (data value change)

The general form of a database system and the role of a DBMS can be illustrated as shown in Figure 2.1.

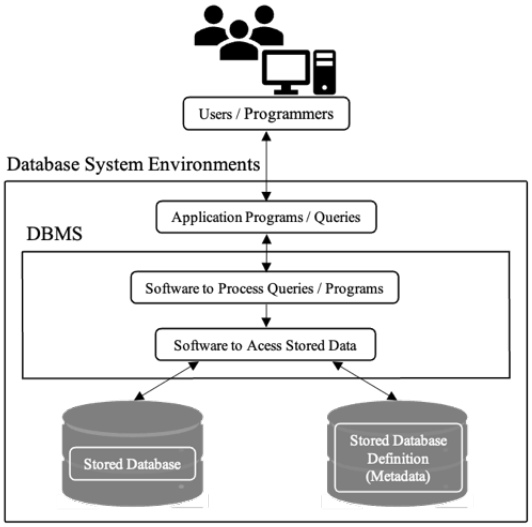


Figure 2.1 A simplified database system environment (Rigaux, 2001)

Geographic Information System (GIS)

A Geographic Information System (GIS) is a key to the spatial analysis of geotechnical data. GIS can be defined differently based on the perspectives.

- Toolbox-based definition: "*a powerful set of tools for collecting, storing, retrieving at will, transforming and displaying spatial data from the real world"* (Burrough, 1987)
- Database-based definition: "*a database system in which most of the data are spatially indexed, and upon which a set of procedures operated to answer queries about spatial entitles in the database"* (Simth et al., 1987)
- Organization-based definition: "*a decision support system involving the integration of spatially referenced data in a problem-solving environment"* (Cowen, 1988)

From the vantage point of toolbox-based definition, GIS is a system that allows us to productively utilize both spatial and non-spatial data (Parker, 1988). The database-based viewpoint focuses more on its ability to handle spatial data. Lastly, the institution-oriented perspective emphasizes the interaction between the users and GIS. Due to its versatility, GIS showed its potential in cartography, geology, civil engineering, urban planning, and even image analysis.

GIS is often used to store geotechnical and geological survey data in geotechnical engineering. As geotechnical engineers' interests in GIS increased, many researchers have developed Web-based GIS. Web-based GIS has several advantages in accessibility, reliability, and standardization of data (Kunapo et al., 2005; Nemoto et al., 2015; Chang and Park, 2004). However, Web-based

GIS often emphasizes the dissemination of data rather than providing analytical tools. On the other hand, many researchers have also focused on developing geotechnical analytical tools to assess hazards, such as earthquakes and debris flow (Sun and Kim, 2017; Delmonaco et al., 2003)

2.3 Geostatistics

Kriging, one of the most representative methods in geostatistics, was first developed to estimate the average grade of mining panels rather than the property values of unknown points (Krige, 1951; Chiles and Delfiner, 2007). Despite its origin and initial purpose, applications of geostatistical concepts are no more limited to the field of mining. Geostatistical methods are capable of taking spatio-temporal variability into account. Due to its advantage over traditional interpolation methods, geostatistics has been used in petroleum geology, hydrology, and meteorology (Pyrcz and Deutsch, 2014; Kitanidis, 1997; Phillips et al., 1992). Last but not least, geotechnical engineering is one of the beneficiaries. Similar to other listed fields of studies in geotechnical engineering, it is common practice to assume engineering properties of unsurveyed locations based on the limited amount of information collected from the areas. Therefore, many researchers have tried to adopt geostatistical estimation methods to approximate geotechnical properties attained from boring and field test results, including SPT, RQD, RMR, subsurface stratification information, and susceptibility to geotechnical hazards, such as liquefaction (Sitharam and Samui, 2007; Madani and Asghari, 2012; Pinheiro et al., 2016; Chun, 2005; Dawson and Baise, 2005). In this study, geostatistical estimation methods were adopted to estimate the original subsurface stratification condition of an urban area that experienced ground subsidence due to tunneling, and their performances were assessed through cross-validation.

2.3.1 Geospatial Interpolation Method

1) Inverse Distance Weighted (IDW)

The Inverse Distance Weighted (IDW) method is one of the traditional interpolation methods. IDW is a deterministic method developed with the idea that *"an improvement on naively giving equal weight to all samples is to give more weight to the closest samples and less to those that are farthest away. One obvious way to do this is to make the weight for each sample inversely proportional to its distance from the point being estimated"* (Isaak and Srivastava, 1989). The general formula of IDW is as follows (Burrough and McDonell, 1998).

$$Z(x_0) = \sum_{i=1}^{n} \lambda_i \cdot Z(x_i) \,, \sum_{i=1}^{n} \lambda_i = 1 \tag{2.1}$$

$$\lambda_i = d_{ij}^{-r} \Big/ \sum_{i=1}^{n} d_{ij}^{-r} \tag{2.2}$$

where $Z(x_0)$ is the estimated value of variable z in the unknown point, $Z(x_i)$ is the sample value of known data point, λ_i is the weight, n is the number of known points used in the estimation, d_{ij} is the distance between the known point and the point to be estimated, and r is the power parameter.

Due to the simplicity and intuitiveness of the estimating algorithm compared to statistical estimation methods, which will be introduced shortly, IDW has been frequently applied, and its performance has been compared with

that of statistical methods. A brief summary of some early works that focused on comparative analysis of IDW and kriging can be found in the works of Zimmerman et al. (1999) and Shahbeik et al. (2013).

2) Kriging

The term kriging, first used as an honorarium of D. G. Krige by Matheron, is used to call the optimal spatial linear estimation method (Matheron, 1963; Cressie, 1990). In addition, Olea (1999) describes kriging as "*a form of generalized linear regression for the formulation of an optimal estimator in a minimum mean square error sense*". Simple Kriging (SK) is the basic form of kriging with the simplest mathematical model. The general formula for simple kriging is written as follows:

$$Z_0^* = m_0 \sum_{i=1}^{n} \lambda_i \cdot [Z_i - m_i]$$

where Z_0^* is the resulting unbiased estimator, Z_i is a known random variable, m is the mean, n is the number of known points, and λ_i is the weights (Journel, 1989).

The minimized error variance for simple kriging can be written:

$$\sigma_{SK}^2 = Var\{Z_0 - Z_0^*\} = C_{00} - \sum_{i=1}^{n} \lambda_i C_{i0} \geq 0$$

Due to its simplicity of mathematical formulation, simple kriging has a well-known weakness, *"limited applicability and suboptimal results if an attempt is made to force usage beyond the embedded assumption"* (Olea, 1999). One of its critical assumptions is that the mean must be known prior to estimation to minimize the variance of estimation error. However, Ordinary Kriging (OK) does not necessitate this step (Journel, 1989).

Ordinary kriging is also known as Best Linear Unbiased Estimator (BLUE). Unlike simple kriging, ordinary kriging utilizes Lagrange multipliers to solve constrained optimization (Isaak and Srivastava, 1989).

The general formula for ordinary kriging does not include the mean and it is written as follows:

$$Z_0^* = \sum_{i=1}^{n} \lambda_i \cdot Z_i \,, \sum_{i=1}^{n} \lambda_i = 1$$

where Z_0^* is the resulting unbiased estimator, Z_i is a known random variable, n is the number of known points, and λ_i is the weights.

The minimized error variance for ordinary kriging can be written:

$$\sigma_{OK}^2 = E\{[(Z(x_0) - Z^*(x_0)]^2\} = C(0) - \sum_{i=1}^{n} \lambda_i C\big(x_i - x_j\big) - \mu \geq 0$$

Kriging is known to have several inherent advantages over the traditional estimation methods, including IDW, which was explained earlier (Journel, 1986; Setianto (2015) pointed out that kriging is more reliable because it uses spatial autocorrelation examined for a specific point instead of applying a global trend with a universal distance power parameter. In addition,

Zimmerman et al. (1999) demonstrated that the kriging method was superior compared to IDW with different types of surfaces, sampling patterns, noise levels, and spatial correlation of data.

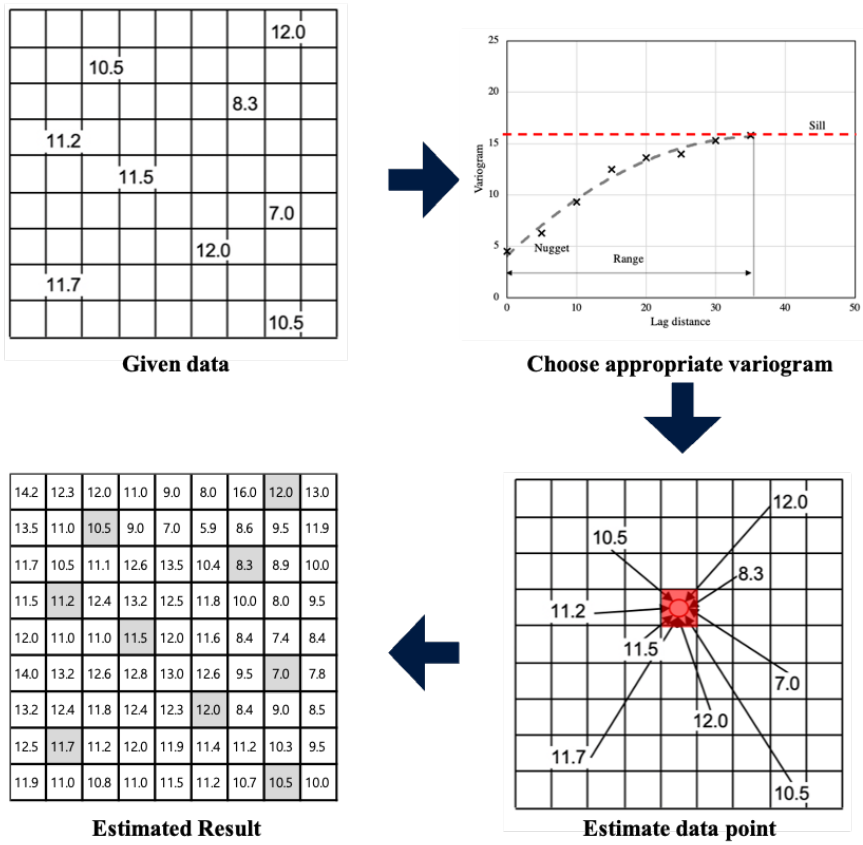


Figure 2.2 General procedure of ordinary kriging

2.3.2 Simulation Method

1) Conditional simulation method

The conditional simulation method is another branch of estimation method that falls into Geostatistics. However, it differs from kriging that "*all moving average-type estimates, including all kriging estimates, provide a smooth image of the underlying reality: the variogram of these estimates would not reproduce the data variogram. The concept of conditional simulation allows generating alternative, equiprobable, images which honor data values at their locations and reflect a series of spatial continuity functions*" (Journel, 1989). The expression "conditional" comes from the last condition stated above, otherwise it is called stochastic simulation.

Sequential Gaussian Simulation (SGS) is one of the representative estimation methods of conditional simulation methods that generates partial realizations with multivariate normal random functions (Olea, 1999; Emery and Pelaez, 2011; Goovaerts, 1997). The main characteristic of SGS is the sequential approach to estimation. Before the estimation process, a sequential, random path is determined. Unlike how kriging provides estimation results in a single step, SGS follows the predetermined path by estimating the unknown point in a sequential order. For each estimation process, both known sample data and previously simulated data are used for the estimation of the following point. Due to its sequential approach, it allows the generation of equally probable realizations by taking different random paths (Verly 1993). For

estimation procedure, simple kriging is sufficient to be used as an implemented estimation method unless it is proved inappropriate (Deutsch and Journel, 1993).

Moreover, SGS is free of smoothing effect, which is a common characteristic of traditional interpolation methods and ordinary kriging (Deutsch and Journel, 1993; Yamamoto, 2005; GhojehBeyglou, 2021). The smoothing effect refers to phenomena where results get flattened with sparse data. However, SGS's shortcoming comes from the inevitable excessive computational cost due to its sequential algorithm. The general procedure of SGS is as follows:

1. Normal score transform the sample data.
2. Compute and fit theoretical variogram.
3. Define a random path for a realization that goes through all the nodes.
4. Do kriging for each node following the designated path.
5. Back transform the estimated result.
6. Repeat steps 3-5 for the desired number of simulations.

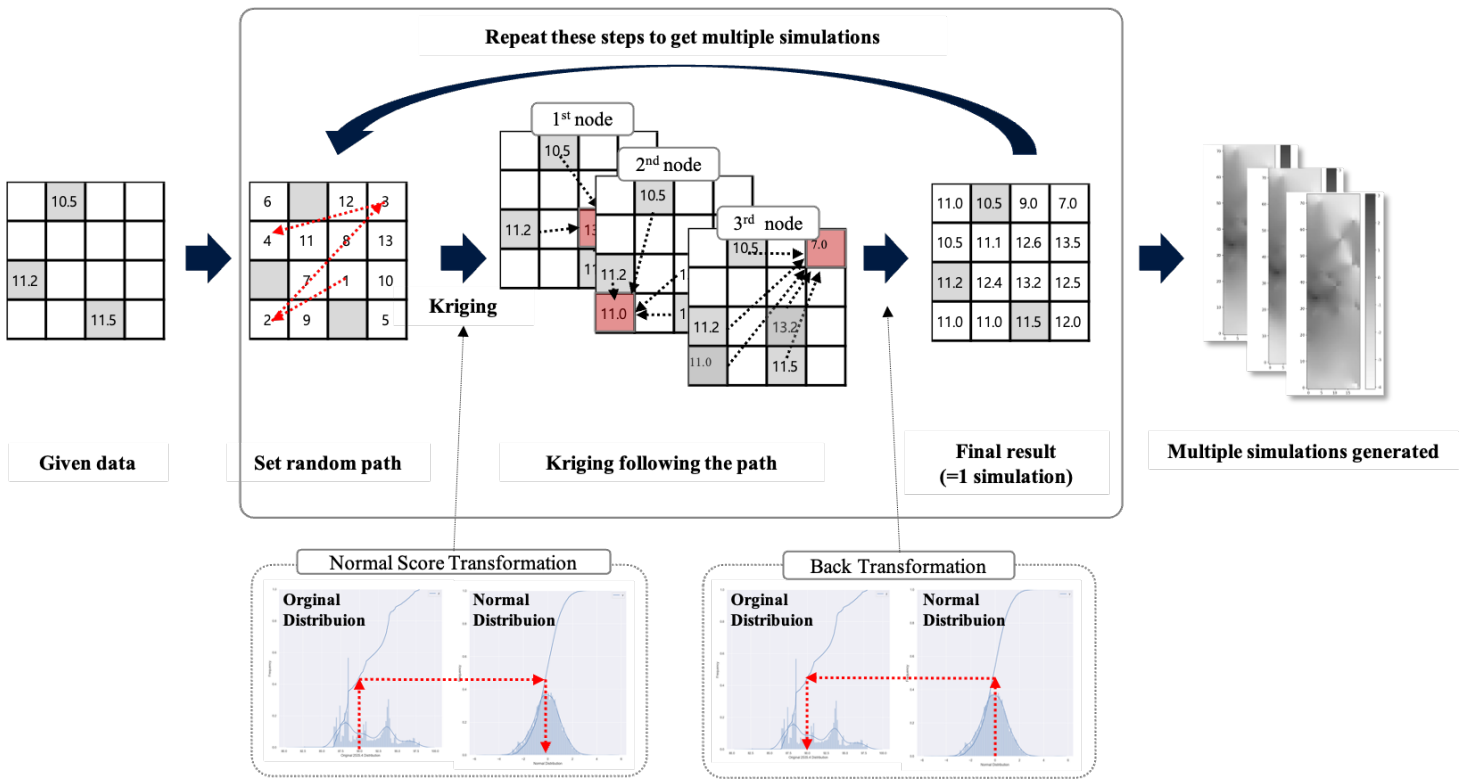


Figure 2.3 General procedure of sequential gaussian simulation

Chapter 3 Preprocessing of 3D Geospatial Integrated Datasets

3.1 Database Development Procedure

The database system allows more flexible uses of data for users. It is a convenient tool for storing, retrieving, analyzing, and updating a massive amount of data. Due to its versatility, the database system has become a standard for handling data in many fields of applications. In geotechnical engineering, it is particularly effective as there are numerous types of data collected from field surveys, field monitoring, field tests, laboratory tests, and so on. To manage and utilize the data, database systems are often used along with GIS, which links the data stored in the database to a spatial and temporal workspace. In general, open source-type GISs offer convenient tools, such as visualization, spatial interpolation, and mapping. However, such programs often come short for applications of modified or advanced analysis methods.

In this study, PostgreSQL, a Database Management System (DBMS), was used to develop a database system for integrated analysis of geotechnical survey data. As illustrated in Figure 3.1, the database was developed to store and effectively retrieve two different types of geotechnical survey data: borehole survey data and geophysical survey data.

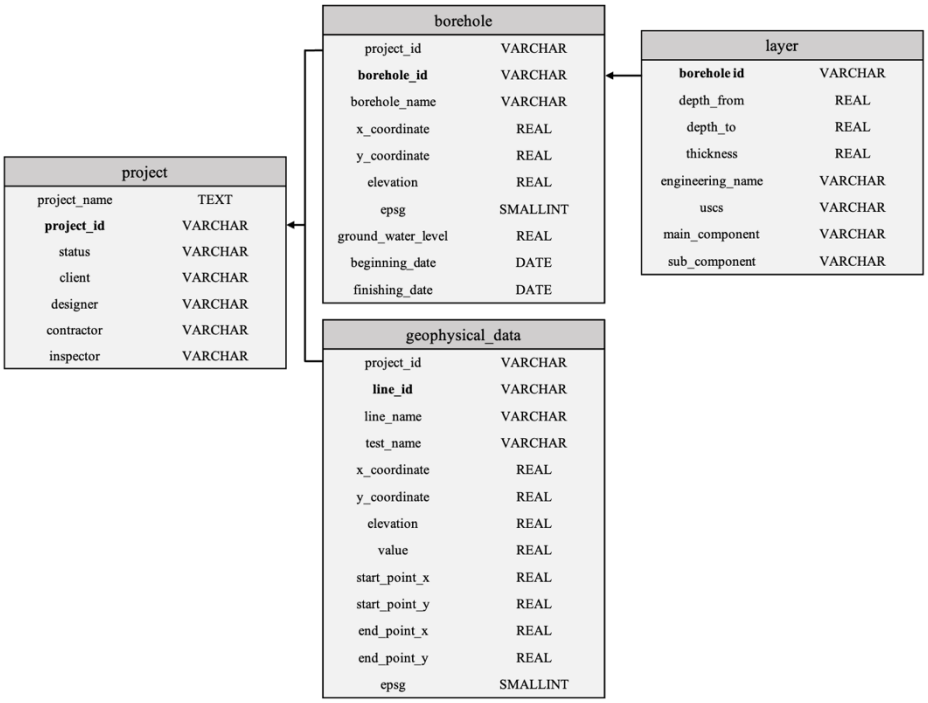


Figure 3.1 Database schema developed for geotechnical integrated analysis

The database was designed to effectively access both borehole and geophysical survey data with the keys: project_id, borehole_id, and line_id**.** Although the geophysical survey data and borehole survey data are closely related considering the locations and survey methods, these two data sets have to be managed separately due to their form of data. In the perspective of 2-dimensional space defined by abscissa and ordinate, a borehole survey data presents as a point data while a single section of geophysical data exists as a line. Therefore, each data requires a different location identification formats and modes of digitization.

3.2 Borehole Datasets Using Outlier Detection Methods

In statistical analysis, the distribution of values of sample data is critical. It is ideal to utilize all of the measured data; however, there is a possibility that the sample data contains some suspicious observations. Such data must be investigated thoroughly to determine whether it is appropriate for further analysis. Beckman and Cook (1983) defined a group of "a contaminant" or "a discordant observation" as outliers. A *contaminant* refers to observations that do not correspond with the target distribution, and a *discordant observation* refers to observations that show a distinguishable discrepancy. It is vital to eliminate outliers as they could lead to misspecification of the model, biased estimation of parameter, and misleading statistical outputs (Kim, 2016)

3.2.1 Outlining Method for Borehole Dataset

In this study, borehole data is considered true data and has a substantial advantage over other survey data. Especially in determining soil profile, it is possible to judge based on the field borehole test results and monitoring retrieved cores by an investigator. However, it does not necessarily mean that borehole data is free of misleading observations. In this sense, performing outlier analysis to detect anomalies is common before further applications. Kim (2012b; 2016) has adopted a cross-validation-based outlier analysis. This outlier detection method consists of three steps. Here, steps 2 and 3 are repeated for all data points to perform leave-one-out cross-validation. Also, Figure 3.2

illustrates the general procedure of the cross-validation method.

1. Compute experimental variogram with whole sample data set and determine appropriate theoretical variogram.
2. Exclude one surveyed data point, then estimate the property value at the removed location with the predetermined theoretical variogram
3. Compute the estimation error by subtracting measured from the original observed value at the point

3.2.2 Outlier Analysis

In this study, a similar outlier detection algorithm was adopted. Outliers were identified based on the average of Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) obtained from cross-validation of each layer boundary. As this study focuses on the optimal estimation of a soil profile that consists of five different layers, cross-validation results must be evaluated on all four types of layer boundaries: F-AS(RS), AS(RS)-WR, WR-SR, and SR-MR[1]. With the conventional method, which compares the measured value and estimated value via cross-validation for each layer boundary distinctively, it is hard to find a solid agreement among several layer boundaries. Therefore, the cross-validation result of each borehole was quantitatively assessed based on MAE and RMSE.

[1] F, AS, RS, WR, SR, MR refers to fill, alluvial soil, residual soil, weathered rock, soft rock, and medium rock, respectively.

The equations for MAE and RMSE are as follows:

$$MAE = \frac{1}{n}\sum_{i}^{n}|y_i - \hat{y}_i|$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

where n is the total number of data, y_i is the measured value, and $\hat{y}_i$ is the estimated value.

The cross-validation result is shown in Table 3.1 in descending order of errors. Values of both MAE and RMSE showed the same trend. The top two boreholes, BH-6 and BH-12, on the list were determined as outliers considering their discrepancy with the rest of the data.

Table 3.1 Cross-validation based outlier detection of borehole data for the elevation of F-AS(RS), AS(RS)-WR, WR-SR, and SR-MR boundary layers.

Borehole id	MAE	RMSE
BH-6	**3.396**	**5.103**
BH-12	**2.403**	**2.827**
BH-11	1.149	1.363
BH-8	1.103	1.340
BH-3	0.812	0.916
BH-10	0.700	0.739
BH-9	0.618	0.668
BH-5	0.504	0.652
BH-7	0.399	0.525
BH-4	0.249	0.278
BH-1	0.112	0.112

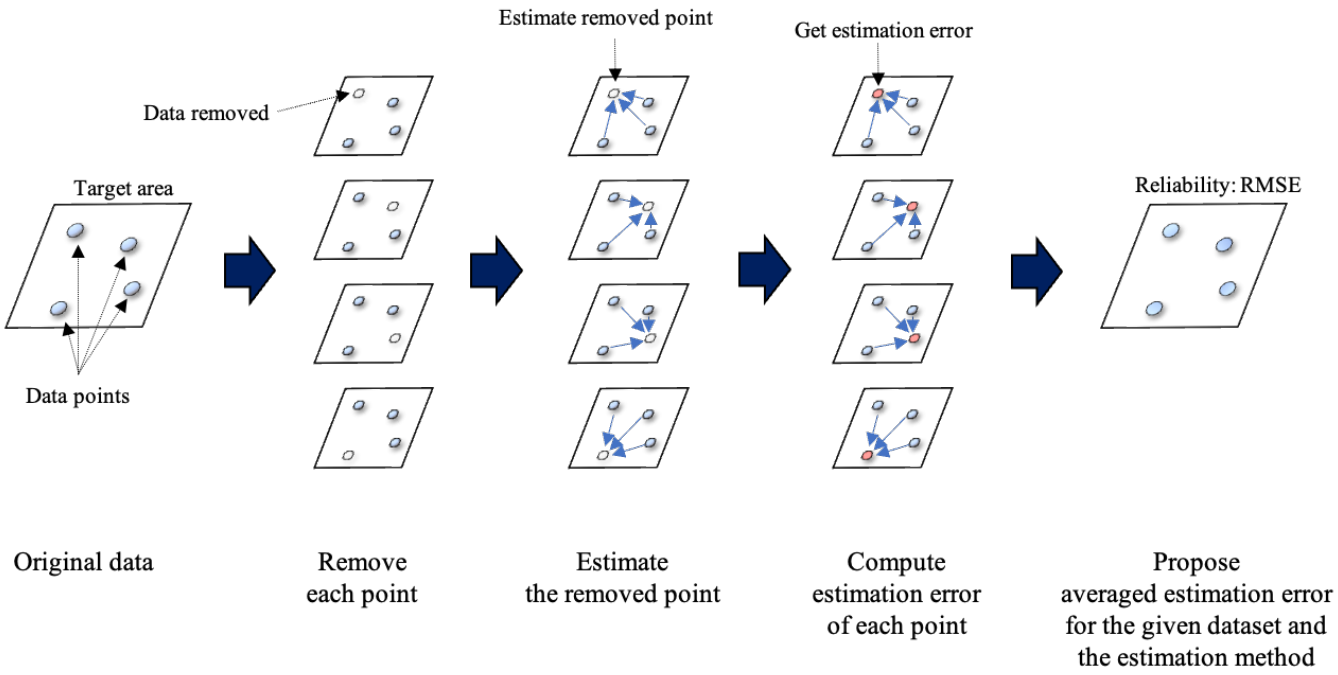


Figure 3.2 General procedure of leave-one-out cross-validation.

3.3 Optimization of The Existing Stratification Classification Criteria Using Cross-validation

Optimizing the subsurface stratification estimation method requires integrating borehole survey data and geophysical survey data. The integrated data set is used for the estimation of specific layer boundaries. To quantitatively evaluate the reliability of the estimated result, the leave-one-out cross-validation method was adopted. During the analysis, different sets of integrated data are created, and each set of data is considered a trial set. Depending on the types of geophysical survey data, appropriate criteria for trials must be determined.

3.3.2 Digitization of Geophysical Survey Results

In this study, cross-hole seismic tomography survey data was collected along with vertical and horizontal boring data. Before integrating the two distinct data sets, the geophysical survey results must be digitized.

Each section of tomography was given a line_id for identification for digitization as described in the database development procedure. Then each P wave velocity line shown in the tomography was digitized based on the scales and locations of the boreholes shown on the edges of the tomography section. In order to convert velocity line data to a set of equivalent point data, points were plotted using the spline method to duplicate the original velocity line. In other words, a group of points, which are defined by x, y, and z coordinates,

represents a velocity line shown in the tomography section, and the comprehensive set of all velocity datasets form a tomography section. After digitization, all of the point data were stored in the database described in the previous section. Figure 3.3 illustrates the digitizing, reconstructing, and extracting process of a tomography section.

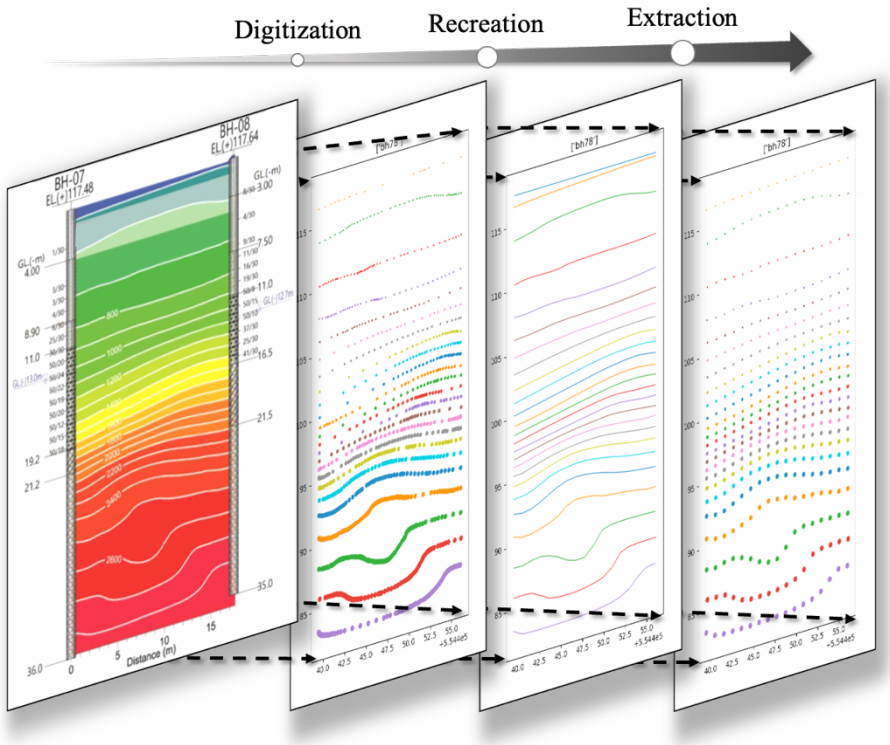


Figure 3.3 Processing of geophysical survey data.

For the integrated analysis, the geophysical data were selected based on the line_id and value, which is the velocity of the P-wave. After retrieving a specific dataset, the points were replotted into a line with the aid of SciPy, an open-source Python library that allows the recreation of smooth lines out of

given data points and extraction of the equally distanced data points from the line. Such programming language-based data manipulation enables users to transform unstructured digitized data into structured data defined by desired data spacing. It has a clear advantage over structured digitization in that data can be restructured effortlessly and repetitively avoiding any misinterpretation of the original data. As a preliminary research, different data spacings (5m, 4m, 3m, 2m, 1m, 0.5m, and 0.1m) were adopted; however, spacings did not show a clear relationship with estimation performance. In this study, data spacing of 1m was adopted considering the computational cost and data density.

3.3.2 Existing Subsurface Stratification Classification Criteria

In the previous studies that adopted integrated analysis for optimal subsurface stratification estimation, geophysical data used for the layer boundary estimation were determined by referring to the conventional geomaterial classification criteria (Kim, 2014; Kim, 2016). The geomaterial classification criteria proposed by Korean government departments and public enterprises are based on the P-wave velocity, V_p.

This study aims to determine optimal V_p for estimating subsurface stratification which comprises of F, AS(RS) [2], WR, SR, and MR. Correspondingly, total four classification criteria are needed. For the

[2] Based on the boring log, residual soil (RS) was observed in only three boreholes with average thickness of 1.5m. Moreover, the SPT-N values of RS were within the range of that of AS. Thus, in this study, RS is considered to be included in AS.

V_p classifying AS(RS)-WR and WR-SR boundaries, the two V_p -based classification guidelines were adopted, which are shown in Table 3.2. However, for the V_p classifying F-AS(RS) and SR-MR, the guidelines suggest an unreasonable range of values for the given site. Due to the discrepancy between the guideline and dataset, the classification criteria for F-AS(RS) and SR-MR were determined based on the boring log and the range of the V_p of seismic tomography. This study used the geomaterial classification criteria of 600 m/s, 1,500 m/s, 2,000 m/s, and 2,200 m/s for F-AS(RS), AS(RS)-WR, WR-SR, and SR-MR, respectively.

Table 3.2 P-wave velocity-based geomaterial classification criteria

Geomaterial classification[3]	**Soil deposit**	**Weathered rock**	**Soft rock**	**Normal rock**	**Hard rock**
$\boldsymbol{V_p}$ **(m/s)**	<1500	1500-2000	2000-3200	3200-4200	4000-5000

After determining the classification criteria for each layer boundary, threshold value for each trial were assumed. Kim (2016) adopted geomaterial classification criteria of 1,500 m/s and 2,000 m/s for RS-WR and WR-SR boundaries and a fixed ratio for trial classification criteria. This study rescaled and differed the ratio for trial classification criteria for each layer boundary to provide optimal estimation for the given site. Tables 3.3 and 3.4 show the assumed classification criteria for the previous study and this study.

[3] Geotechnical investigation handbook (Seoul Metropolis, 2006)
Construction estimate manual (Korea Ministry of Land, Infrastructure and Transport, 2012)

Table 3.3 Seven assumed P-wave velocity classification criteria suggested by Kim (2016).

Assumed criteria V_p (m/s)	No.1	No.2	No.3	No.4	No.5	No.6	No.7
RS-WR	900 (60%, 3/5)	1125 (75%, 3/4)	1312.5 (88%, 7/8)	**1500 (100%)**	1687.5 (112%, 9/8)	1800 (120%, 6/5)	1875 (125%, 5/4)
WR-SR	1200 (60%, 3/5)	1500 (75%, 3/4)	1750 (88%, 7/8)	**2000 (100%)**	2250 (112%, 9/8)	2400 (120%, 6/5)	1500 (125%, 5/4)

Table 3.4 Seven assumed P-wave velocity classification criteria suggested in this study.

Assumed criteria V_p (m/s)	No.1	No.2	No.3	No.4	No.5	No.6	No.7
F-AS(RS)	498 (83%)	525 (87.5%)	570 (95%)	585 (97.5%)	**600 (100%)**	630 (105%)	675 (112.5%)
AS(RS)-WR	750 (50%)	1125 (75%)	1312.5 (87.5%)	**1500 (100%)**	1687.5 (112.5%)	1875 (125%)	2250 (150%)
WR-SR	1000 (50%)	1500 (75%)	1750 (87.5%)	**2000 (100%)**	2250 (112.5%)	2500 (125%)	3000 (150%)
SR-MR	1320 (60%)	1650 (75%)	1925 (87.5%)	**2200 (100%)**	2475 (112.5%)	2750 (125%)	3080 (140%)

3.3.3 Selection of Site-specific Layer Boundary Criteria

After preparing the data and determining analysis conditions, the following step is to perform the estimation. A quantitative measure of the estimation performance can be provided through cross-validation. This study adopted the leave-one-out cross-validation method. Leave-one-out method refers to a methodology that tests estimation error for all given sample data points. Initially, one data point is eliminated, and the property value of that point is predicted with the rest of the data using the specified estimation method. Then, the estimation error for the data point is stored and move onto a different data point that has not been excluded yet. After computing all points' estimation errors, the estimation method's reliability is evaluated based on the RMSE value.

For integrated analysis, cross-validation is performed for each integrated dataset. Here, the integrated dataset refers to a dataset formulated with boring data and geophysical data that corresponds to a single assumed classification criterion. Therefore, cross-validations for all seven assumed criteria are performed for each layer boundary for an estimation method. The optimal site-specific layer boundary classification criterion can be determined by plotting the RMSE-V_p plot, and then, finding the V_p of minimum RMSE. Figure 3.4 shows an example of RMSE - V_p plot used for determining optimal V_p value for a layer boundary.

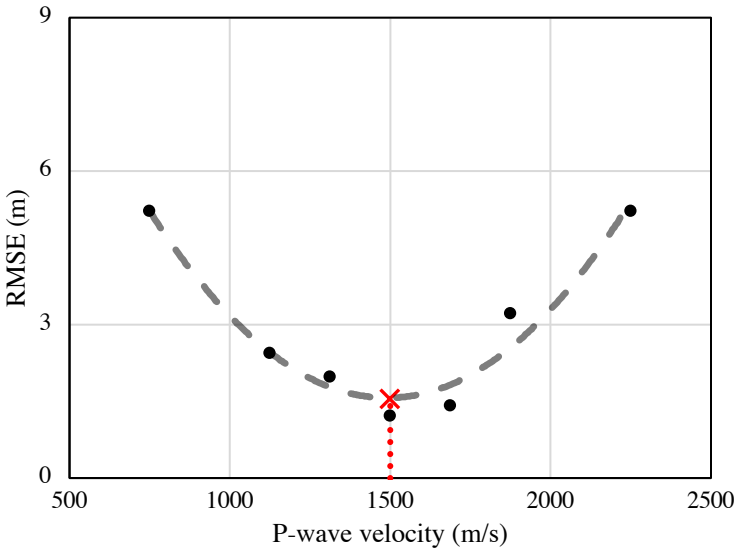


Figure 3.4 An example of RMSE – P-wave velocity plot.

Chapter 4 Optimal Subsurface Stratification Estimation Method

4.1 Site Characterization

An urban tunnel excavation site for subway construction in Gyomun-dong, Guri-si, Gyeonggi province was investigated to determine the cause of ground subsidence. In this study, the subsurface investigation data of the site was used to estimate the original state of subsurface stratification before the incident. The approximate location of the site is shown in Figure 4.1.

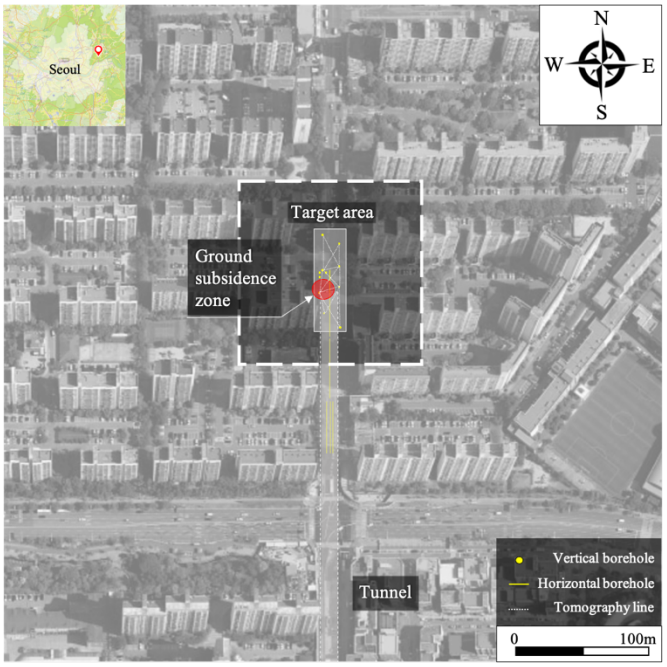


Figure 4.1 Site overview shown over a 500m × 500m sized satellite image.

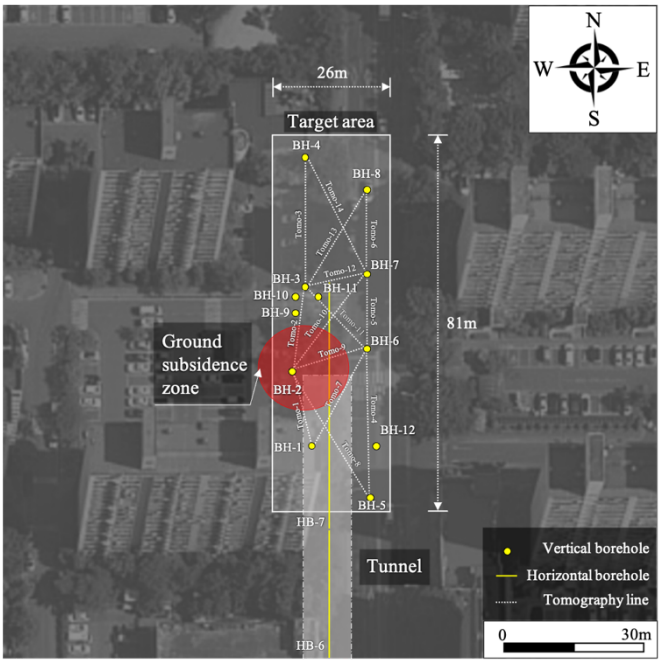


Figure 4.2 Site overview shown over a 150m × 150m sized satellite image.

As shown on the top left of Figure 4.1, the site is placed in Guri-si, which is located next to the northeast perimeter of Seoul, the capital city of Republic of Korea. According to the construction management reports, several boreholes were surveyed along the planned subway line, which spans from south to north. BH-12, shown in Figure 4.2, is one of the boreholes used for preliminary investigation. In addition, horizontal boreholes were also bored during the excavation. Other vertical boreholes numbered 1 to 11 were surveyed after the incident for the investigation. Along with the borehole survey, a cross-hole seismic tomography survey was conducted. In total, 12 vertical boreholes, 1 horizontal borehole, and 14 topographies sections were collected from the target area.

4.1.1 Geological History

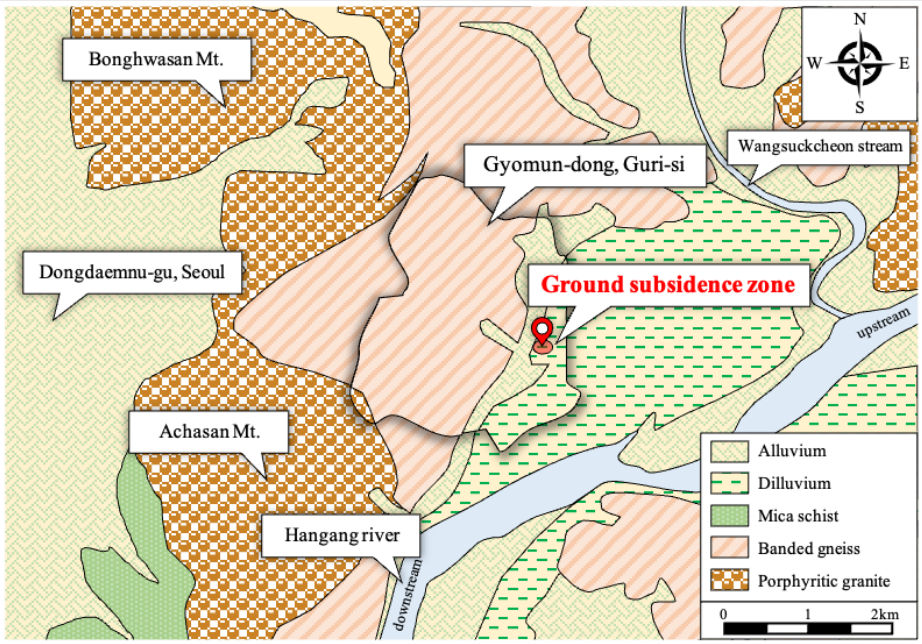


Figure 4.3 Geologic map of the area near Gyomun-dong, Guri-si, Gyeonggi province. (Geological map of Korea, 1981)

A geologic map of the area near Gyomun-dong is shown in Figure 4.3, along with the boundary of Gyomun-dong and the approximate location of the subsided area. As illustrated in the figure, the majority of the Guri-si outcrop consists of the Quaternary alluvium, diluvium, and the Precambrian banded gneiss. The subsidence took place in the area covered by diluvium deposited in the Quaternary Pleistocene epoch. The Pleistocene deposit is a non-lithified layer formed by the accumulation of sediments supplied from floodings of nearby rivers: Wangsuckcheon stream and Hangang river. Generally,

unconsolidated or non-lithified deposits have higher permeability and void ratio. Therefore, it is more likely to have high groundwater flow and susceptible to physical erosion. In addition, the Ministry of Land, Infrastructure, and Transport has published an investigation report of the incident.[4] According to the report, satellite imagery taken in 1947 showed that the Jangjamot pond located about 400 meters from the subsided area was part of an ancient stream channel that linked Wangsuckcheon stream and Hangang river. Moreover, the report states that the bedrock near Jangjamot pond primarily consists of gneiss while that of the investigated area consists of mostly granite and some gneiss. Also, it was presumed that the granite intruded the pre-existing gneiss.

4.1.2 Ground Subsidence

On August 25th, 2020, a 16m wide and 21m deep sinkhole occurred on the road near apartment complexes in Gyomun-dong, Guri-si. The ground subsidence zone was located directly above the tunnel excavation site for the Byeolnae subway line, which is planned to connect three different cities, Seoul, Guri-si, and Namyangju-si. Due to the event, the central subsurface incident investigation committee was formed to investigate the cause of the incident. According to the investigation report, it was concluded that the initiation of the subsidence had not been influenced by the nearby groundwater conditions such as leakage from sewages; however, the unforeseen variability of the subsurface

[4] Ministry of Land, Infrastructure and Transport (2021) Investigation report of Guri-si ground subsidence by the central subsurface incident investigation committee.

stratification condition led to collapse at the face of excavation. Fortunately, no fatalities were reported as the subsidence developed gradually, and there was enough time to evacuate the excavation site both above and below the surface. Figure 4.4 shows the approximate shape and location of the ground subsidence zone in 3-dimensional space with the tunnel and the data points.

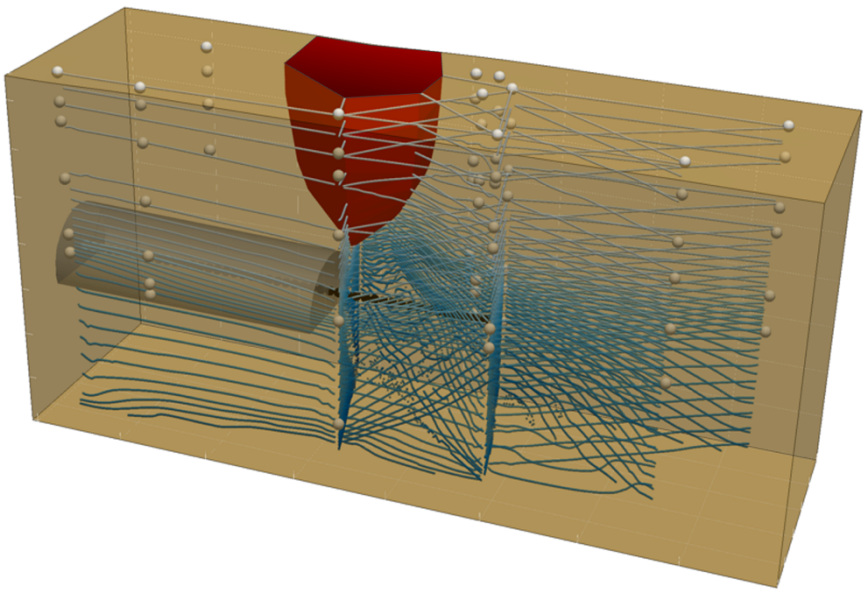


Figure 4.4 Ground subsidence zone visualized with the tunnel and the geotechnical survey data points.

4.2 Geotechnical Information Data

4.2.1 Types of Surveyed Data

Subsurface investigations were conducted during the design of the subway construction and after the investigation. Before the tunnel excavation, a few borehole surveys were performed along the planned subway line. During the excavation, horizontal boreholes were bored at the face of the excavating tunnel. After the incident, several additional geotechnical surveys were conducted, including a borehole survey, Standard Penetration Test (SPT), high-resolution optical televiewer borehole logging, gamma-gamma density logging, and seismic wave tomography survey, Ground Penetrating Radar (GPR), basic geotechnical property tests, and uniaxial compression test.

4.2.2 The Density of Survey Data

The size of the target area and locations of all surveyed points are shown in Figure 4.2. As this study aims to propose a site-specifically optimized subsurface stratification in its original form, it is necessary to use data selectively based on their locations and histories. First, data obtained outside of the target area were eliminated. Then, data points within the ground subsidence zone were removed. Lastly, tomography lines that crossed the excavated tunnel were also removed as it is hard to fathom the influence of the tunneling, grouting, and reinforcement. In total, one borehole (BH-2), three tomography

lines (Tomo-1, -7, and -8), and some portions of the other three tomography lines (Tomo-2, -9, and -10) were removed from the dataset. The entire dataset before outlining is shown in Figure 4.5. As shown in the figure, the geotechnical surveys were done in a relatively smaller area considering the general scale of construction sites.

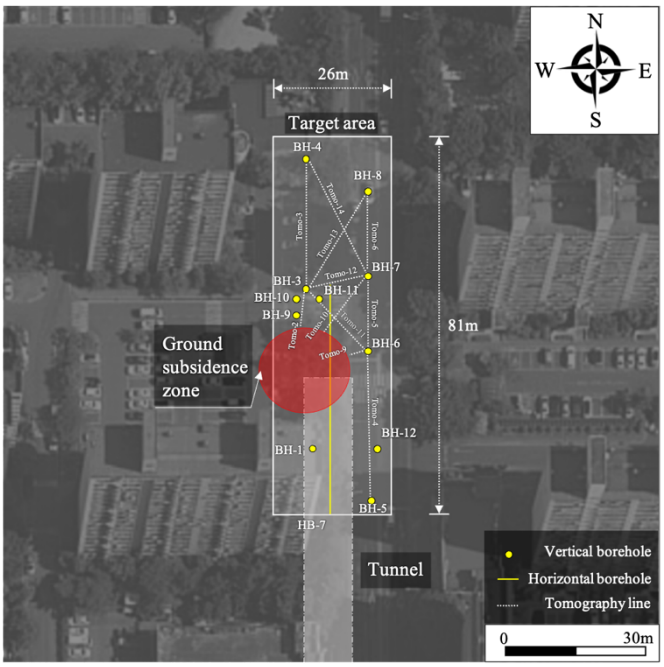


Figure 4.5 Target area and locations of data before outlining.

To minimize the effects of potential errors and provide a more reliable estimation of original subsurface stratification conditions, borehole data outliers were determined using leave-one-out-based cross-validation method. As a result of the outlier analysis, two boreholes (BH-6 and -12) were identified as outliers. Moreover, three tomography lines (Tomo-2, -9, and -10) were also

removed from the dataset, as they have possibly been influenced by grouting at the face of excavation according to the seismic wave tomography report. Figure 4.6 shows an example of a tomography section that exhibits the grouting effect. Compared to the adjacent sections, the influenced section in the middle has a higher velocity near the face of the tunnel. Moreover, such a phenomenon leads to inconsistency of the velocity line near the boreholes as it can be recognized by observing 1500m/s velocity line. There were total three tomography sections showed high velocity in the middle. Thus, the three lines were removed from the dataset to minimize any potential source of errors in analyzing the original subsurface stratification condition prior to the ground subsidence incident.

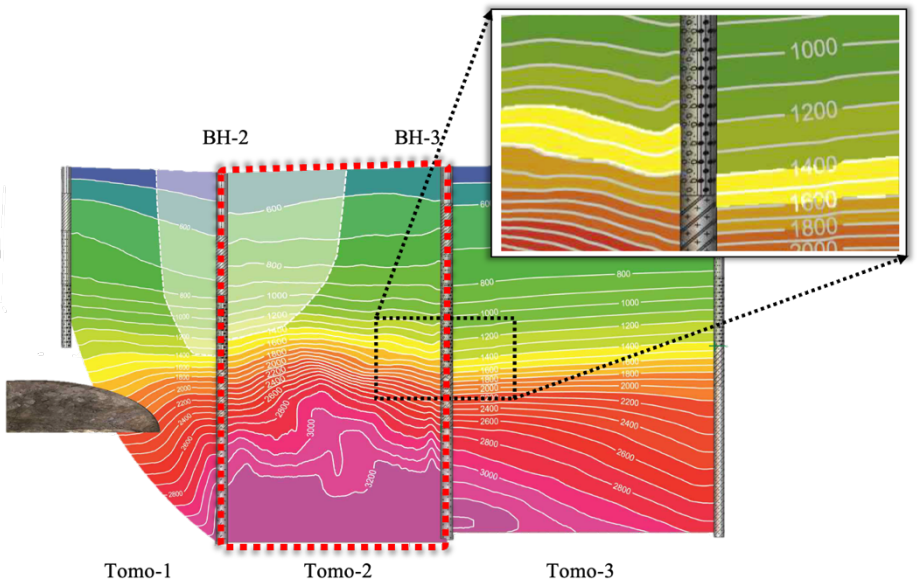


Figure 4.6 An example of a tomography section affected by grouting.

The target area for the outlined dataset is the same as that of the dataset before outlining. The locations of outlined data are shown in Figure 4.6. The amounts of data before and after outlining are summarized in Table 4.1.

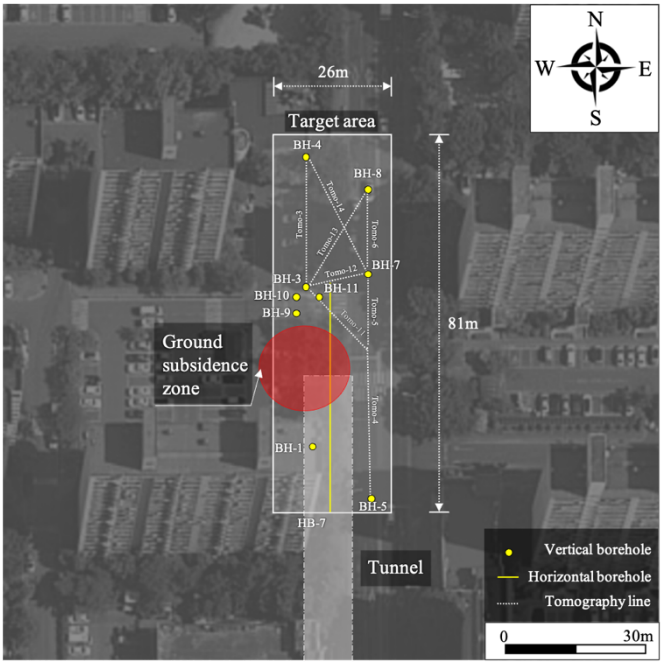


Figure 4.7 Target area and locations of data after outlining.

Table 4.1 Summary of the amount of data before and after outlining.

Type of survey	Type of data	Before outlining	After outlining
Borehole survey	Vertical (hole)	11	9
	Horizontal (hole)	1	1
Geophysical survey	Tomography (section)	11	8

4.3 Subsurface Stratification Estimation with Borehole Dataset

Before the application of the integrated analysis of subsurface stratification, the soil profile of the target area was estimated only with borehole data. Borehole data-based estimation is similar to the conventional estimation process using kriging. First, the target layer boundary is set. Then, the spatial information of the layer boundary is selected from each borehole. The spatial information includes x-, y-, and z-coordinates of the point that indicates the boundary of the layers. Since there were no inversions of the layers based on the boring log, there were no redundant data points. However, some types of layers were not observed in several boring locations. For these points, the missing layer was assumed to have zero thickness. Thus, it was assumed that there are no inversions between the layers in vertical directions. In the target area, five different soil and rock layers were observed. For simplicity, the five soil and rock layers, fill, alluvial soil, residual soil, weathered rock, soft rock, and medium rock, were abbreviated to F, AS, RS, WR, SR, and MR. The average thickness of each layer was 2.49m, 15.19m, 0.56m, 2.59m, and 2.86m for F, AS, RS, WR, and SR, respectively. As it can be noticed from the thickness, there is relatively small amount of residual soil present at the site. Moreover, according to the boring log, the SPT-N value of the RS layer falls into the range of SPT-N values of AS layer. Also, the main components of AS are clay, sand, and gravel mixed with silt, while that of RS is sand with some silt. Therefore, this study assumed that the RS of this site is part of AS, considering their

similarity in engineering perspective and reliability of the estimation results. Since the RS layer is too thin, it is hard to guarantee the reliability of the layer as the prediction error can exceed the thickness of the layer, which means that the estimation result is unreliable in any sense.

4.4 Site-specific Subsurface Stratification Estimation with Integrated Dataset

In this study, integrated analysis refers to a subsurface stratification estimation method that uses integrated geotechnical survey data set. The primary spatial interpolation method used for integrated analysis are geostatistical methods such as kriging and simulation methods. Both methods follow certain common steps before the estimation. These steps are primarily required to get the optimal estimation of the subsurface stratification based on cross-validation. Through the cross-validation-based quantitative optimization, integrated analysis is expected to provide a site-specific optimized estimation result, which takes the site's geotechnical characteristics into account.

4.4.1 Integrated Analysis using Kriging Method

The kriging based integrated analysis requires several additional steps compared the rather conventional borehole data-only estimation method. First of all, it uses different dataset defined by layer boundaries and assumes thresholds listed in Table 3.4. Then, each data set is ready to be applied to the analysis. For a layer boundary, seven assumed thresholds are quantitatively evaluated to determine the optimal threshold velocity value that represents the P-wave velocity, V_p, at the boundary of the layers. The optimal V_p is determined by finding the minimum point of the 2nd order polynomial regression curve of the RMSE-threshold velocity plot. In this study, the ratios

of assumed thresholds were modified to enhance the curve fitting performance.

The finalized V_p is considered *optimal* as it is expected to be the velocity that gives minimum error when used for estimation of the target layer boundary. Moreover, it is also regarded as *site-specific* as the finalized V_p is not the same as conventional classification criteria proposed for general purposes by government departments or public enterprises. Unlike such criteria, the optimal V_p can differ by a target area. Thus, it can be said that the classification criteria suggested through integrated analysis is a site-specific value. The integrated analysis has several advantages over the conventional methods that it uses more data points to propose a unique set of layer classification criteria.

The V_p -based optimal layer boundary classification criteria determined through cross-validation of the IK method with datasets before and after outlining are shown in Figures 4.8 and 4.9, respectively. The figures show 2nd-order polynomial regression curves with their minimum points marked in red.

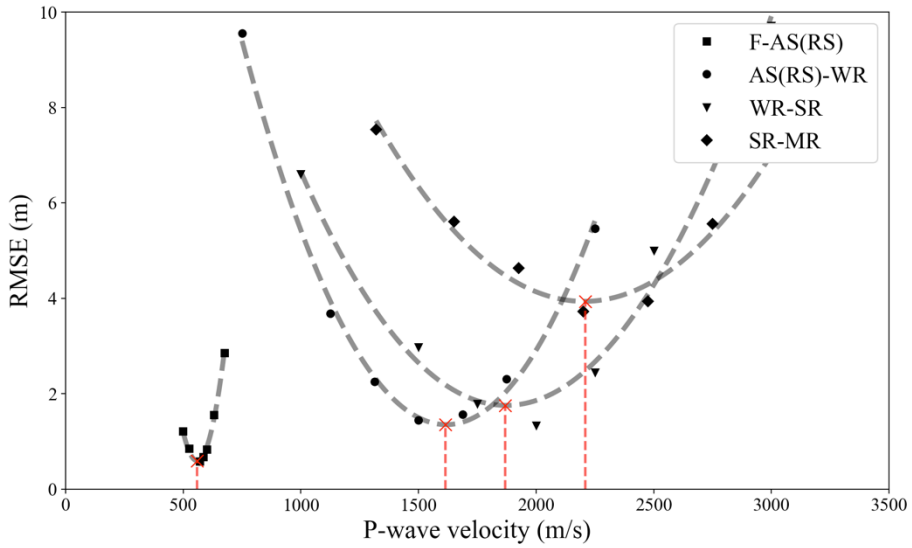


Figure 4.8 Site-specifically optimized P-wave velocity-based classification criteria determined with IK method using original dataset.

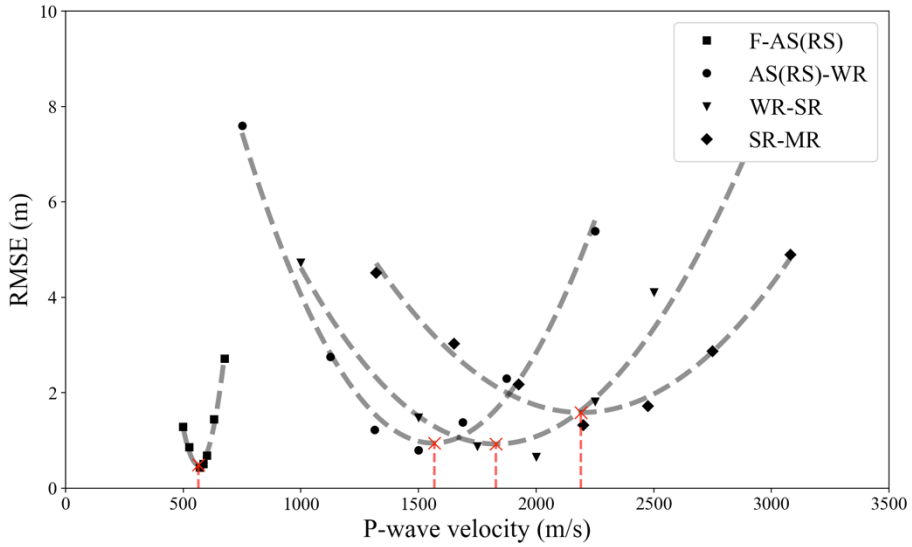


Figure 4.9 Site-specifically optimized P-wave velocity-based classification criteria determined with IK method using outlined dataset

4.4.2 Integrated Analysis using Simulation Method

The simulation method differs from the kriging-based method that it can give more realistic estimation results. Kriging is known to have a smoothing effect which flattens out the estimated values in between the sampled data points. Simulation method, however, estimates one point at a time by following a randomly generated path for a single realization. Single realization refers to the completion of estimations of entire nodes following the random path. Each realization has a unique random path; therefore, each of them shows different result. In this study, simple kriging was used for the node estimation. Moreover, the datasets were normal score transformed before the estimation, and then they were back-transformed using the original cumulative distribution function after one realization is completed.

Kim (2020) adopted ordinary kriging based SGS to integrated analysis for an dam emergency spillway construction site. In the previous study, the method of permissible error was introduced for SGS-based integrated analysis. The term *permissible error* refers to the maximum error allowed for each realization. In this study, total thousands of simulations were generated, and the top hundred realizations were selected and averaged to propose the optimal estimation result of SGS-based integrated analysis. For the simulation method, it is important to set an appropriate number of simulations as computational costs increase rapidly with the size of the target area, data density, and cell sizes. From the preliminary research, it was observed that higher data density and higher resolution, which is achieved by using smaller cell size, do not

necessarily guarantee improvements in estimation performance.

The V_p based optimal layer boundary classification criteria determined through cross-validation of the SGS method with datasets before and after outlining are shown in Figures 4.10 and 4.11, respectively. The figures show 2nd-order polynomial regression curves with their minimum points marked in red.

The summary of site-specifically optimized classification criteria for all six conditions are shown in Table 4.2. With these optimal classification criteria, the integrated dataset is selected and then applied to the final cross-validation to quantitatively evaluate and compare the performance of each method and the effects of outlier analysis.

Table 4.2 Summary of the optimal P-wave velocity determined from the cross-validation results of three methods (OK, IK, SGS) using datasets before and after outlining.

Optimal V_p	Before outlining		After outlining	
(m/s)	IK	SGS	IK	SGS
F-AS(RS)	558.8	560.6	564.2	564.2
AS(RS)-WR	1613.6	1598.5	1568.2	1522.7
WR-SR	1868.7	1909.1	1828.3	1949.5
SR-MR	2208.9	2013.3	2191.1	2440.0

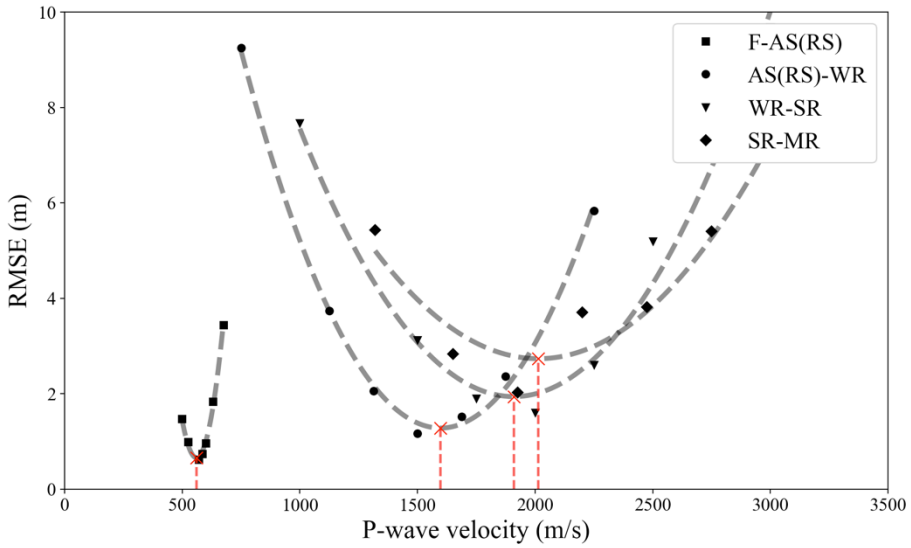


Figure 4.10 Site-specifically optimized P-wave velocity-based classification criteria determined with SGS method using the original dataset

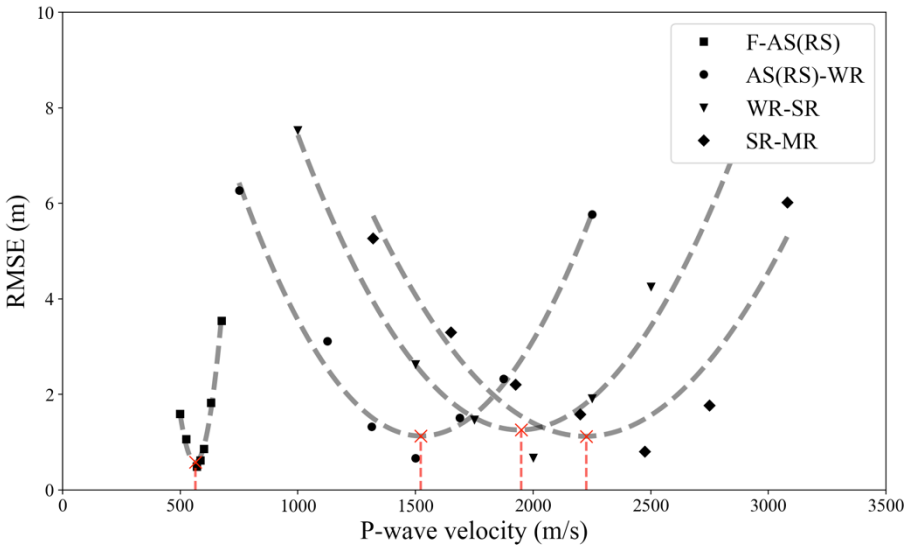


Figure 4.11 Site-specifically optimized P-wave velocity-based classification criteria determined with SGS method using outlined dataset

4.5 Performance Evaluation of the Estimation Methods

For all three methods (OK, IK, and SGS), leave-one-out cross-validation was performed to compare the performance of the estimation results quantitatively. The leave-one-out methodology requires estimation errors for all data points. Since the size of the target area and the number of tested data points are relatively small, leave-one-out cross-validation was performed without computational performance problems. In this study, borehole survey data are considered true data, and thus, cross-validation was done for vertical borehole data only. For each method and layer boundaries, one borehole data was removed, and then estimated without the point to get the estimation error at the removed point. This step was repeated for all vertical borehole data points possible.

The performance of each method was evaluated for all four layer boundaries before and after outlining both borehole and geophysical survey data. Table 4.3 shows the summary of the performance of each condition. As it can be noticed from the summary, there is a clear improvement in the performance of each method after outlining.

For F-AS(RS), AS(RS)-WR, and SR-MR, the SGS method showed the smallest RMSE followed by IK and OK after outlining. However, for estimation of the WR-SR layer boundary, the OK method, which only used borehole data, showed the best performance. There are two possible justifications for such a phenomenon. Firstly, the curves for WR-SR layer boundaries are spread laterally, which makes it harder to find an appropriate

minimum point. A potential weakness of such an argument can arise from the curves of SR-MR, which is also laterally spread. However, for SR-MR curves, by looking at the data points that were used for regression, there is a solid noticeable minimum point even without the regression curves. Thus, the widening of the SR-MR curve is due to nearby points and the process of regression, not by the characteristics of the data itself. Unlike SR-MR, which has an evident minimum, WR-SR curves show no apparent minimum without the regression curve; therefore, it is hard to tell that the finalized classification criteria are the optimal V_p that identifies the WR-SR boundary of the target stie.

Table 4.3 Summary of RMSE values of cross-validation results of three methods (OK, IK, SGS) using datasets before and after outlining.

RMSE	Before outlining			After outlining		
(m)	OK	IK	SGS	OK	IK	SGS
F-AS(RS)	0.76	0.67	0.60	0.55	0.54	0.50
AS(RS)-WR	1.12	1.97	1.25	1.23	1.20	0.71
WR-SR	0.92	1.60	1.70	0.62	0.93	0.70
SR-MR	4.70	4.60	1.78	1.19	0.95	0.80
Average	1.88	2.21	1.33	0.90	0.91	0.68

4.6 Site-specific Subsurface Stratification Analysis Result

After performing cross-validation-based performance analysis for all conditions, the site-specifically optimized subsurface stratification models were developed for all three methods (OK, IK, and SGS) using outlined datasets and optimal classification criteria depending on the methods.

For each method, 2-dimensional elevation maps of the four layer boundaries were constructed. Then the estimation errors obtained from leave-one-out cross-validation were interpolated with ordinary kriging for each layer boundary. In order to construct a 3-dimensional profile, the surface level map was developed using ordinary kriging and ground elevation data collected from the boring logs. After developing 2-dimensional map for each layer boundary, including the ground surface, all of the soil and rock layer volumes were constructed by stacking layers boundaries on 3-dimensional space based on the x- and y- coordinates and estimated property value, which is the elevation of the point. A 37m × 26m × 81m sized soil volume was formulated by setting the minimum elevation to the deepest point of boring. The 2-D maps were plotted with the popular python library Matplotlib, and 3-D volumes were visualized with PyVista, a python library that offers powerful tools for 3-D visualization.

The 2- and 3-dimensional visualizations of the site-specifically optimized soil profiles for all three methods are shown in the Appendices.

Chapter 5 Conclusions

As only limited information about the ground is available, many researchers have adopted geostatistical approaches to interpolate uninvestigated points. This study applied three different geostatistics-based subsurface stratification estimation methods to propose the optimal estimation method and site-specifically optimized soil profile of a small section of the subway line construction site near the ground subsidence zone. The following conclusions are drawn from this study.

(1) The borehole and geophysical survey data were digitized and standardized. The digitization procedure for geophysical survey data was simplified by reconstructing unstructured tomography with python, a popular programming language.

(2) Outlier analysis was performed for borehole datasets through the leave-one-out cross-validation method. For geophysical survey data, tomography lines affected by the excavation of the tunnel and grouting at the face of the tunnel were eliminated.

(3) Threshold P-wave velocity was determined considering the guidelines provided by Seoul Metropolis, Korea Ministry of Land, and previous studies. However, the ratios of assumed classification criteria were modified based on the site conditions through preliminary experiments. Based on the assumed classification criteria, the optimal P-wave velocities that indicate layer boundaries were determined through integrated analyses.

(4) Three estimation algorithms were developed with python. The three estimation methods are borehole-based kriging, integrated analysis using kriging, and integrated analysis using the simulation method. Their estimation performances were assessed by RMSE values computed by leave-one-out cross-validation.

(5) After outlining and preprocessing the datasets, the RMSE of all three methods decreased significantly. Therefore, outlining data was shown to have a positive effect on estimation performance.

(6) SGS, the simulation-based integrated analysis method, showed the most reliable estimation results for layer boundaries of F-AS(RS), AS(RS)-WR, and SR-MR. For these three boundaries, IK was slightly better than OK.

(7) OK, the borehole-based kriging method was most reliable for estimating the WR-SR layer boundary. Integrated analysis methods were particularly struggling for this boundary as finding the appropriate minimum point of the 2^{nd} order polynomial curve was difficult for the WR-SR case.

(8) Site-specifically optimized estimation method for subsurface stratification has been quantitatively analyzed, and the final results were visualized in 2-D and 3-D. Therefore, further studies are needed to enhance the reliability of integrated methods in various conditions.

References

Ajvazi, B., & Czimber, K. (2019). A comparative analysis of different dem interpolation methods in GIS: Case study of rahovec, Kosovo. *Geodesy and Cartography, 45*(5), 43-48.

Arun, P. V. (2013). A comparative analysis of different dem interpolation methods. *Geodesy and Cartography, 39*(4), 171-177.

Beckman, R. J., & Cook, R. D. (1983). Outlier ..... s. *Technome Trics, 25*, 119-149.

Burrough, P. (1987). *Principles of Geographical Information Systems for land resources assessment*. Oxford: Clarendon Pr.

Chilès, J., & Delfiner, P. (2007). *Geostatistics: Modeling spatial uncertainty*. New York u.a.: Wiley.

Chun, S., Sun, C., & Chung, C. (2005). Application of Geostatistical Methods for Geo-Layer Information. *KSCE Journal of Civil and Environmental Engineering Research*.

Cowen, D. J. (1988).). GIS versus CAD versus DBMS: What are the differences? *Photogrammetric Engineering and Remote Sensing, 54*, 1551-1554.

Cressie, N., & Hawkins, D. M. (1980). Robust estimation of the variogram: I. *Journal of the International Association for Mathematical Geology, 12*(2), 115-125.

Dawson, K. M., & Baise, L. G. (2005). Three-dimensional liquefaction potential analysis using geostatistical interpolation. *Soil Dynamics and Earthquake Engineering, 25*(5), 369-381

Delmonaco, G., Leoni, G., Margottini, C., Puglisi, C., & Spizzichino, D. (2003). Large scale debris-flow hazard assessment: A geotechnical approach and GIS modelling. *Natural Hazards and Earth System Sciences, 3*(5), 443-455.

Deutsch, C. V., & Journel, A. G. (1993). *GSLIB: Geostatistical Software Library and user's guide*. Oxf. U.P. (N.Y.).

Emery, X., & Peláez, M. (2011). Assessing the accuracy of sequential gaussian simulation and Cosimulation. *Computational Geosciences, 15*(4), 673-689.

Eurocode 7: Geotechnical design, European Committee for Standardization (2007).

Gallerini, G., & De Donatis, M. (2009). 3D modeling using geognostic data: The case of the low valley of Foglia River (Italy). *Computers & Geosciences, 35*(1), 146-164.

GhojehBeyglou, M. (2021). Geostatistical modeling of porosity and evaluating the local and global distribution. *Journal of Petroleum Exploration and Production Technology, 11*(12), 4227-4241.

Goovaerts, P. (1997). Geostatistics for natural resources evaluation. *Oxford University Press on Demand*.

Isaaks, E. H., & Srivastava, R. M. (1989). *Applied geostatistics*. New York: Oxford University Press.

Journel, A. G. (1986). Geostatistics: Models and tools for the Earth Sciences. *Mathematical Geology, 18*(1), 119-140.

Journel, A. G. (1989). Fundamentals of geostatistics in five lessons.

Kim, H., Chung, C., & Kim, H. (2016). Geo-spatial data integration for subsurface stratification of dam site with outlier analyses. *Environmental Earth Sciences, 75*(2).

Kim, H., Kim, M., Kim, J., Kim, K., & Chung, C. (2012a). Geostatistical integration of borehole and geophysical data for design of offshore-foundation. *Journal of the Korean Geotechnical Society, 28*(5), 109-120.

Kim, H., Kim, H., Shin, S., & Chung, C. (2012b). Application of statistical geo-spatial information technology to soil stratification in the Seoul Metropolitan Area. *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards, 6*(4), 221-228.

Kim, H., Kim, J., & Chung, C. (2014). Geostatistical Integration Analysis of geophysical survey and borehole data applying Digital Map. *Journal of the Korean Geoenvironmental Society, 15*(3), 65-74.

Kim, J. (2015). *3D Geostatistical Integration Based on Site-specific Correlation with Borehole data and Geophysical data*(Unpublished doctoral dissertation, 2015). Seoul National University.

Kim, M. (2020). *Optimization of Three-dimensional Geostatistical Integration of Site Investigation Information*(Unpublished doctoral dissertation, 2020). Seoul National University.

Kitanidis, P. K. (1997). Introduction to geostatistics: Applications in hydrogeology. *Cambridge University Press*.

Krige, D. G. (1951). A statistical approach to some basic mine valuation problems on the witwatersarand. *Journal of the Chemical Metallurgical & Mining Society of South Africa, 52*(6).

Korea Ministry of Land, Infrastructure and Transport (2012) Construction estimate manual. Korean Geotechnical Society, Seoul, pp 152–155

Madani Esfahani, N., & Asghari, O. (2012). Fault detection in 3D by sequential gaussian simulation of Rock Quality Designation (RQD). *Arabian Journal of Geosciences, 6*(10), 3737-3747.

Matheron, G. (1963). Principles of Geostatistics. *Economic Geology, 58*, 1246-1266.

McDonnell, R. A., & Burrough, P. A. (1998). *Principles of Geographical Information Systems*. Oxford: OUP.

Nemoto, T., Masumoto, S., Nonogaki, S., Raghavan, V., & Yonezawa, G. (2015). Development of a Distributed Database System for Borehole Data using Free and Open Source Software. *International Journal of Geoinfomatics, 11*(3).

Oh, S., Chung, H., & Lee, H. (2004). Geostatistical integration of MT and borehole data for RMR Evaluation. *SEG Technical Program Expanded Abstracts 2004*.

Olea, R. A. (1999). *Geostatistics for engineers and Earth scientists*. Boston, MA: Springer US.

Parker, H. D. (1988). The unique qualities of a geographic information system: A commentary. *Photogrammetric Engineering and Remote Sensing, 54*, 1547-1549.

Phillips, D. L., Dolph, J., & Marks, D. (1992). A comparison of geostatistical procedures for spatial analysis of precipitation in mountainous terrain. *Agricultural and Forest Meteorology, 58*(1-2), 119-141.

Pinheiro, M., Vallejos, J., Miranda, T., & Emery, X. (2016). Geostatistical simulation to map the spatial heterogeneity of geomechanical parameters: A case study with rock mass rating. *Engineering Geology, 205*, 93-103.

Pyrcz, M. J., & Deutsch, C. V. (2014). Geostatistical reservoir modeling. *Oxford University Press*.

Rigaux, P., Scholl, M., & Voisard, A. (2001). *Spatial databases with application to GIS*. San Francisco u.a.: Kaufmann.

Seoul Metropolis (2006) Geotechnical investigation handbook. Seoul Metropolis, Seoul, pp 7–9

Setianto, A., & Triandini, T. (2015). Comparison of Kriging and inverse distance weighted (IDW) interpolation methods in lineament extraction and analysis. *Journal of Applied Geology, 5*(1).

Shahbeik, S., Afzal, P., Moarefvand, P., & Qumarsy, M. (2013). Comparison between ordinary kriging (OK) and inverse distance weighted (IDW) based on estimation error. case study: Dardevey iron ore deposit, Ne Iran. *Arabian Journal of Geosciences, 7*(9), 3693-3704.

Sitharam, T. G., & Samui, P. (2007). Geostatistical modelling of spatial and depth variability of SPT data for Bangalore. *Geomechanics and Geoengineering, 2*(4), 307-316.

SMITH, T. R., MENON, S., STAR, J. L., & ESTES, J. E. (1987). Requirements and principles for the implementation and construction of large-

scale geographic information systems. *International Journal of Geographical Information Systems, 1*(1), 13-31.

Sun, C., & Kim, H. (2017). GIS-based regional assessment of seismic site effects considering the spatial uncertainty of site-specific geotechnical characteristics in coastal and inland urban areas. *Geomatics, Natural Hazards and Risk,8*(2), 1592-1621.

Verly, G. (1993). Sequential gaussian simulation: A Monte Carlo method for generating models of porosity and permeability. *Generation, Accumulation and Production of Europe's Hydrocarbons III,* 345-356. d

Yamamoto, J. K. (2005). Correcting the smoothing effect of ordinary Kriging estimates. *Mathematical Geology, 37*(1), 69-94.

Zimmerman, D., Pavlov, C., Ruggles, A., & Armstrong, M. P. (1999). An Experimental Comparison of Ordinary and Universal Kriging and Inverse Distance Weighting. *Mathematical Geology, 31*(4).

한국동력자원연구소 (1981) 한국 지질도(1:50,000): 뚝섬도폭

[Geological map of Korea (1:50,000): Ttukseom], Seoul, Republic of Korea

Appendix

A.1 Visualization of the site-specifically optimized soil profiles for all three methods.

The target area is illustrated in Figure A.1 with the outlined data. The 2-dimensional elevation maps of the layer boundaries at the target area estimated using OK, IK, and SGS methods are shown in Figures A.2, A.3, and A.4, respectively. The reliability maps shown in Figures A.5, A.6, and A.7 were constructed based on the estimation errors obtained from the cross-validation of the estimation methods. The purpose of the reliability map is to provide information about which part is more likely to give an accurate estimation. To make a reliability map, estimation errors were interpolated over the target area with ordinary kriging.

The final site-specifically optimized soil profiles for all three methods are shown in Figures A.8, A.10, and A.12. The same soil profiles are also shown in Figures A.9, A.11, and A.13 with decreased opacity. In these figures, both vertical and horizontal boring lines are also shown along with the excavated tunnel

Since it is hard to grasp the distribution of the stratification near the ground subsidence zone with the given figures, longitudinal section and cross-section views are also visualized in Figures A.15, A.16, and A.17. Plan view of the target area is shown in Figure A.14. In these figures, soil profiles along the

excavation direction are shown in the longitudinal section views, and the soil profiles at the face of the tunnel, where ground subsidence occurred, are shown in the cross-sectional views.

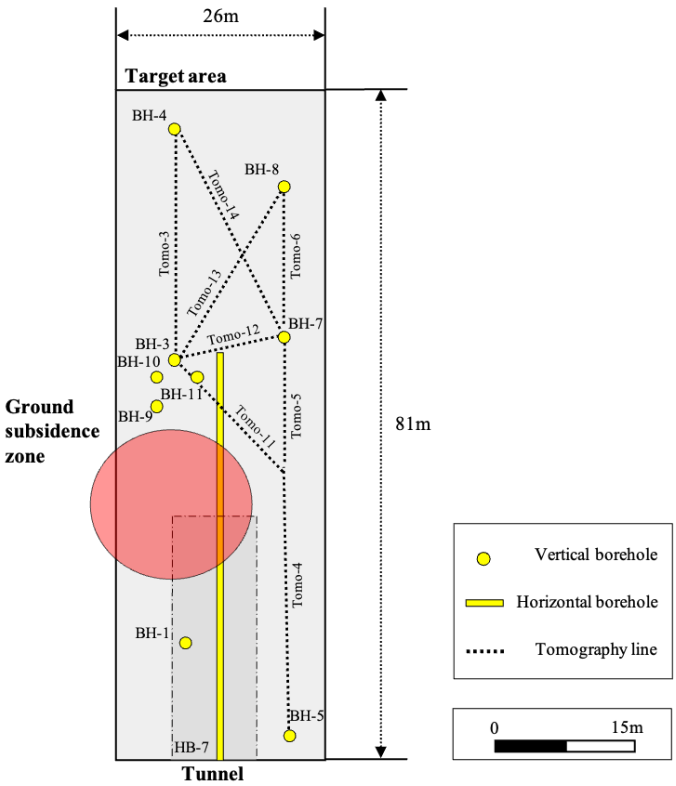
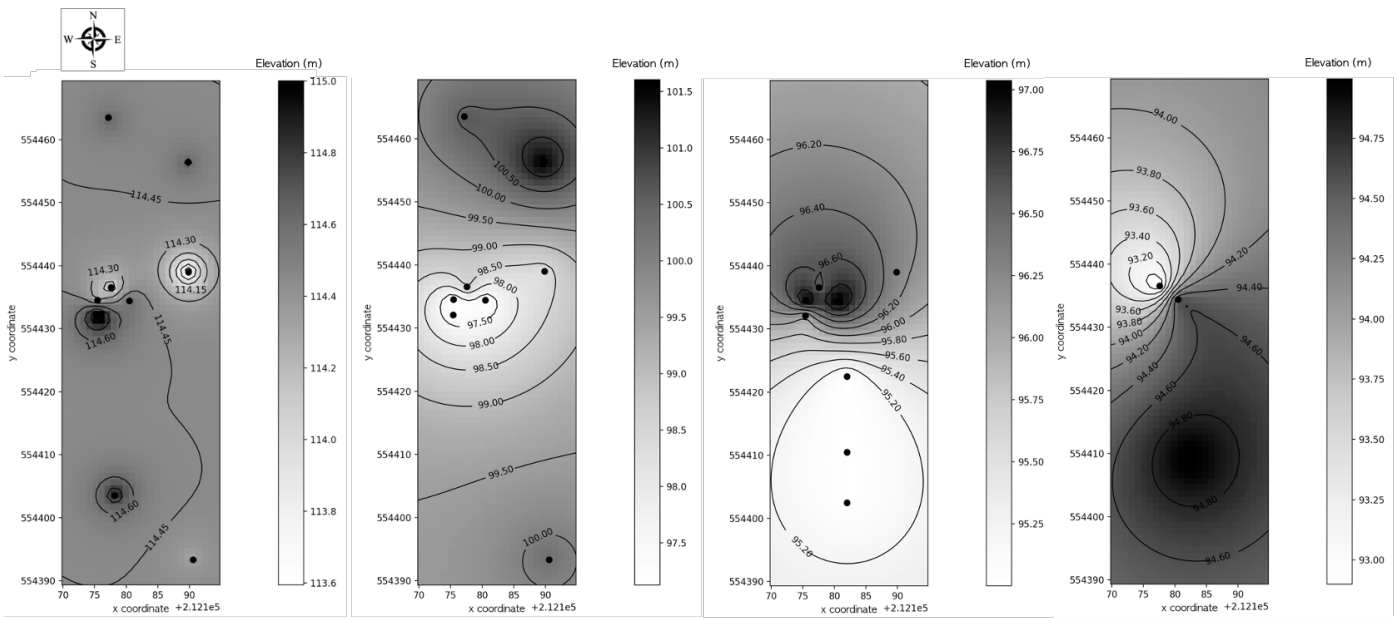


Figure A.1 Plan view of the target area.



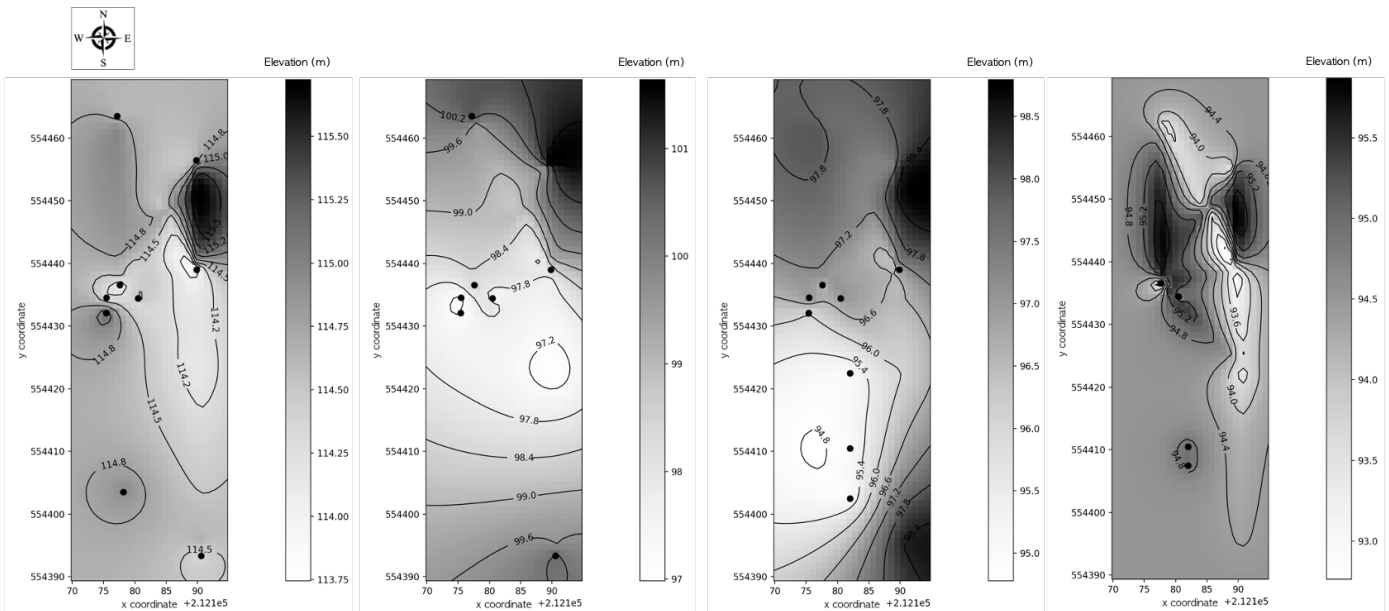
FS-AS(RS)
(RMSE: 0.55)

AS(RS)-WR
(RMSE: 1.23)

WR-SR
(RMSE: 0.62)

SR-MR
(RMSE: 1.19)

Figure A.2 2-D elevation maps of the final four layer boundaries estimated with the OK method.



FS-AS(RS)
(RMSE: 0.54)

AS(RS)-WR
(RMSE: 1.20)

WR-SR
(RMSE: 0.93)

SR-MR
(RMSE: 0.95)

Figure A.3 2-D elevation maps of the site-specifically optimized four layer boundaries estimated with the IK method.

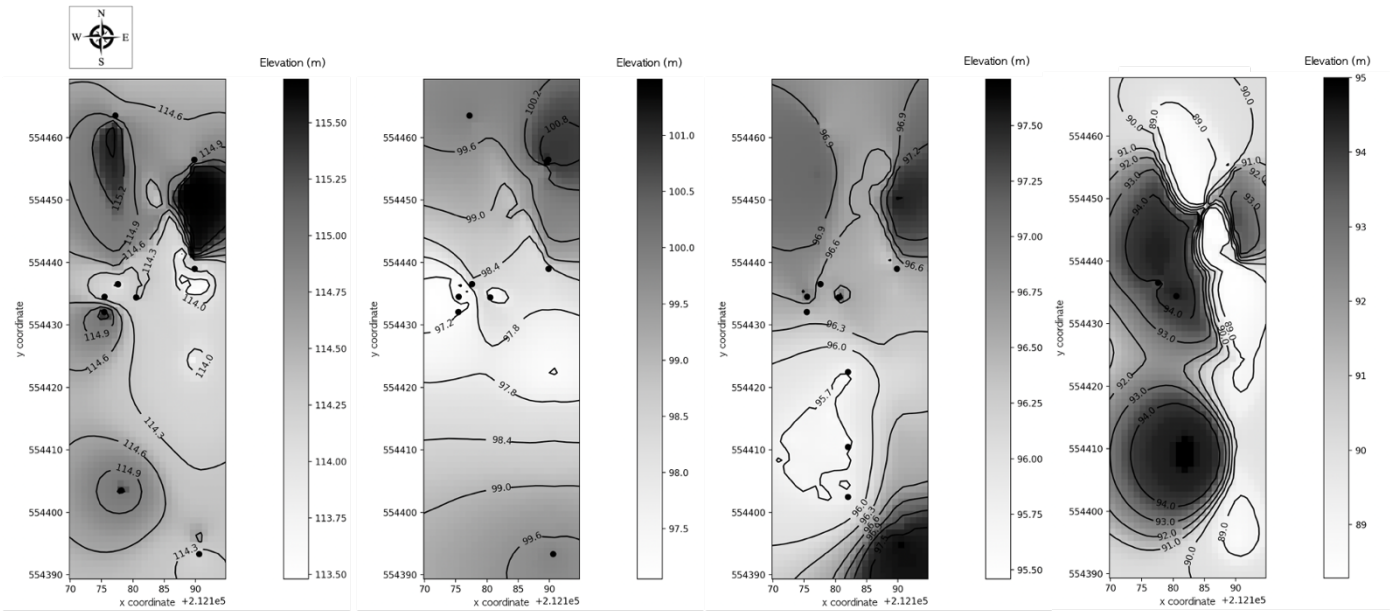


Figure A.4 2-D elevation maps of the site-specifically optimized four layer boundaries estimated with the SGS method.

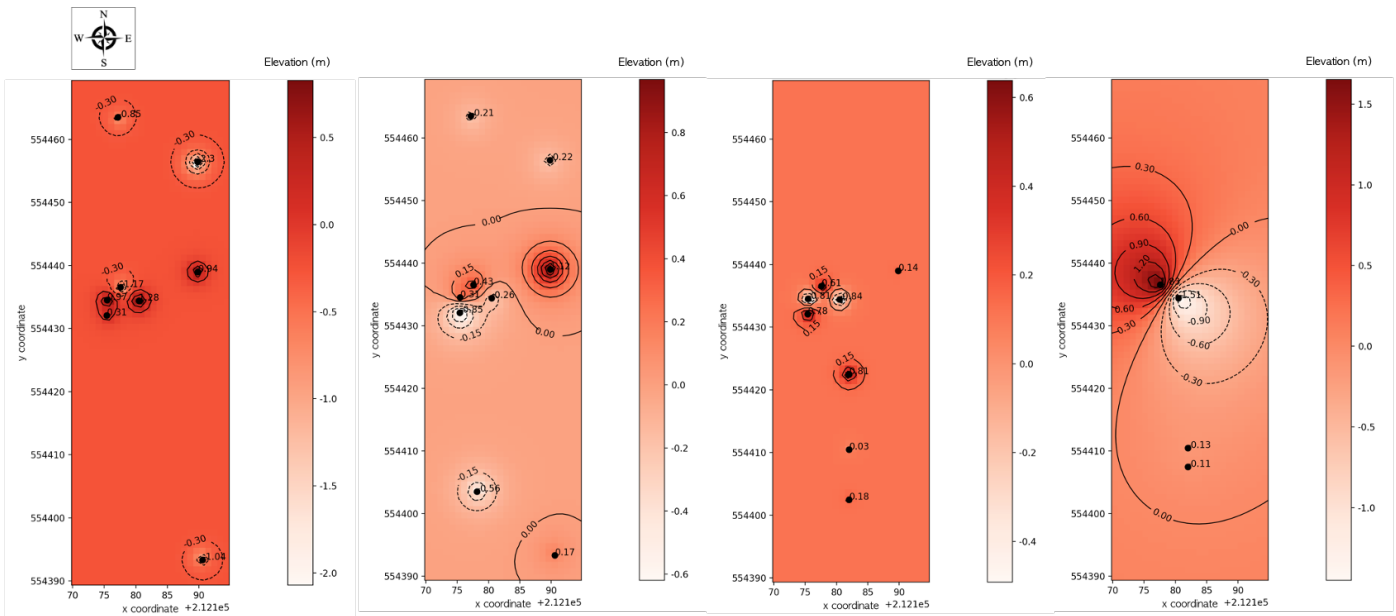
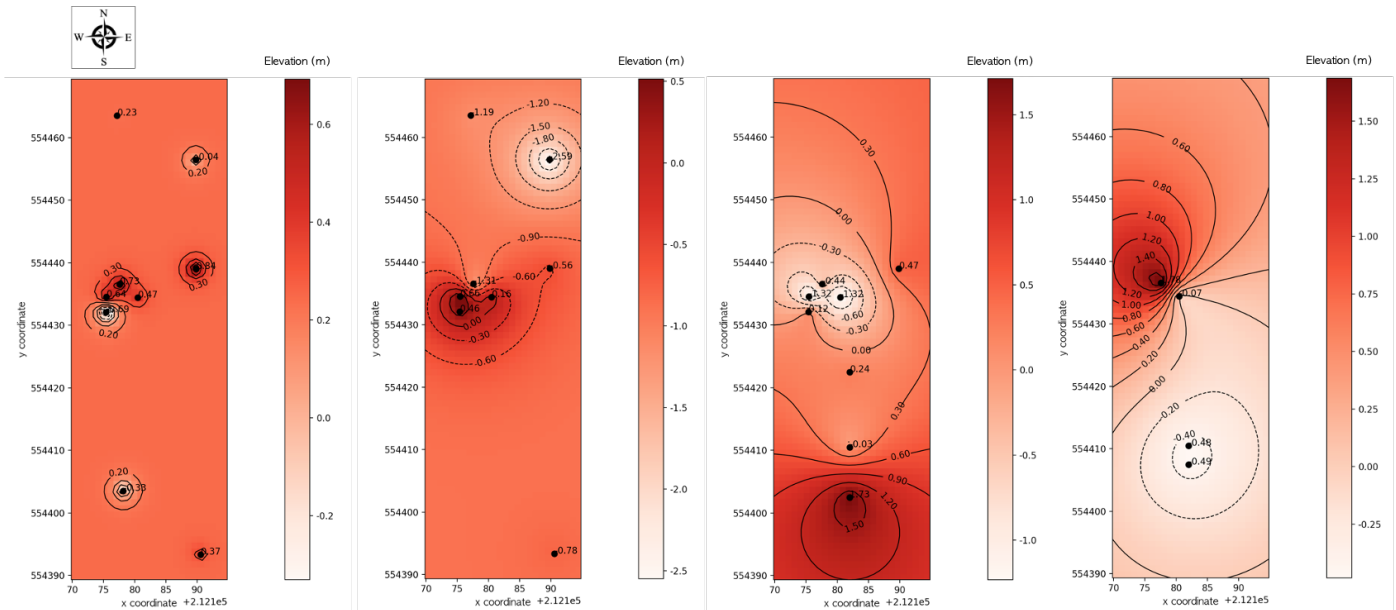


Figure A.5 2-D reliability map constructed by kriging estimation errors obtained from cross-validation of the OK method.



FS-AS(RS)
(RMSE: 0.54)

AS(RS)-WR
(RMSE: 1.20)

WR-SR
(RMSE: 0.93)

SR-MR
(RMSE: 0.95)

Figure A.6 2-D reliability maps constructed by kriging estimation error obtained from cross-validation of the IK method.

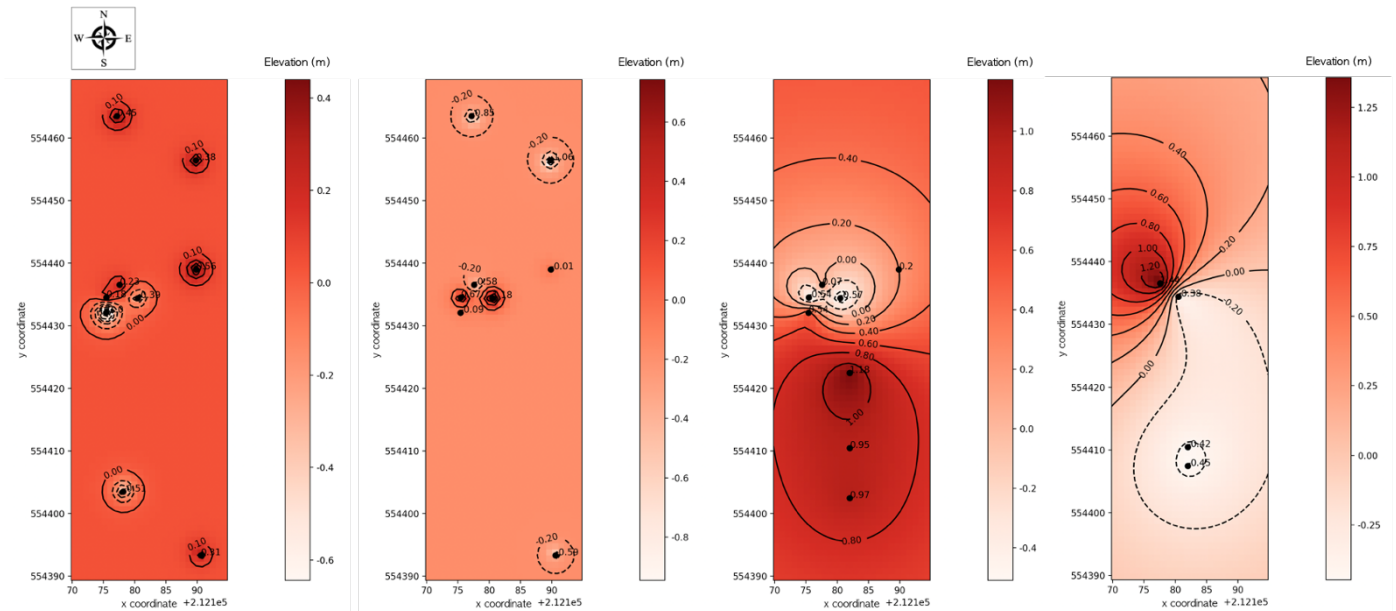


Figure A.7 2-D reliability maps constructed by kriging estimation error obtained from cross-validation of the SGS method.

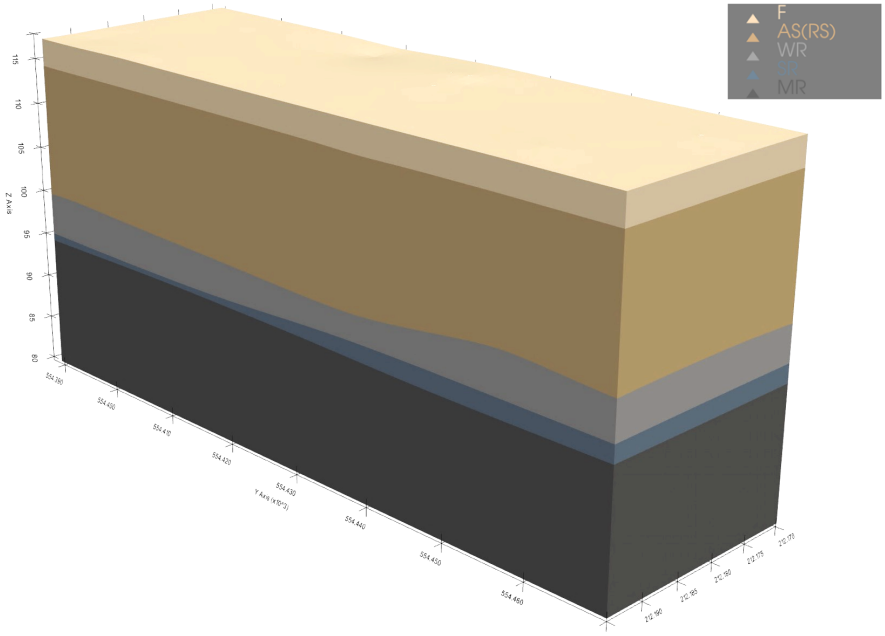


Figure A.8 The final subsurface stratification model developed using the OK method.

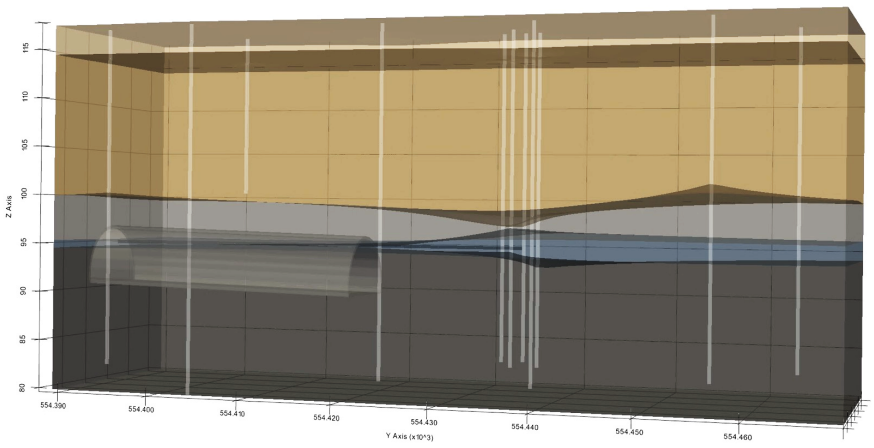


Figure A.9 Transparent version of the final subsurface stratification model developed using the OK method.

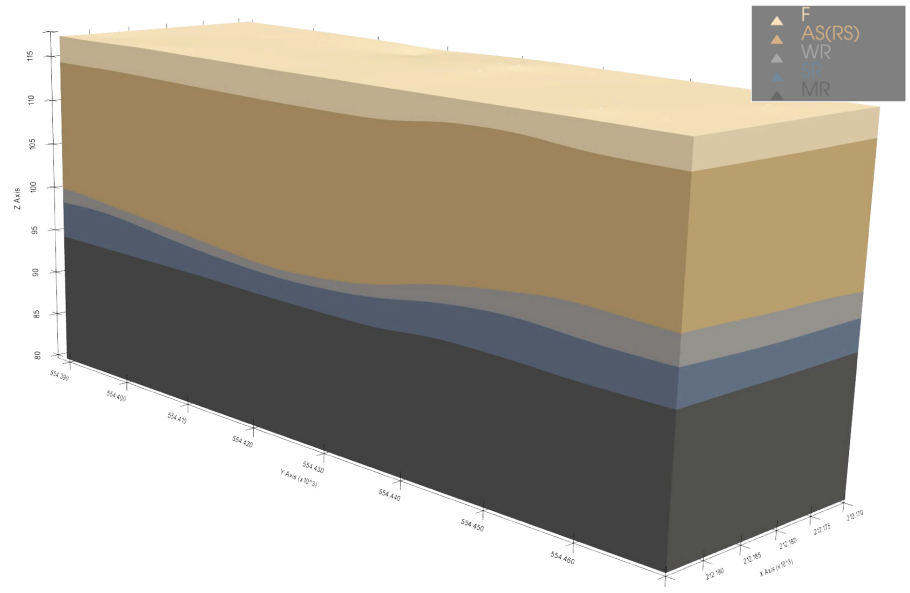


Figure A.10 The site-specifically optimized subsurface stratification model developed using the IK method.

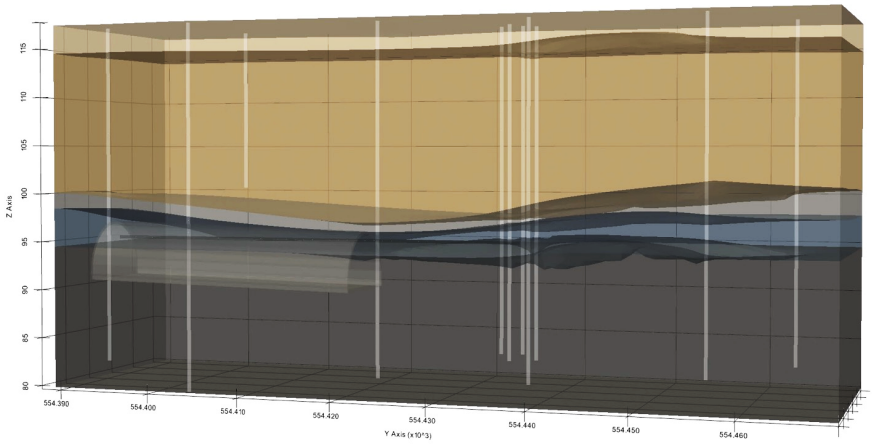


Figure A.11 Transparent version of the site-specifically optimized subsurface stratification model developed using the IK method.

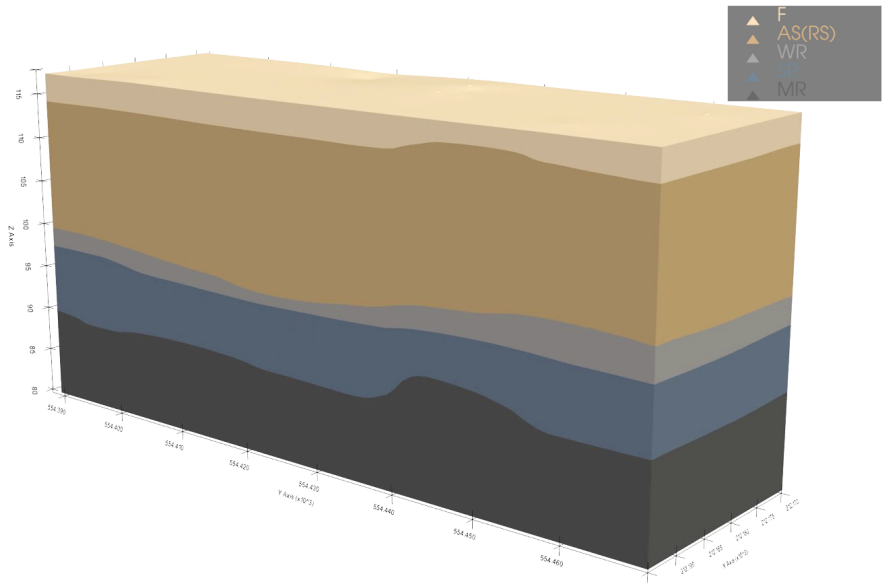


Figure A.12 The site-specifically optimized subsurface stratification model developed using the SGS method.

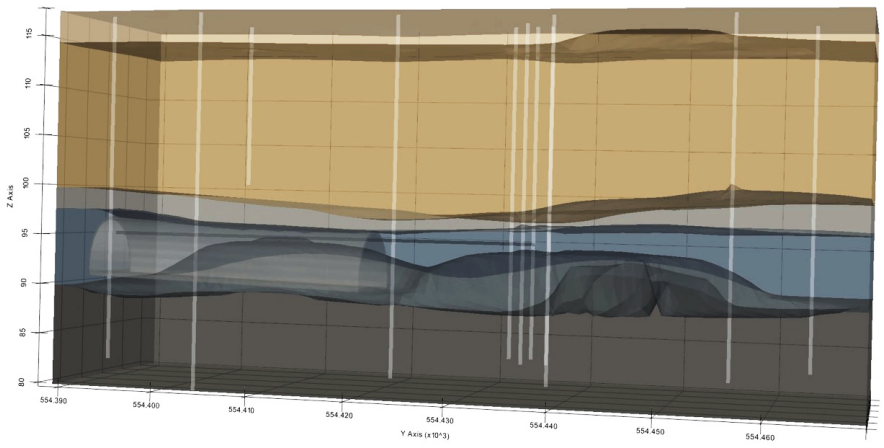


Figure A.13 Transparent version of the site-specifically optimized subsurface stratification model developed using the SGS method.

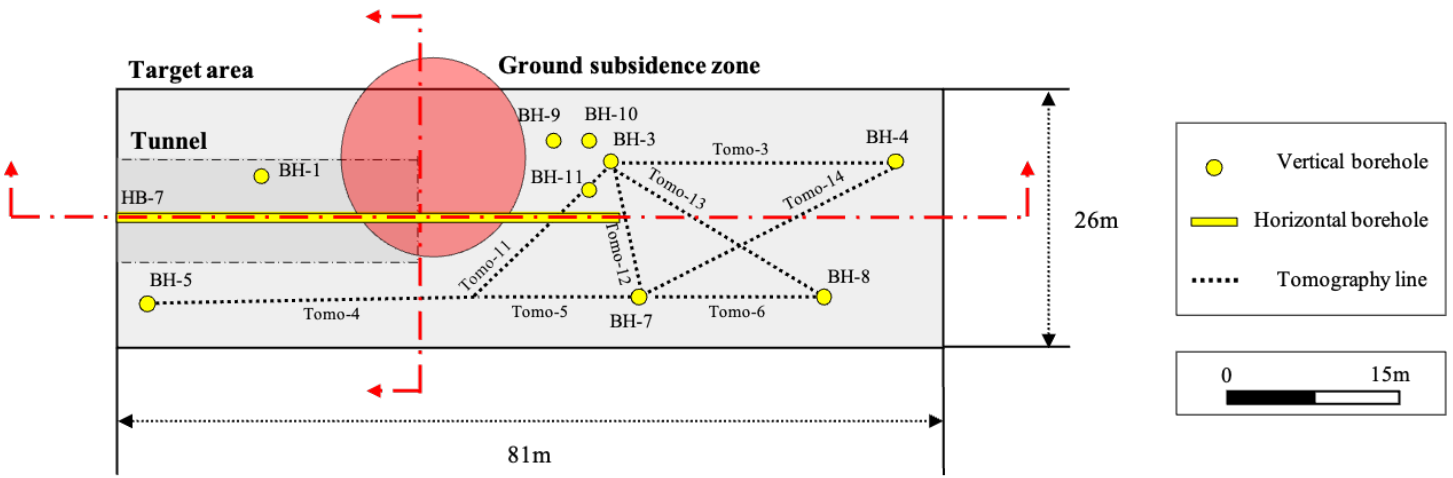
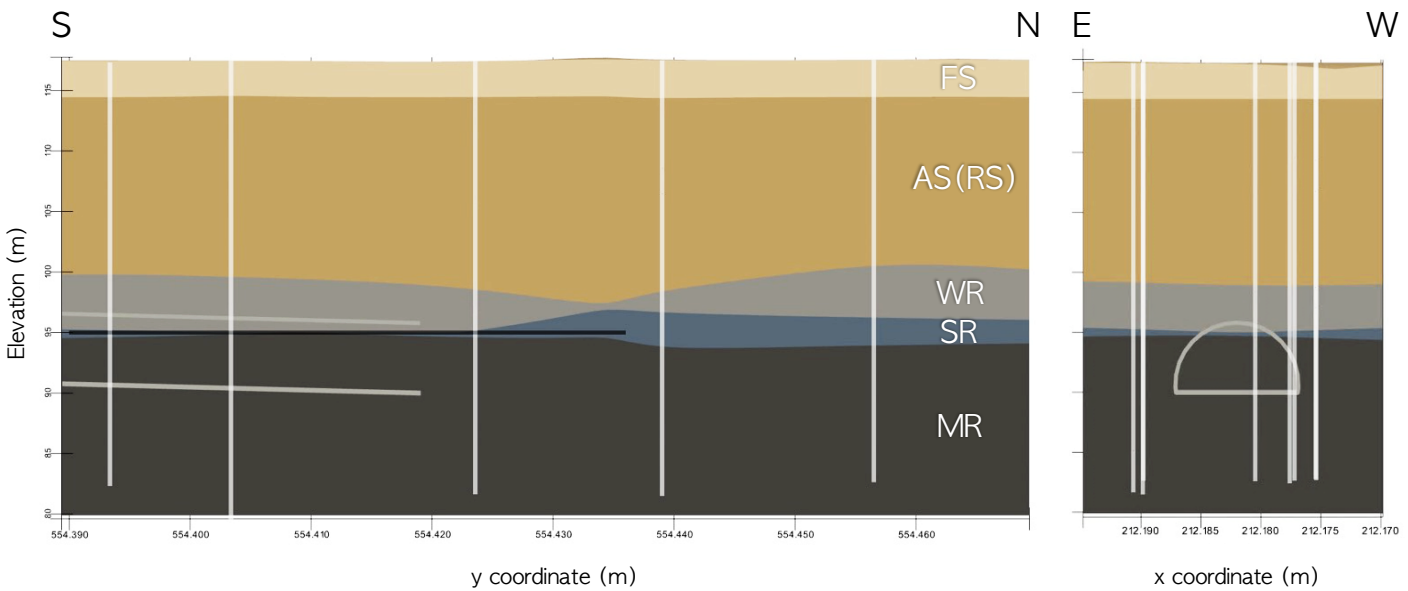


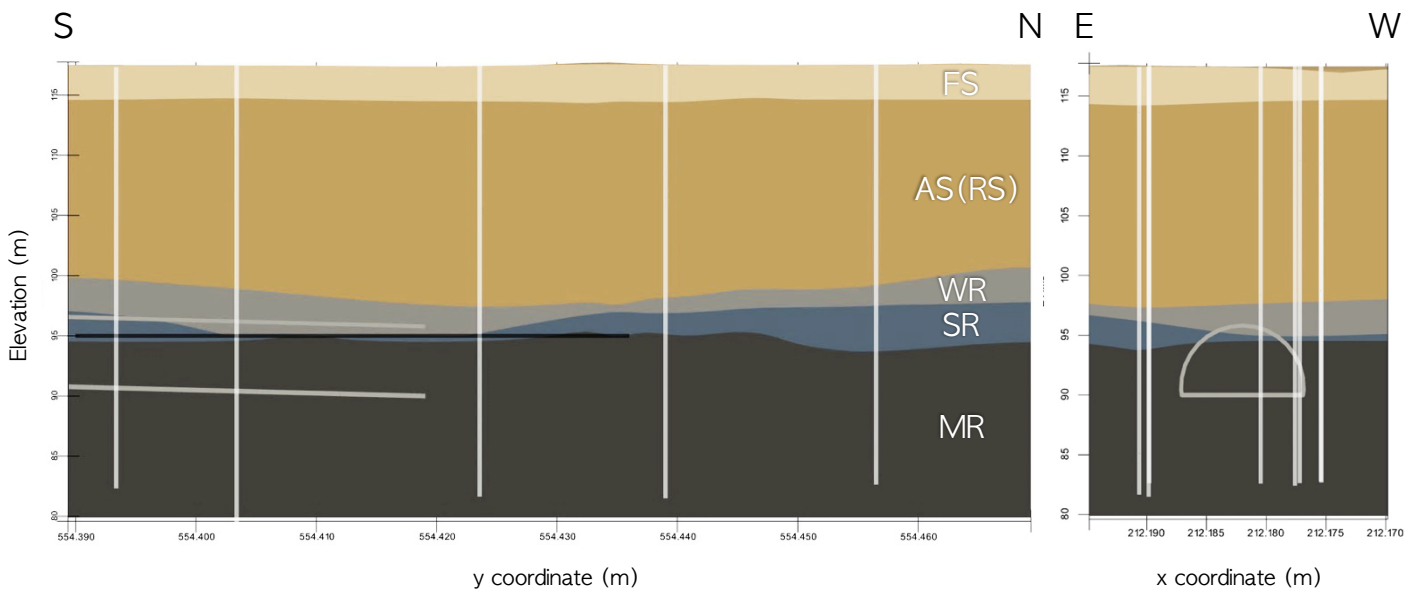
Figure A.14 Plan view of the target area with the cross-section lines.



Longitudinal section view from east

Cross section view from north

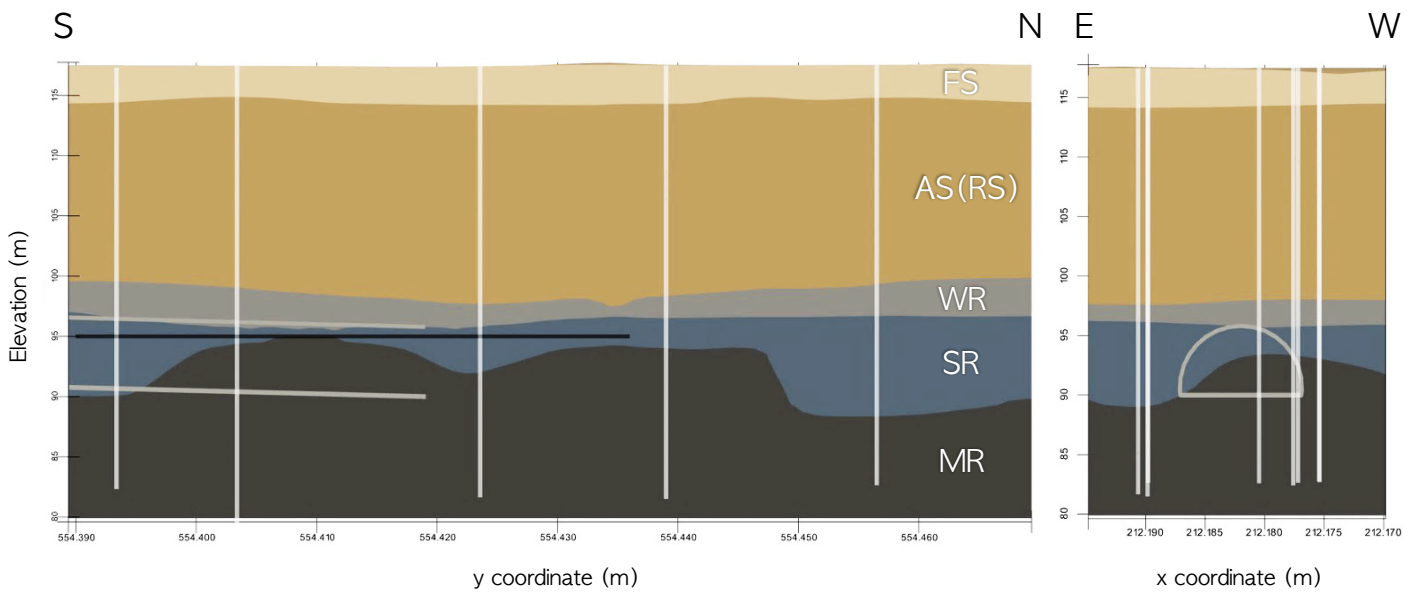
Figure A.15 Longitudinal section and cross-section of the site-specifically optimized soil profile developed by the OK method.



Longitudinal section view from east

Cross section view from north

Figure A.16 Longitudinal section and cross-section of the site-specifically optimized soil profile developed by the IK method.



Longitudinal section view from east

Cross section view from north

Figure A.17 Longitudinal section and cross-section of the site-specifically optimized soil profile developed by the SGS method

초 록

공간적 불확실성은 지반공학에서 가장 중요한 문제 중 하나이다. 특히 중요한 구조물을 건설하는 경우 예상치 못한 위험요소를 최소화 하기 위하여 충분한 양의 지반조사가 수행되어지는 것이 필수적이다. 그러나, 대부분의 경우 경제적 및 시·공간적인 한계로 인하여 수행 되어지는 지반조사의 수가 제한되게 된다. 따라서, 흙의 공간적 변동성과 지반에 대한 정보의 부족이란 한계를 극복하기 위해 지구통계학적 공간 보간기법들이 널리 사용되어져 왔다. 지반층상정보의 공간보간이 지구통계학적 기법의 가장 보편적인 적용 예이다. 지반층상정보는 가장 기본적인 지반공학적 정보로 공학자들에게 포괄적인 정보를 제공함에 있어 그 중요도는 간과될 수 없다. 그러므로, 신뢰도 높은 공간 공간분석을 통해 정확한 지반층상정보를 파악하는 것이 중요하다고 볼 수 있다. 선행연구에서는 현장 최적화된 지반층상정보 예측을 위해 통합분석법이 개발된 바 있다.

본 연구에서는 터널 굴착 중 지반함몰이 발생한 지하철 건설 현장 원지반에 대해 시추조사 정보 기반의 공간보간 기법과 두 종류의 통합분석법을 활용해 신뢰할 수 있는 현장 최적화된 지반층상정보 모델을 제시하고자 하였다. 기존의 방법과는 달리 통합분석법은 시추조사와 물리탐사로 얻어진 정보를 통합하여 활용한다는 특징이 있으며, 본 연구에서는 크리깅과 시뮬레이션에 기반한 통합분석법이 적용되었다.

분석에 앞서 지반조사 자료들은 디지타이징 (Digitizing) 후 표준화 하였다. 세 공간보간 기법들의 예측 신뢰도를 평가하기 위해 leave-one-out 방식의 교차검증 기법을 적용하였으며, 지반조사 자료들의 이상치 분석을 수행하여 예측 결과의 신뢰성에 끼치는 영향 또한 평가 하였다. 상기의 결과들을 바탕으로 각 기법별로 현장 최적화된 지반층상정보를 제시하였다.

주요어 : 현장 최적화, 지반층상, 지구통계학, 통합분석, 이상치 분석, 지반조사 자료의 통합
학 번 : 2020-28616