The use of R programming languages for geotechnical data processing and visualization: Techniques to manage geotechnical data for a post-closure tailings storage facility

Y Sun *Stantec, USA*

H Kalm *Stantec, USA*

E Hsiao *Stantec, USA*

PE Crouse *Stantec, USA*

Abstract

Mine closure requires regular updates of the site-specific knowledge base of the facility. Effective processing and visualization of the data acquired in tailings storage facilities (TSF) are vital in mine closure data management to ensure that all relevant information is evaluated. Digital geotechnical data can be stored and presented in systems such as Geographical Information Systems (GIS). Although GIS is a powerful tool for record keeping and spatial data visualization, it usually requires significant initial efforts to integrate the data into GIS. This study presents an alternative technique using R language to sort and present the digital data. A hypothetical TSF will be used to demonstrate the workflow that includes specifying the suitable data formats, pre-processing the data to check for inconsistencies, loading the data into the R scripts, extracting the relevant data columns for the desired calculations, performing the calculations, interpreting the results, and exporting the results. R language has a comprehensive library that provides a wide selection of statistics and graphics packages that makes the data analysis and visualization flexible. For the data visualization, this study will demonstrate several examples to display the geotechnical data in a one-dimensional drilling log that documented soil properties at one spot, a two-dimensional subsurface cross-section can be used to delineate subsurface layering, and a three-dimensional model that can be used to identify potential areas of concern. The choice of data processing and visualization techniques will depend on the specific geotechnical data being analyzed and the questions being asked. It is often useful to use a combination of techniques to get a comprehensive understanding of the subsurface conditions at a site. The techniques discussed in this study have been demonstrated to be a process that is flexible to expand and apply to knowledge bases of other mine closure disciplines.

Keywords: *geotech data analysis; data visualization; mine closure knowledge base; CPT data processing*

1   Introduction

A knowledge base is the site-specific information that is developed throughout the Life of Asset (LoA) for the various elements of the mine. In the *Integrated Mine Closure: Good Practice Guide* published by International Council on Mining and Metals (ICMM), the knowledge base is considered as one of the key elements of the mine closure planning cycle as shown in Figure 1 (ICMM, 2019). Regarding the knowledge base, the ICMM members considered the following questions are important aspects in the *Good Practice Guide*, including the following:

- Is the knowledge base updated regularly as data is collected and reviewed in an iterative process over the mine life?

- Is the data management protocol in place to ensure that ongoing monitoring and field trials are incorporated into the knowledge base?

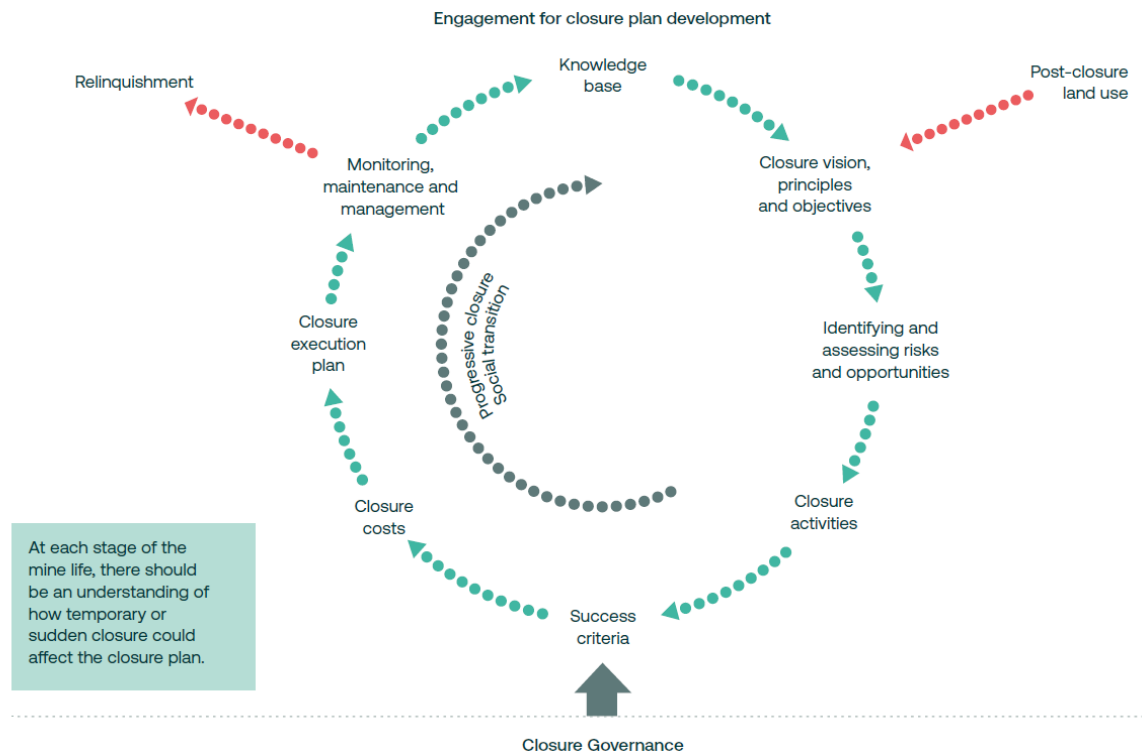- Has the spatial data been generated showing the mining operation area?

The use of R programming languages for geotechnical data processing and visualization: Techniques to manage geotechnical data for a post-closure tailings storage facility

Y Sun & H Kalm & E Hsiao & PE Crouse

Figure 1   Elements of closure planning (ICMM, 2019)

To achieve these objectives, we must have a powerful system designed to meet the ongoing data management and visualization needs. One of the most challenging types of data to manage is the geotechnical data since the geotechnical investigations and the instrumentation monitoring generate large volumes of data through the life of the mine. It especially difficult to integrate such wide ranges of data types when they come from various sources and formats.

Geographic Information System (GIS) technology can be used as an efficient tool to store, organize, analyze, and display this type of data in the form of digital layers and maps (Player, 2006). GIS provides promising outcomes for integrating multiple data types, such as photos, scanned maps, Shapefiles, or computer-aided drafting (CAD) files (Singh, 2018). However, GIS is not designed for geotechnical data analysis and is mostly used to visualize the final results of the geotechnical analysis. It lacks the specialized functionalities tailored to the unique characteristics of geotechnical data, such as geotechnical calculations, which hinders the real-time visualization of the analysis results.

This study proposes an alternative tool for geotechnical data management using the R programming language (R Core Team, 2021). As a popular programming environment for statistical computing and graphics, R can be the one-stop solution for both powerful data calculations and high-quality plots. It is especially valuable for geotechnical data where visualization of data through plots, maps, and charts helps understand patterns, trends, and relationships in the data. R's flexible plotting functionalities allow customization and interactive exploration of the data. R is an open-source software, which means it is freely available to use and easier to access compared to some other commercial applications, such as MATLAB. Companies can encourage geotechnical engineers to explore the capabilities of R without incurring substantial software licensing expenses.

Figure 2 illustrates a simplified process of data management discussed in this work that includes three steps: data acquisition, data integration and data validation. Instead of providing a synthesized manual to data management, this study aims to demonstrate a workflow using R and therefore only the key information

The use of R programming languages for geotechnical data processing and visualization: Techniques to manage geotechnical data for a post-closure tailings storage facility

Y Sun & H Kalm & E Hsiao & PE Crouse

within the geotechnical engineering discipline is listed in Figure 2. Section 2 of this paper discusses in more details each step in the workflow.
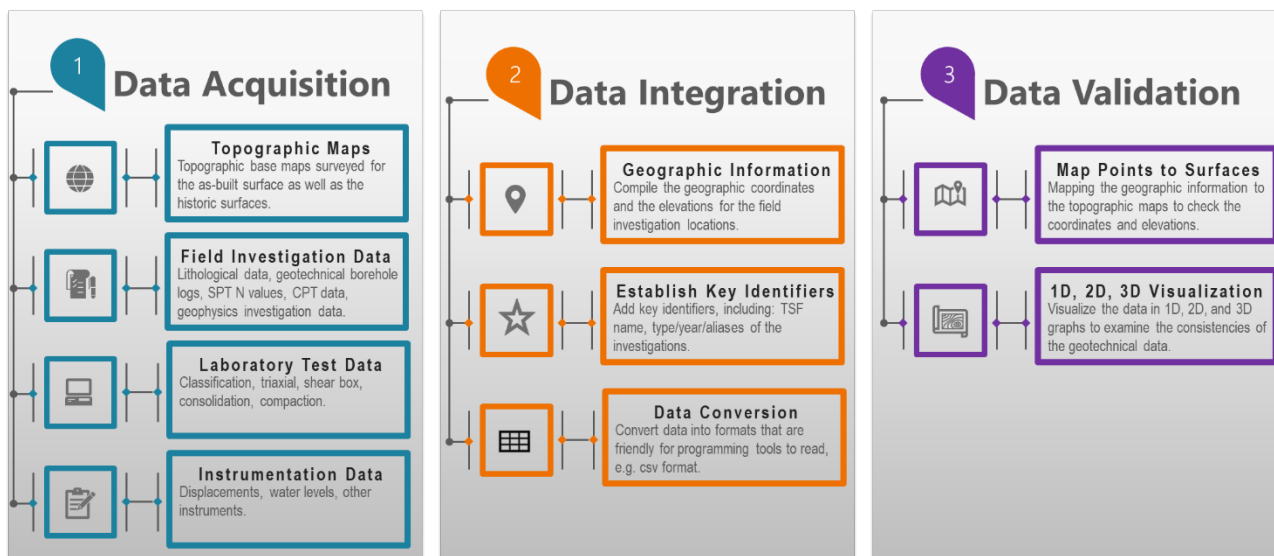

Figure 2   Diagram of the streamlined data acquisition and process workflow

2   Geotechnical data management

2.1   Data acquisition

Data acquisition can involve several sources of information. As the first step, a three-dimensional (3D) model is the basis to map the geotechnical information into a spatially represented database. The topographic maps of the area not only for the as-built surface but also for the surfaces over the life of the facility. The data should be gathered to improve the accuracy and resolution of the 3D model.

After the 3D model is established, field investigation data, laboratory test results, and instrumentation monitoring data can vary in complexity depending on the specific investigation or monitoring instrument. Prior to performing a geotechnical investigation, a desktop study should be conducted to collect existing information about the site. This helps provide insights to select the investigation locations and define the objectives of the subsequent investigation. Electronic formats of the collected data should be gathered to minimize manual data entry in the next steps of the analysis.

Coordinating the field investigation involves managing various tasks and resources to ensure an efficient data acquisition process. Table 1 provides an example of roles and responsibilities for the field team and the office crew to improve efficiencies in the data acquisition process. In this example, the team was tasked to complete the Cone Penetration Tests (CPT) and drilling activities. The personnel from the drilling company and one field geologist from the consulting company were required per rig. Additionally, the consulting company had one field coordinator working on-site in the field office, and the technical team working off-site.

In this study, when geotechnical data need to be shown on a site base map or a cross-section, the as-built topographical data were extracted from a digital elevation model of the site that was built based on LiDAR survey. The topographic data should be regularly updated when the site features change.

The use of R programming languages for geotechnical data processing and
visualization: Techniques to manage geotechnical data for a post-closure
tailings storage facility

Y Sun & H Kalm & E Hsiao & PE Crouse

Table 1   Example of roles and responsibilities for the field and office crews

	Field Geologist	Field Coordinator	Technical Team
Data Quality Assurance	Overseeing investigation activities to verify that the work plan was implemented appropriately.	Quality assurance and review of data gathered daily during the investigation.	Performing quality assurance of data gathered throughout the investigation program.
Data Acquisition and Data Processing	Completing borehole lithology logs. Maintaining photographic logs. Collecting global positioning system (GPS) coordinates using a handheld Trimble GPS unit.	Converting borehole lithology logs to spreadsheet using Excel template. Preparing daily field reports.	Preliminary data processing. Comparing the new data against past data in nearby locations Preparing fence diagrams for cross-sections of interest.
Field Testing Program Coordination	Communicating to drilling company the depths for pore pressure dissipation (PPD) tests, field tests, and sampling.	Communicating sequencing and schedule of the program with the Stantec technical team.	Providing to field team depths for PPD tests, field tests and sampling. Adjusting testing locations and sampling depths as necessary.
Laboratory Testing Coordination	Assigning labels to the collected soil samples.	Preparing soil sample chains of custody with the technical team to select sample intervals and laboratory geotechnical testing analysis. Preparing soil samples for shipment to the off-site laboratory.	Working with the Engineer of Record (EoR) to determine investigation, sample locations and laboratory testing.

2.2   Data integration

Data integration is a process to map the geotechnical data into the 3D model. The first step is to compile the geographic data of the investigation to align them spatially with the coordinate system used in the 3D model. This step ensures that the geotechnical data can be accurately referenced within the 3D model. The second step is to identify the key identifiers that can be used to sort, filter, and differentiate the data. The examples of the commonly used identifiers include names of the facilities, types of the investigations, dates of the investigations. The third step is to convert the data to appropriate computer-readable formats. It is crucial to minimize manual data entry, which will reduce not only time and cost but also potential for human error. R supports integration with various data formats commonly used in geotechnical engineering, such as comma separated value (CSV) files, Excel (Microsoft Corporation, 2018) spreadsheets, and SQL databases. This enables seamless importing and exporting of data from various sources and tools. The following are three strategies to minimize manual work during the data integration process.

The use of R programming languages for geotechnical data processing and visualization: Techniques to manage geotechnical data for a post-closure tailings storage facility

Y Sun & H Kalm & E Hsiao & PE Crouse

The first strategy is automated data importing. This is the most preferred method for data entry as data is imported directly from the equipment. Data can be downloaded from various sources such as the laboratory test apparatus, the monitoring devices, and the CPT rigs. The data files can be reformatted in R and the useful information are captured automatically. The data reformatting is often achieved through string manipulations through R packages such as *tidyverse* (Wickham, 2018) and *stringr* (Wickham, 2022).

The second strategy is optical character recognition (OCR). This method applies to texts stored in PDFs or images in printed fonts. An example of this type of data is historic laboratory report. The *tesseract* package (Ooms, 2023) in R is a powerful tool to extract texts, and it supports over 100 languages. Internet tools are available to analyze the scanned documents or images (Ankit, 2022) and convert the texts into machine-readable data using technologies such as artificial intelligence (Jaksa and Liu, 2018).

If the above two mentioned strategies cannot be applied, we can use template-based data entry. This is the least preferred method for data entry and should be used only if handwritten texts are the sole source of information. It is recommended that standardized data entry templates are designed to match specific data requirements and to be read by a computer (Deaton, 2018). Figure 3 provides an example of the lithology log to guide the data input for the borehole information, the field description of the soil, and the field blow counts. It is noted that the example in Figure 3 is not structured as a typical statistical data frame which is made up of various columns with the labels and appended rows with the data. The reason for not using statistical data frame for the lithology log is that number of the rows under each column could vary significantly, which is not friendly for manual data entry. The objective for building a template that is consistent through the data entry process by different people. A template-based data entry is useful when migrating data from the legacy reports to the new database. In this study, the borehole data were entered manually and CPT data were downloaded from the drill rigs into spreadsheets.

	A	B	C	D	E	F	G	H	I	J
1	**BH-ID:**	**BH-TSF-2023-1**		BORING LOG					**SPT**	
2	Northing:	3907543.764		Depth (ft)		USGS/ASTM	Description		Depth (ft)	N,field
3	Easting:	1554289.806		From	To					
4	Datum:	WGS84				Classification				
5	Start Date:	44927		0	6	SW-SM	Well-graded sand w silt and gravel, olive brown (2.5Y 4/3)…		23	10
6	Finish Date:	44927		6	7.75	SP	Poorly graded sand, yellowish brown (10YR 5/6)…		41	64
7	Drilling Method:	Terra Sonic TSI 150		7.75	8	ML	Silt, light gray (10YR 7/2), moist, non-plastic, non-cohesive…		61	36
8	Driller:	Company		8	10	SP	Poorly graded sand, yellowish brown (10YR 5/6)…		76	58
9	Logged by:	C. Charles		10	14	SP-SM	Poorly graded sand w silt, very dark gray (10YR 3/1)…			
10	Reviewed by:	D. David		14	16	ML	Inorganic silt, very dark gray (10YR 3/1)…			
11	Total Boring Depth (ft):	85		16	34	SP-SM	Poorly graded sand w silt, very dark gray (10YR 3/1)…			
12	Borehole Diameter (in):	4"/6"		34	38	SP	Poorly graded sand, olive brown (2.5YR 4/4)…			
13	Water Depth (ft):	Not encountered		38	40	SP-SM	Poorly graded sand w silt, gray (10YR 5/1)…			
14				40	75	SP	Poorly graded sand, medium dense…			
15				75	85	SW	Well-graded sand w gravel…			
16										

Figure 3   Example of the lithology log

2.3   Data validation

R can support data validation by assisting with statistical analysis and data visualization. The example provided in study focus on data validation through visualizing data in one-dimensional, two-dimensional, and three-dimensional plots.

Figure 2 provides an example of data validation conducted in two rounds. The first round of data validation aims to check the locations and elevations of the investigation locations. This process compares major features (e.g., depth of the tailings, thickness of the cover material) indicated by the CPTs or boreholes and if they agree with or contradict to the topographic surfaces established. The second round of data audit provides an in-depth review of data collected by different methods, including field investigation data, laboratory test results and instrumentation monitoring data.

The use of R programming languages for geotechnical data processing and visualization: Techniques to manage geotechnical data for a post-closure tailings storage facility

Y Sun & H Kalm & E Hsiao & PE Crouse

R provides extensive capabilities for data visualization, which is particularly valuable in geotechnical data management, where visualizing data through one-dimensional (1D), two-dimensional (2D) and 3D plots can help understand patterns, trends, and relationships in the data. Section 3 illustrates the examples that utilize the R's plotting functionalities to explore the geotechnical data. The packages used in the data visualisation included *ggplot2* (Wickham, 2016), *scatter3d* (Ligges, 2003), and *plotly* (Sievert, 2020).

3 Geotechnical data visualization

3.1 1D geotechnical model

The 1D geotechnical model is to visualize the geotechnical data along the depth axis. Figure 4 provides an example of a comparison between the information gathered from a sonic drilled borehole (TSF-BH-A12) and a companion (SCPTu TSF-CPT-A08) that is 10 ft away. The plot can be customized to show desired parameters of the field and the laboratory data. The example in Figure 4 includes the field description of the soils, the field blow counts, the laboratory fines, and the composition of soil types by weight from the borehole investigation as well as the tip resistance, the normalized shear wave velocity, and the dynamic pore water pressure from the CPT investigation. The vertical axis is labelled with both the elevations and the depths for the various geotechnical properties, which allows comparison of the data at similar depths. Figure 4 also includes additional annotations to enhance the desired emphasis on important thresholds or reference values. For example, in the fourth plot from left that shows the percentage by weight for the various soil composition, a reference line of 50% is highlighted to differentiate the fine-grained soil and the coarse-grained soil.

The 1D geotechnical models can be used for data audit and validation. The 1D geotechnical model allows information to be compiled in the same plot, which simplifies the cross-refences between the geotechnical data acquired by different methods. Questions can be asked through the comparison including if a stratigraphic layer that is observed in the borehole data also exists in the CPT data, and if the laboratory data contradicts or is in line with the field data. Analysis of the 1D geotechnical model helps identify any inconsistencies during the data validation process.

The 1D models are also useful to delineate of material breaks. While the geotechnical engineers look for contact surfaces between stratigraphic layers, the 1D geotechnical model facilitates a comprehensive understanding of the data and an evaluation of the related variables at the same time.

Comparison between the borehole and CPT data is a common practice for geotechnical data analysis. Using the script streamlined the process and produced batch-generated plots more efficiently than the typical spreadsheets.

In addition, the 1D geotechnical model can be used to compare the historic and new geotechnical data to evaluate if any geotechnical parameters at close-by locations and depths have changed in time due to consolidation, aging, variation of the pore pressure conditions, etc.
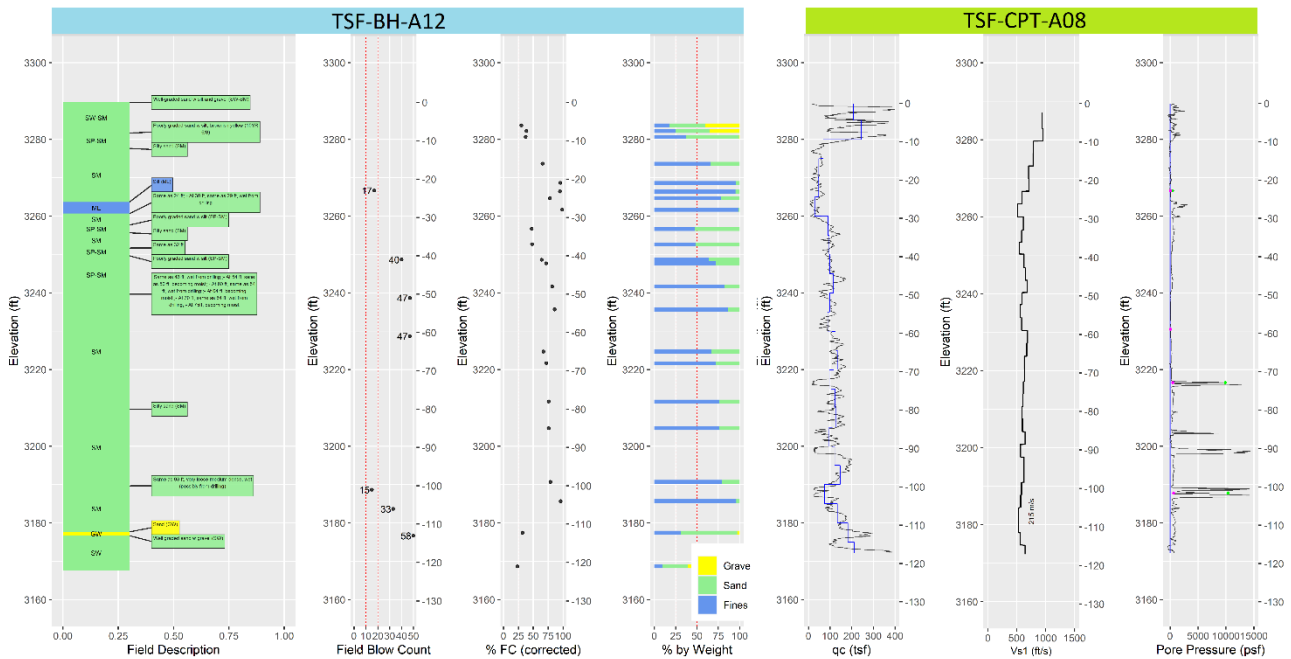
The use of R programming languages for geotechnical data processing and visualization: Techniques to manage geotechnical data for a post-closure tailings storage facility

Y Sun & H Kalm & E Hsiao & PE Crouse

Figure 4   Example of 1D geotechnical model

3.2   2D geotechnical model

The 2D geotechnical model allows for data visualization in a cross-section or a map. Figures 5 and 6 provide two examples of the 2D cross-sections of a tailings storage facility (TSF) with the information gathered from CPT. In Figures 5 and 6, two variables were projected to the cross-sections, residual shear strength ratio Su_{Liq}, and dynamic pore water pressure u_2, respectively. The foreground black lines combined with the background color bars indicate both the trend and the magnitudes of the parameters. Although only two examples are presented, the 2D geotechnical model is very versatile and can be quickly adapted to show any desired parameters. An essential step to ensure the correct location of the projected data is to interpolate the spatial coordinates along the selected cross-section to calculate the lateral distance, which is shown as the horizontal axis in Figures 5 and 6. Instead of showing multiple geotechnical parameters across the board, we noted that the 2D geotechnical model needs to focus on one or two parameters in the same plots to maintain readability. Due to the spatial complexities of the model and potential overlaps of the projected data, there is less clarity of the data if the plots become crowded. To improve understanding of the relative magnitudes, the 2D plots may need to include contour intervals and color-coded reference lines.

The 2D geotechnical models can be used to delineate material zones. The 2D geotechnical model facilitates the delineation of the material zones based on the various geotechnical parameters. In Figures 5 and 6, a total of three tailings groups were identified between the as-built topographic surface and the bottom of the tailings. Once the material zones are delineated, the 2D geotechnical model can be used to define the saturated zones based on the laboratory test data and the dynamic pore water pressure. In the u2 plot in Figures 5 and 6, the pressure head at the end of pore pressure dissipation tests are also shown to assist in evaluating the saturated zones. The geotechnical data within each material zone can be extracted for further data analysis. Figure 7 shows the cumulative distribution curves for the residual shear strength ratio calculated for the tailings sand, the interbedded tailings, and tailings slimes delineated in the cross-section. The values in Figure 7 can be used to develop the geotechnical parameters that can be used in subsequent stability analyses of this section.
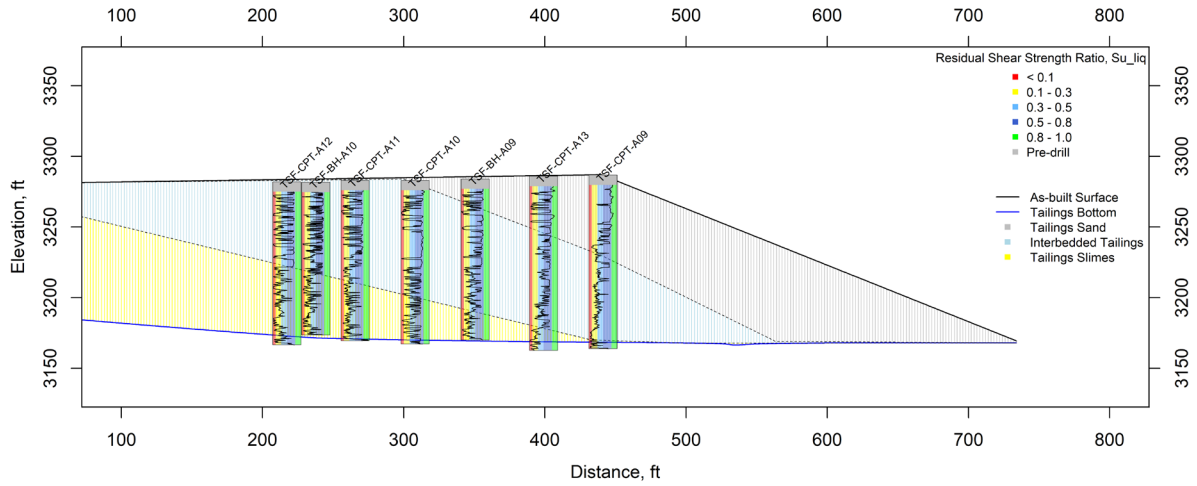
The use of R programming languages for geotechnical data processing and
visualization: Techniques to manage geotechnical data for a post-closure
tailings storage facility

Y Sun & H Kalm & E Hsiao & PE Crouse

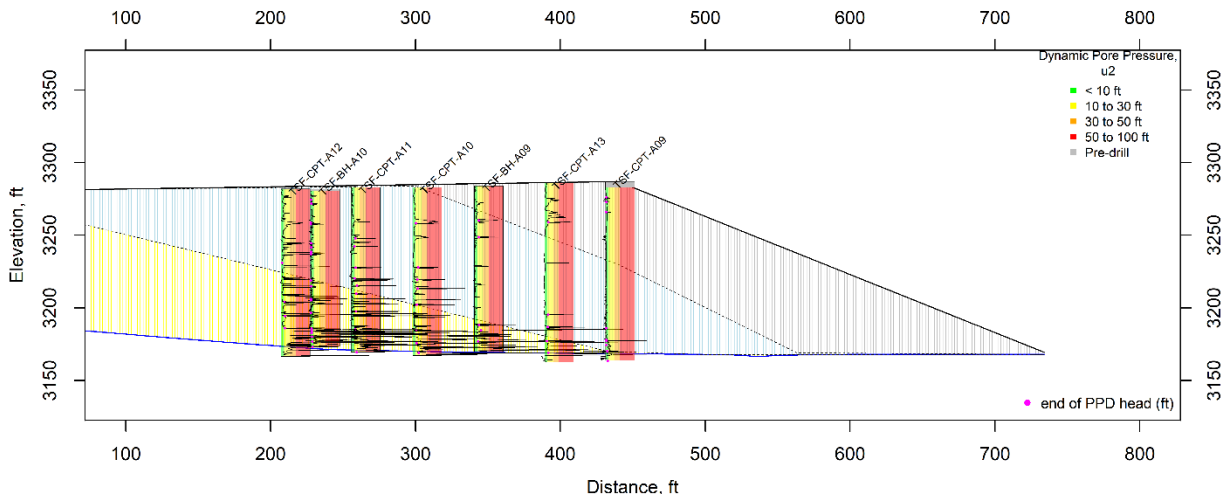**Figure 5   Example 2D geotechnical model of residual shear strength ratio**

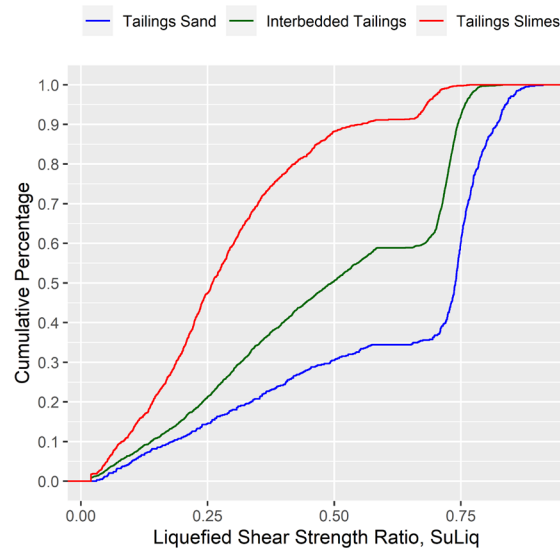**Figure 6   Example 2D geotechnical model of dynamic pore water pressure**

Figure 7   Cumulative distribution curve of Liquefied Shear Strength Ratio

The use of R programming languages for geotechnical data processing and visualization: Techniques to manage geotechnical data for a post-closure tailings storage facility

Y Sun & H Kalm & E Hsiao & PE Crouse

3.3   3D geotechnical model

The 3D geotechnical model is to visualize the spatial distribution of the data. Figures 8 and 9 provided two examples of the 3D models of the information gathered from CPT within the TSF. In Figures 8 and 9, normalized cone tip resistance Qtncs and dynamic pore water pressure u2 are shown as color columns that are associated with their northings, eastings, and elevations. The magnitudes of the geotechnical parameters are coded by the colors. It is important that spatial reference points are included in the 3D model so that the positions of the data are readable. In Figures 8 and 9, the locations of the CPT data are vertically projected to the terrain contours to represent their relative positions within the TSF. As shown in Figures 8 and 9, the Qtncs and u2 plots can be compared to examine the subsurface conditions and locate areas with the contractive and saturated tailings. The views can be rotated, zoomed in and out to provide a comprehensive view of the geotechnical data within the site context.

The 3D geotechnical models can be to optimize investigation areas. Through the course of the geotechnical investigation, information can be gathered to create the 3D models to understand the subsurface conditions. Considering the representation of different soil types, the investigation locations can be optimized. The 3D models are useful to select representee samples to cover the critical areas of the TSF where contractive and saturated tailings exist.

In addition, the 3D model can assist in the process of selecting slope cross-sections for the stability analyses to determine the factors of safety against slope failures. The 3D model assists in the review of important information for stability analysis, such as critical soil layers, geological features, and pore water pressure conditions.
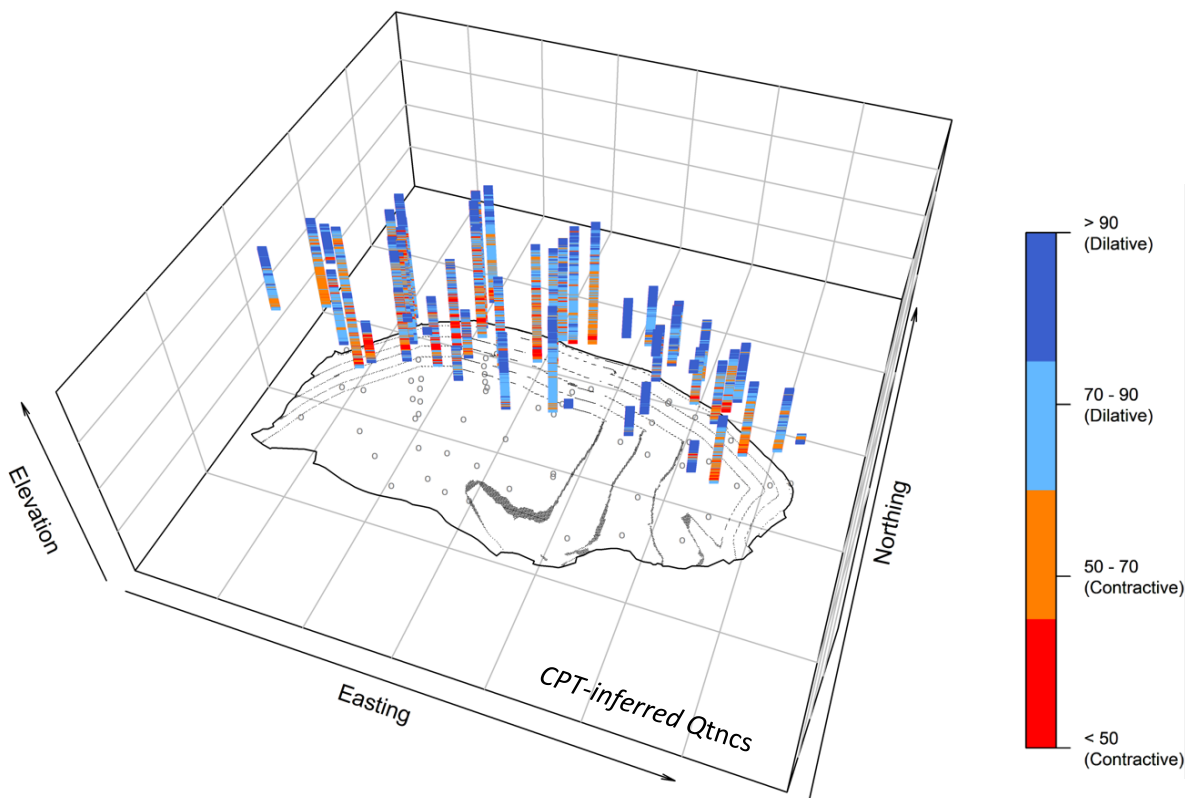

Figure 8   Example of 3D geotechnical model for CPT-inferred Clean Sand Corrected Tip Resistance
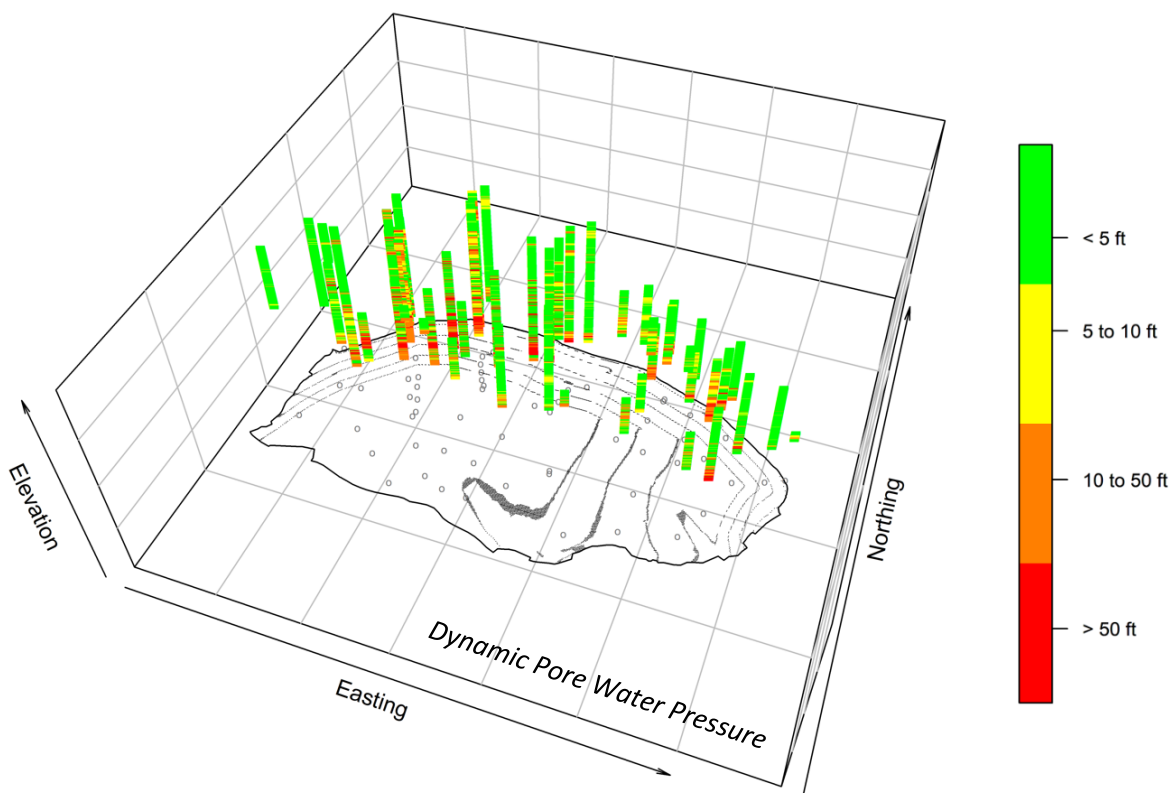
The use of R programming languages for geotechnical data processing and visualization: Techniques to manage geotechnical data for a post-closure tailings storage facility

Y Sun & H Kalm & E Hsiao & PE Crouse

Figure 9   Example of 3D geotechnical model for dynamic pore water pressure

4   Discussion and conclusion

This paper presented a workflow to integrate geotechnical data into a database and utilized R to visualize the geotechnical data in 1D, 2D, and 3D models. The examples demonstrated the efficiency of R that enables automation of the repetitive tasks involved in the data processing, and the flexibility of R that designs plots specifically tailored to incorporate desired geotechnical parameters for visualization.

The techniques discussed in this study are demonstrated to be easy to adapt, which can be utilized for not only the TSFs but also other mine domains, such as the waste stockpile. Like developing a model for the TSF, topographic surfaces of the waste stockpile can be compiled into a 3D model to show the geometry of the stockpile at various stages of the construction. Operational data can be calculated from the 3D model, such as volumes currently and planned to be deposited, as well as the characterization data of the waste. The types of data can be expanded to include other disciplines of the mine closure process, including water quality data, ecological data, and financial data.

It is noted that the complexity of the models will require basic knowledge in programming and, most importantly, the developer's technical expertise in geotechnical engineering, mining, and related engineering disciplines. The site and engineering knowledge are more important than the programming skills. To adapt to the new needs in the industry and to improve the quality of the solutions, continuous learning for geotechnical engineers in programming and data science should be encouraged by consulting companies.

References

Ankit, R. 2022, '*WebPlotDigitizer Version 4.6*', California, USA, https://automeris.io/WebPlotDigitizer, Pacifica.
Deaton, S. L. 2018, 'What are the benefits of geotechnical data interchange', *69th Highway Geology Symposium*.
Microsoft Corporation, 2018. 'Microsoft Excel', Retrieved from https://office.microsoft.com/excel

The use of R programming languages for geotechnical data processing and visualization: Techniques to manage geotechnical data for a post-closure tailings storage facility

Y Sun & H Kalm & E Hsiao & PE Crouse

Jaksa, M.; Liu, Z. 2021, 'Editorial for Special Issue Applications of Artificial Intelligence and Machine Learning in Geotechnical Engineering', *Geosciences,* vol. 11, iss. 10, pp. 399.

ICMM, 2019. Integrated Mine Closure Good Practice Guide, 2nd Edition.

Ligges, U.; Mächler, M. 2003, 'Scatterplot3d - an R Package for Visualizing Multivariate Data', *Journal of Statistical Software*, 8(11), 1–20.

Ooms, J. 2023, 'tesseract: Open Source OCR Engine', https://docs.ropensci.org/tesseract.

Player, R.S.V 2006, 'Geographic Information System (GIS) Use in Geotechnical Engineering', *American Society of Civil Engineers*, Reston, VA, USA.

R Core Team, 2021. 'R: A language and environment for statistical computing', *R Foundation for Statistical Computing*, Vienna, Austria.

Singh, A.; Noor, S.; Chitra, R.; Gupta, M 2018, 'Applications of GIS in Geotechnical Engineering: Some Case Studies', *Int. J. Sci. Eng. Sci.*, vol.2, pp. 34–38.

Sievert, C 2020, 'Interactive Web-Based Data Visualization with R, plotly, and shiny', *Chapman and Hall/CRC*, ISBN 9781138331457, https://plotly-r.com.

Wickham, H. 2016, 'ggplot2: Elegant Graphics for Data Analysis', *Springer-Verlag New York,* ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org.

Wickham, H.; Averick, M.; Bryan, J.; Chang, W.; McGowan, LD.; François, R.; Grolemund, G.; Hayes, A.; Henry, L.; Hester, J.; Kuhn, M.; Pedersen, TL.; Miller, E.; Bache, SM.; Müller, K.; Ooms, J.; Robinson, D.; Seidel, DP.; Spinu, V.; Takahashi, K.; Vaughan, D.; Wilke, C.; Woo, K.; Yutani, H. 2019, 'Welcome to the tidyverse', *Journal of Open Source Software*, 4(43), 1686.

Wickham, H. 2022, 'stringr: Simple, Consistent Wrappers for Common String Operations', https://stringr.tidyverse.org, https://github.com/tidyverse/stringr.