BIG – Branschsamverkan i grunden

Forskningsprogram för effektiv och säker grundläggning av vägar och järnvägar

Projekt A2019:17

Rock and Groundwater Surface Modelling Using Artificial Intelligence

CHALMERS UNIVERSITY OF TECHNOLOGY

KTH

TRAFIKVERKET

STATENS GEOTEKNISKA INSTITUT

LULEÅ TEKNISKA UNIVERSITET

BIG – Branschsamverkan i grunden
Forskningsprogram för effektiv och säker grundläggning av vägar och järnvägar

Rapport BIG projekt A2019:17

Rock and Groundwater Surface Modelling Using Artificial Intelligence





Framtagen inom ramen för BIG av Kungliga tekniska högskolan och Tyréns

Stockholm 2022

Rapport publicerad av BIG, Branschsamverkan i grunden
Beställning Web: www.big-geo.se

ISBN 978-91-986927-3-0
Upplaga Digital

Framtagen av Kungliga Tekniska Högskolan
Tyréns
Författare Chunling Shan, KTH and Tyréns
Abbas Abbaszadeh Shahri, Tyréns
Stefan Larsson, KTH

Förord

This report contains a study of applying artificial intelligence to bedrock and groundwater surface modelling. The project started in January 2020 and finished in December 2021 and is funded by BIG-Branschsamverkan i Grunden, BeFo-Rock Engineering Research Foundation, Tyréns and KTH.

The following project group members have made valuable contributions to the project:
Stefan Larsson: KTH project leader and expert in geotechnology
Abbas Abbaszadeh Shahir, Tyréns AI expert and Ph.D in geotechnology
Chunling Shan, Tyréns/KTH expert and Ph.D in geophysics
Emma Zäll, Tyréns project leader
Jennifer Wänseth, Tyréns expert in rock techniques
Olof Friberg, Tyréns BIM/GIS specialist
Lars Marklund, Tyréns expert in hydrogeology
Beatriz Machado, Tyréns expert in rock mechanics
Maria Duvaldt, Tyréns expert in geotechnology
Ida Samuelsson, Tyréns/KTH Ph.D in geotechnology

The following reference group members have given valuable comments and suggestions:
Per Tengborg, BeFo
Robert Sturk, Skanska
Torleif Dahlin, LTU
Olle Båtelsson, Trafikverket
Mats Svensson, Tyréns
Rikard Gothäll, Tyréns
Alireza Malehmir, Uppsala University
Diego Mas Ivars, SKB
Paul Evins, WSP
Fardad Maghsoudi Moud, Twente University, Ntherlands

Sammanfattning

Att kartlägga bergtopografi, grundvattennivåer och överlagrade sediment är en viktig och svår uppgift i geotekniska tillämpningar på grund av komplexa stora variationer och glesa data. Modeller utgör ett verktyg för geo-ingenjörer i sitt dagliga arbete och är ett värdefullt underlag för diskussioner mellan olika yrkesroller och beslut. Resultatet av prediktiva modeller innehåller dock givetvis osäkerheter. Därför behövs formella ramverk för kvantifiering av osäkerheter (UQ, uncertainty quantification) för att kunna bedöma modellernas tillförlitlighet och minska tveksamheter i både beräkningar och tillämpningar. Moderna beräkningstekniker som artificiell intelligens-baserade modeller (AI, artificial intelligence) är lämpliga alternativ för att övervinna bristerna med idag använda metoder. Syftet med denna studie är att undersöka genomförbarheten av AI för att kunna prediktera rumslig 3D-fördelning av berggrundens topografi och grundvattennivåer i stora områden i Stockholm. Denna studie var också motiverad att ta itu med utmaningen med kvantifiering osäkerheter i geotekniska projekt genom att använda en adekvat AI-modell. Två datauppsättningar från Tvärförbindelsen och E20 Bälinge-Vårgårda användes för prediktion av berggrundtopografi och grundvattenytor och kvantifiering av osäkerheterna. Från jämförelsen av de predikterade ytorna med AI och den geostatistisk metoden Ordinary Kriging (OK), fann man att AI-metoden kunde förutsäga mer korrekta rumsliga 3D-ytor och ge mer korrekta förutsägelser vid vilken punkt som helst. Tre metoder för AI-analys och modellering testades i denna studie och de jämfördes med OK-metoden: Monte Carlo Dropout (MCD), Quantile Regression (QR) och Random Deactivating Connective Weights (ARDCW). Dessa AI-osäkerhetsmetoder såsom MCD och QR används inte mycket i bergtopografi- och grundvattenmodellering och ARDCW är en ny föreslagen ny metod i denna studie. Det påvisas att MCD-metoden förbrukar mer beräkningstid och de uppskattade osäkerheterna är inte lika exakta som QR och ARDCW. QR- och ARDCW-metoderna har visat båda överlägsenhet i uppskattningen av osäkerheter från datamängderna från de två studerade områdena. Därför har resultaten visat fler möjligheter för QR och ARDCW för framtida analyser av osäkerheter i bergtopografi- och grundvattenmodellering.

Summary

Delineating and mapping bedrock, groundwater levels and overlaid deposits is a vital task in geo-engineering applications, yet difficult due to complex spatial patterns and sparse data. As the outcome of predictive geo-engineering models contains uncertainties, formal frameworks are needed for the uncertainty quantification (UQ), which assesses the reliability of the models and reduces hesitancy in both computations and real-world applications. Modern computing techniques such as artificial intelligence-based models (AI) are an appropriate alternative that can overcome the deficiencies of current methods. The objective of this study is therefore to investigate the feasibility of AI in the prediction of the 3D spatial distribution of subsurface bedrock and groundwater levels in large areas in Stockholm, Sweden. This study was also motivated to address the uncertainty quantification challenge in geo-engineering projects by using adequately accurate spatial AI models. Two datasets from Tvärförbindelsen and E20 Bälinge-Vårgårda were used for the prediction of bedrock and groundwater surfaces and uncertainty quantifications. By comparing the surfaces predicted by AI and the Ordinary Kriging (OK) geostatistical method, it was found that the AI method could more accurately predict the spatial 3D surface and provide more appropriate predictions at any point in the subsurface. Three AI uncertainty analysis methods, Monte Carlo Dropout (MCD), Quantile Regression (QR) and Random Deactivating Connective Weights (ARDCW), were tested in this study and then compared with the OK method. These AI uncertainty methods such as MCD and QR are not frequently used in bedrock and groundwater surface modelling, and ARDCW is a state-of-the-art ensemble method newly proposed in this study. It showed that AI uncertainty quantification methods can quantify the uncertainties more accurately and contain truer values inside the intervals than the OK method. It proves that the MCD method consumes more computational time and that the estimated uncertainties are not as accurate as QR and ARDCW. The QR and ARDCW methods have both shown superiorities in the estimation of uncertainties from the two tested datasets. Therefore, the results have demonstrated more possibilities for the use of QR and ARDCW in future uncertainty analyses of bedrock and groundwater modelling. The results also show that the AI method is a flexible and efficient alternative approach for DTB and GWT surface mappings and can account for associated uncertainties.

Innehåll

FÖRORD I

SAMMANFATTNING II

SUMMARY III

INNEHÅLL V

LIST OF PUBLICATIONS: VI

1 INTRODUCTION 1

1.1 Background of bedrock surface modelling 1
1.2 Background of groundwater surface modelling 1
1.3 AI modelling 2
1.4 Benefits of accurate modelling 2
1.5 Aims 2

2 DATA 3

2.1 Tvärförbindelsen 3
2.2 E20 Bälinge-Vårgårda 3

3 METHODOLOGY 4

3.1 Deep Learning 4
3.2 Uncertainty analysis 4
3.2.1 Monte Carlo dropout 6
3.2.2 Quantile Regression 6
3.2.3 ARDCW 7

4 RESULTS 7

4.1 Bedrock surface modelling and uncertainty quantification 7
4.2 GWT modelling and uncertainty quantification 10

5 IMPLEMENTATION OF THE WORK 12

6 CONCLUSION 12

7 REFERENSER 13

BILAGA A CONFERENCE ABSTRACTS 16

A.1 Artificial intelligence-based models to predict the spatial bedrock levels for geoengineering application (extended abstract published) 16
A.2 Uncertainty analysis of an optimum predictive neural network model in subsurface bedrock level modelling (extended abstract published) 18

BILAGA B FULL PAPERS 21

B.1 Spatial distribution modelling of subsurface bedrock using a developed automated intelligence deep learning procedure: A case study in Sweden (article published) 21
B.2 A new approach to uncertainty analysis using automated predictive deep learning in groundwater (accepted by Natural Resources Research) 33
B.3 Visualisering av bergtopografi med artificiell intelligens (published) 69
B.4 3D modelling and uncertainty analysis of subsurface bedrock using a hybrid automated deep learning approach- A case study in Sweden (first draft) 73

List of publications:

Conference proceedings:

1. Shan C, Shahri AA, Larson S (2020). Artificial intelligence-based models to predict the spatial bedrock levels for geoengineering application. *In Proc. 3RD Conference of the Arabian Journal of Geosciences (CAJG), Springer Nature.*
2. Shan C, Shahri AA, Larson S, Zäll E (2021). Uncertainty analysis of an optimum predictive neural network model in subsurface bedrock level modeling. *In proc. 3rd International Symposium on Machine Learning and Big Data in Geoscience, Machine Learning & Risk Assessment in Geoengineering, MLRA 2021, 48-52, Wroclaw, Poland.*

Full papers:

1. Abbaszadeh Shahri A, Shan C, Zäll E, Larsson S (2021). Spatial distribution modeling of subsurface bedrock using a developed automated intelligence deep learning procedure: A case study in Sweden (2021). *Journal of Rock Mechanics and Geotechnical Engineering, 13(6):1300-1310, https://doi.org/10.1016/j.jrmge.2021.07.006.*
2. Abbaszadeh Shahri A, Shan C, Larsson S (2022). A new approach to uncertainty analysis using automated predictive deep learning in groundwater. *In Press, Natural Resources Research, Springer.*
3. Abbaszadeh Shahri A, Shan C, Larsson S (2022). Visualisering av bergtopografi med artificiell intelligens. *Bygg & Teknik, 114 (1), 44–46.*
4. Abbaszadeh Shahri A, Shan C, Larsson S (2022). 3D modelling and uncertainty analysis of subsurface bedrock using a hybrid automated deep learning approach - A case study in Sweden. *First draft.*

1 Introduction

1.1 Background of bedrock surface modelling

Understanding the subsurface of our earth is important for many different applications within the geosciences. The creation of three-dimensional models can assist in understanding a conceptual and quantitative model of the subsurface. The level of bedrock is valuable information for both the design and construction phase of many projects. Unexpected conditions regarding the bedrock level can increase risks, cause delays and affect cost estimation. Bedrock level information is usually obtained from sparse measurements such as borehole drilling, or continuous 2D profiles such as geophysical measurement. In Sweden, soil-rock soundings are usually carried out to determine the bedrock level. It is an accurate method to determine the exact level of the bedrock, but one major disadvantage is the high cost of drilling. Furthermore, it is a time-consuming process and only measures the bedrock level at sparse points. Geophysical methods are cheaper to apply and can usually cover a larger part of the study area; however, the bedrock levels interpreted from geophysical profiles are not exact and contain uncertainties (e.g., Fournier et al., 2013). They need to be correlated with information from direct geotechnical methods, as data are generally interpreted qualitatively, and useful results can only be obtained by experts familiar with the particular testing method (Shahri et al., 2021). With limited bedrock information, the idea of having an accurate prediction of the rock surface with the same number of boreholes is a valuable and attractive one (Machado, 2019). To estimate the depth of the bedrock at unsampled locations, interpolation and extrapolation are often needed. This is usually performed using several mathematical methods. Kriging interpolation is one of the most commonly used interpolation methods in the fields of geology and hydrology (Samui and Sitharam, 2011; Viswanathan et al., 2014; Kitterød, 2017; Li et al., 2020), and it is generally considered to produce good fitting surfaces for a variety of datasets (Kitanidis, 1997). Other examples of commonly used interpolation methods are inverse distance weighting and minimum curvature (Li & Heap, 2014). As each interpolation method has its own characteristics and level of complexity, results produced with different interpolation methods for the same data can differ. In literature, interpolations are commonly evaluated according to how accurately the modelled dataset fits to the measured dataset (Li & Heap, 2014). However, uncertainties related to model parameters, underlying conceptual assumptions, model structure, geological conditions and observed spatial data (i.e., the lack of data on natural variability) cannot be quantified by these methods.

1.2 Background of groundwater surface modelling

Groundwater is defined as the water that completely fills the pore spaces in soil and fractures in rock (Todd & Mays, 2005; Hu and Jiao, 2010; Tang et al., 2017; Salvo et al., 2020). Of all the freshwater on earth, around 30% is available as groundwater (Nordström, 2005). Groundwater is an important source of drinking water in many parts of the world. Furthermore, the occurrence of groundwater affects the geotechnical conditions of soil; for example, lowering the groundwater table (GWT) may cause subsidence (SGU, n.d.). Over the past several decades, the use of groundwater modelling has increased to better evaluate the complexities inherent in hydrogeological calculations. Information required for groundwater modelling concerns, for example, the elevation of soil and bedrock layers and the groundwater level (Agerberg, 2020). This data is usually collected at a limited number of sample points, for both practical and economic reasons. To approximate values at unknown points, known values at measured sample points can be interpolated over the study area (Kitanidis, 1997). This is usually performed by using mathematical methods. These methods can be Kriging, inverse distance weighting and minimum curvature, as mentioned in the section above. In this report, only the ground surface level and groundwater level are considered for groundwater modelling.

1.3 AI modelling

In recent years, artificial intelligence (AI) techniques have shown remarkable computational and learning capabilities in addressing geotechnical problems. As depth to bedrock (DTB) and groundwater table (GWT) modelling deals with various uncertainties (Shahri et al., 2020; Hood et al., 2019), subcategories of AI such as machine learning, neural networks, evolutionary computation, and deep learning techniques are appropriate alternatives to overcome the limitation and simplifications (e.g., Hegle et al., 2017; Chang and Chao, 2009). Furthermore, hybridising the AI techniques with metaheuristic algorithms can significantly optimise model performance. Currently, AI is used for systematic uncertainty analysis that can be effectively applied to generate DBT and GWT models. It can account for uncertainties related to model parameters, underlying conceptual assumptions, model structure, geological conditions and observed spatial data (Appendix B.2). For more about AI, see Marsland (2018).

1.4 Benefits of accurate modelling

Depth to bedrock (DTB), as the thickness of the sediments above the bedrock, plays an important role in many different contexts. One example is in the mining industry where DTB information can help in cost estimation for underground mining. Another example concerns infrastructure building projects where DTB information can help in subsurface geo-engineering modelling and risk assessment (Kitterod and Leblois, 2019; Shahri et al., 2020). These examples imply that producing highly accurate DTB predictive models is a critical task that can have significant effects on the costs and risks of geo-engineering projects. An accurate modelling of the bedrock levels at unsampled locations can also save the cost of extra borehole drillings, which usually is very expensive to carry out. In recent years, the uncertainty of estimated bedrock levels has caught the attention of many engineers (e.g., Trafikverket) due to the risk of failures in infrastructure projects. The inaccurate bedrock level predictions can cause big financial problems when worse case scenarios are unplanned and occur in reality (Carlsson, 2005; Koc et al., 2020). The advantage of AI modelling is that it is a flexible and efficient alternative approach that can account for associated uncertainties, thus creating more accurate spatial 3D models and providing an appropriate prediction at any point in the subsurface.

With increasing human and climate pressures on groundwater resources, accurate and reliable predictions of groundwater flow and groundwater levels are essential for sustainable groundwater management practices. GWT modelling is often required to simulate the impacts of management scenarios on groundwater levels. It can help evaluate the costs, benefits and risks of a proposed water resources development plan, and it has been proven useful for a variety of groundwater related engineering problems and supported decision making. An accurate GWT modelling can help, for example, in slope stability analysis, geoengineer's planning of additional safety elements when the groundwater level changes and the prediction of groundwater depletion.

1.5 Aims

- Automatically analyse and model geotechnical sounding data, geophysical data and groundwater level data to study the spatial distribution of bedrock and groundwater surfaces with artificial intelligence
- Estimate the uncertainties of estimated groundwater and bedrock levels
- Compare the AI modelling method with the traditional interpolation method such as the geostatistical method Ordinary Kriging

2 Data

2.1 Tvärförbindelsen

For the groundwater surface modelling, the data was collected from project Tvärförbindelsen. The study area is 20 km of an ongoing highway project in Stockholm, Sweden and contains more than 300 investigated groundwater wells. Among the monitored GWT data, 244 points were available and selected in two-week intervals between 19 to 25 September 2020. The focused area for groundwater surface modelling is 5 km x 5 km, as shown in the digital elevation map (Figure 1A) where we modelled the groundwater surface from the input data as shown in Figure 1B (black dots). Figure 1C shows the distribution of measured geotechnical borehole data, which is used for bedrock surface modelling. The spatially distributed GWT are from short time interval in space. Therefore, it is assumed that the acquired GWTs are stable and not influenced by rainfalls. This implies that the report does not focus on any GWT flow and aquifer conditions as are beyond the scope of project. Accordingly, as the GWT is unconfined, therefore, there is no-flux and does not require any differential equation as the advective and diffusive fluxes must exactly balance.

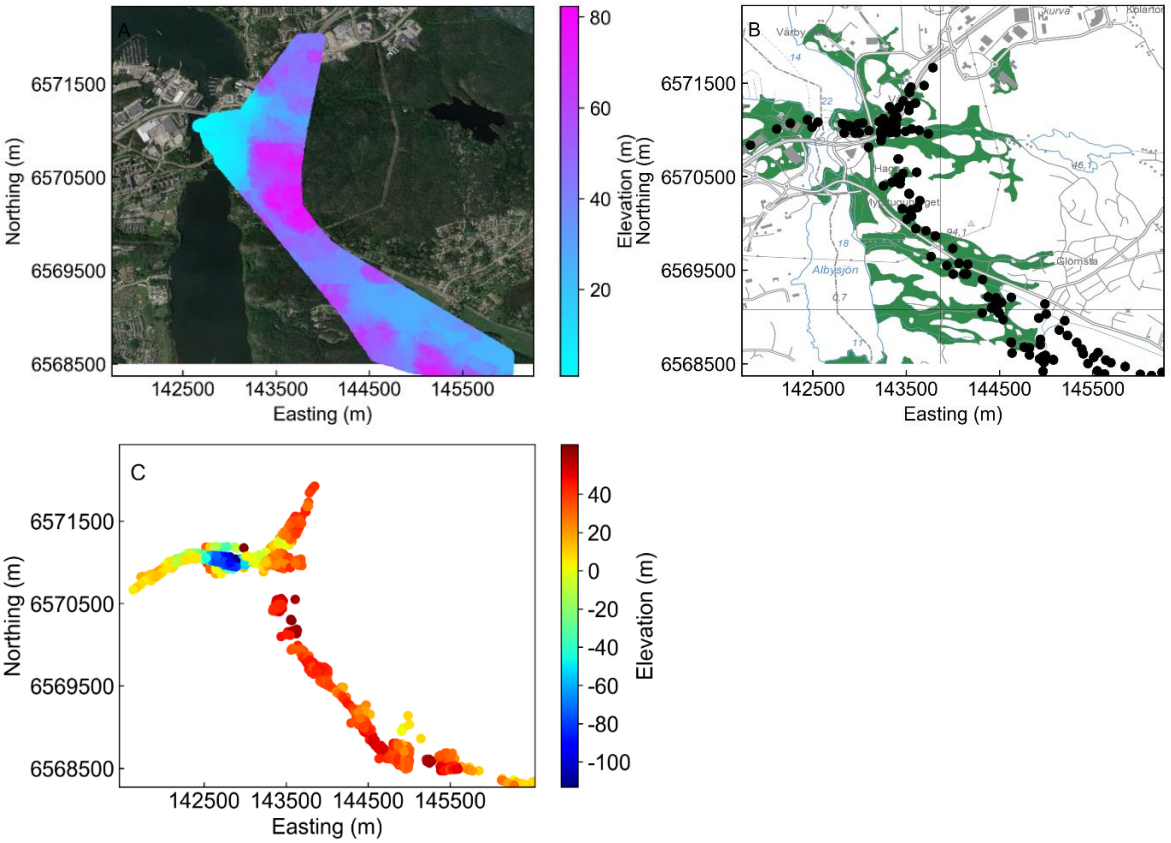


Figure 1. The overlaid digital elevation map and satellite image of the study area (A), the mapped aquifers from SGU (green) and monitored GWT data in the area (B) and the colour map of the spatial distribution of drilled geotechnical boreholes showing the depth information of the bedrock level (C).

2.2 E20 Bälinge-Vårgårda

Trafikverket has expanded E20 between Bälinge-Vårgårda to build a new highway with crossings for increased traffic load and safety concerns. In this project, Tyréns has drilled more than 200 soil-rock sounding boreholes, mapped the bedrock outcrops and carried out geophysical measurements to study the bedrock levels in the area. See Figure 2 for the

distribution of borehole locations and geophysical profiles. All these data were used as the input data for deep learning modelling. In part of the areas, the overlaying soil was excavated and the bedrock surface totally uncovered. This uncovered bedrock surface was later scanned by airborne laser scanning to acquire the point cloud for the true bedrock surface. This allowed us to compare our results from deep learning modelling with the true scanned bedrock surface. There are around 45667 points representing the uncovered bedrock surface.

3 Methodology

3.1 Deep Learning

Deep learning is part of the broader family of machine learning methods based on artificial neural networks. It attempts to mimic the human brain and enables systems to cluster data and make predictions with high accuracy (Charniak, 2019). In this study, the prediction of deep learning is used through all the studies to predict bedrock levels and groundwater levels at unsampled points. Figure 3 shows the structure of deep learning networks we used in this study. The input layers contain 3 variables: X (northing), Y (Easting), Z (Ground surface level). We developed an automated process with the ability to check numerous structures subjected to different numbers of neurons, model arrangements and internal hyper parameters. Monitored results showed that a four-layer deep learning topology, including 28 and 12 neurons managed into hidden layers 1 and 2, gives the best performance amongst captured models and thus can optimally be used to predict bedrock levels at unsampled locations (Figure 4). The output layer gives the results of the predicted bedrock level. A more complete introduction to deep learning can be found in this book: Introduction to Deep Learning (Charniak, 2019).

3.2 Uncertainty analysis

Since we lack sureness about the predicted value (interpolated levels), there is always uncertainty imbedded in the predictions. These uncertainties can be caused by a lack of bedrock information and knowledge, or errors in the measured data and mathematical modelling process. Estimating the uncertainties of predicted bedrock levels is very important for contemporary infrastructure building projects due to the impact of uncertain bedrock levels. A good estimation of bedrock level uncertainty can help in project risk management, construction material planning, better project control and fewer conflicts among the various parties. In this study we tested three uncertainty analysis methods in deep learning to quantify the uncertainties in the predicted values. Figure 5 shows the illustration of the prediction interval.

Uncertainty for predictive models refer to the variability in estimations due to plausible alternative input values. Therefore, both inherent noises in the estimates of the regression parameters and the new measurement are included in uncertainty. The prediction intervals (PI) provide an interval for the value of a single new measurement from the process. In this report, we used PI instead of confidence interval, because confidence interval is usually used for input parameters/measured data and prediction interval is used for a new measurement or a random variable. Therefore, using appropriate inputs the probability distributions propagates through the model to form a probability distribution for predictive model.

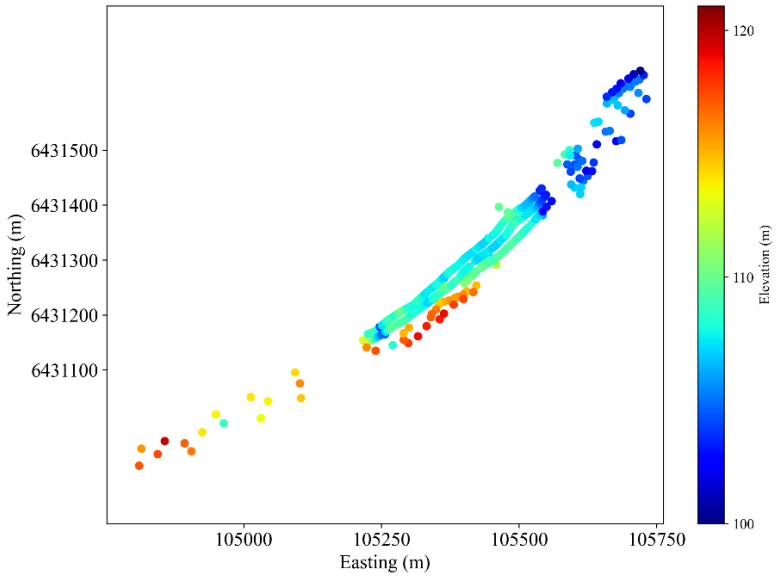


Figure 2. The spatial distribution of drilled geotechnical boreholes and geophysical profiles. Different colours indicate the depth information of the bedrock at each borehole.

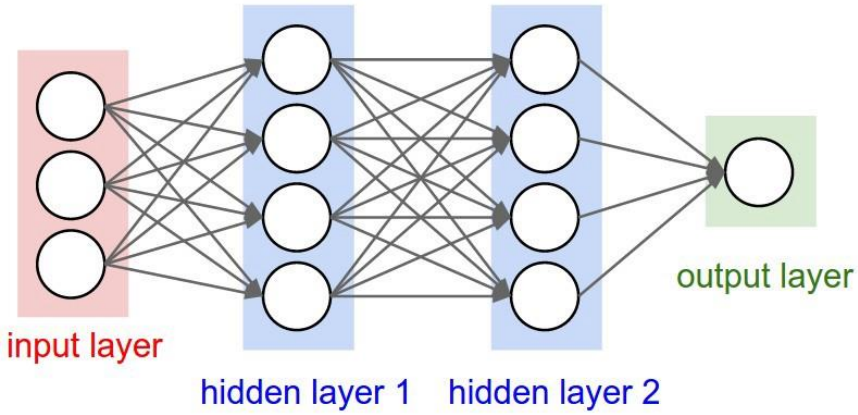


Figure 3: Deep learning neural networks with two hidden layers. Source: https://towardsdatascience.com/coding-neural-network-forward-propagation-and-backpropagtion-ccf8cf369f76

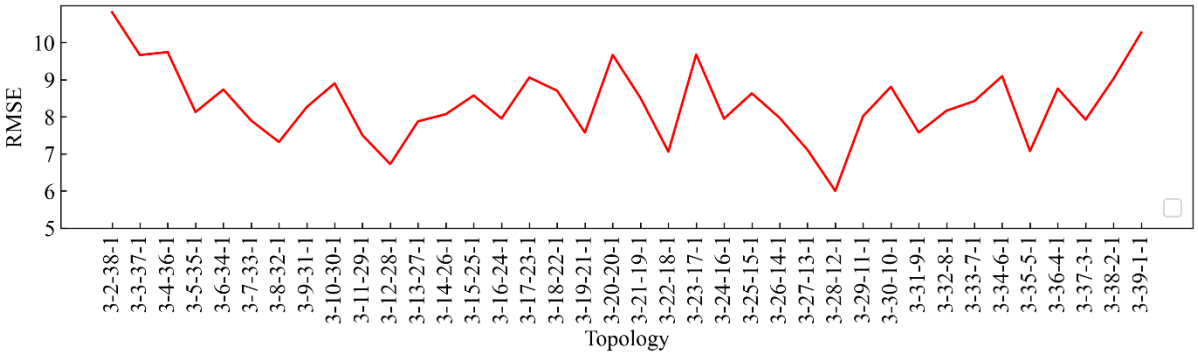


Figure 4. Variation of network RMSE for developed optimum models as a function of network structures.

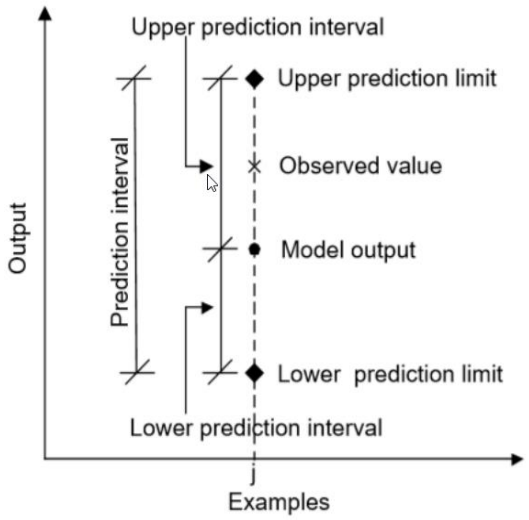


Figure 5. Definition of prediction interval and upper and lower prediction limits (Durga and Dimitri, 2006).

3.2.1 Monte Carlo dropout

Monte Carlo dropout (MCD) randomly disables neurons during training. This implies that in each training step a different subset of the network architecture is evaluated and adjusted. This kind of randomness is also introduced to the prediction process. We trained multiple networks to get a distribution of the output value and use this distribution to estimate the uncertainties. Figure 6 shows the illustration of MCD. The number of disabled neurons is a user-define value and set at 20% in this study. To learn more about MCD, see Seoh (2019).

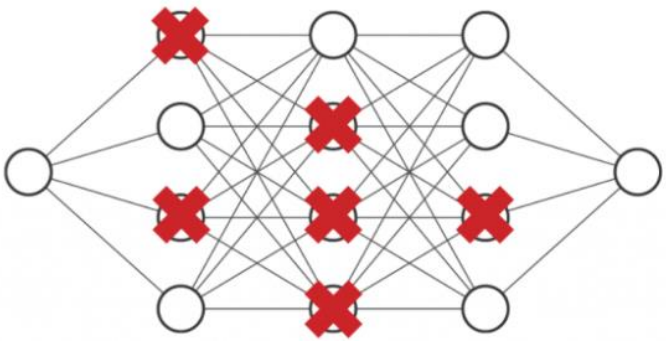


Figure 6. Illustration of MCD. It randomly disables networks during training and prediction time. Source: https://www.inovex.de/de/blog/uncertainty-quantification-deep-learning/

3.2.2 Quantile Regression

Quantile regression (QR) allows for an understanding of the relationships between variables outside the mean of the data, making it useful in understanding outcomes that are non-normally distributed and have nonlinear relationships with predicting variables. Figure 7 shows an example of different estimated quantiles for a simulated dataset. Between Q (0.05) and Q (0.95) lies the 90% prediction interval. For more reading on QR, see Koenker and Hallock (2001).

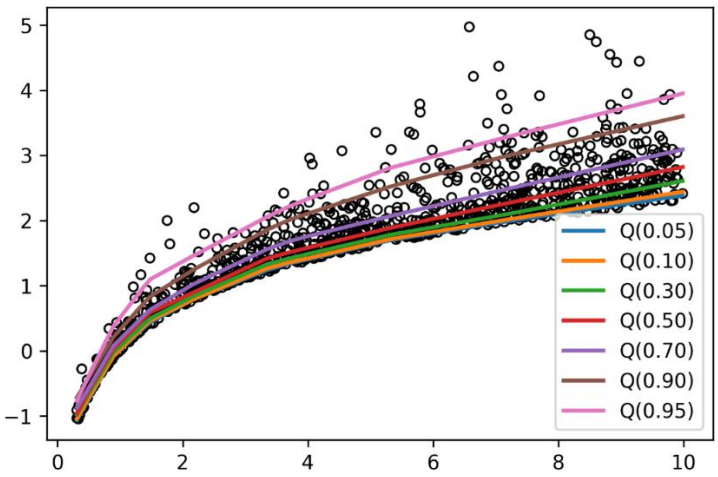


Figure 7. An example showing the different estimated quantiles. Between Q (0.05) and Q (0.95) lies the 90% prediction interval. Source: https://hackmd.io/@cgarciae/quantile-regression.

3.2.3 ARDCW

The acronym ARDCW stands for random deactivating connective weights approach (see Appendix B.2). This is a novel state-of-the-art ensemble-based deep learning method that was proposed and developed during this study. Unlike the MCD method, instead of dropping neurons, the connective weights between layers are randomly deactivated. The deactivation runs several times without changing the structure of the network, and for each run the deactivation of weights starts from the pre-found optimum model. The number of deactivated weights is a user-define value, and 15% was used in this study. More theories and information about this method was written in our manuscript in Bilaga B (B.2).

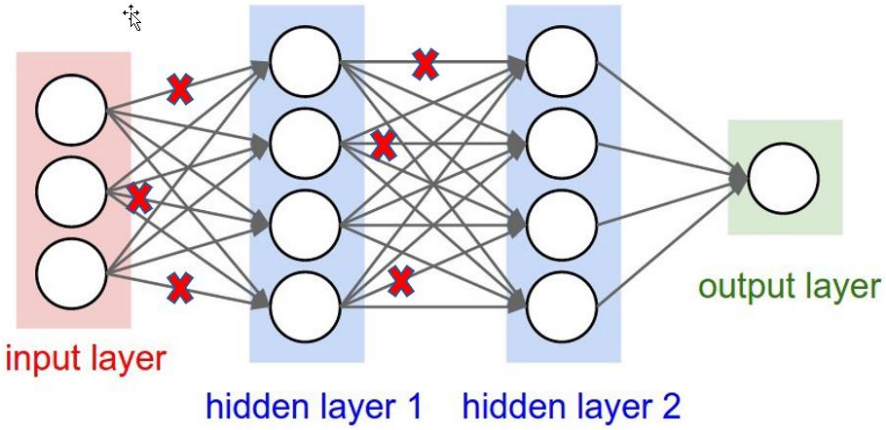


Figure 8: Illustration of ARDCW with two hidden layers. Modified from source: https://towardsdatascience.com/coding-neural-network-forward-propagation-and-backpropagtion-ccf8cf369f76

4 Results

4.1 Bedrock surface modelling and uncertainty quantification

After the optimum model was founded, as explained in the Deep Learning section, the model was applied to both the Tvärförbindelsen and E20 datasets to predict bedrock levels at unsampled locations. For the Tvärbindelsen dataset, a validation dataset was split from the whole dataset to evaluate the performance of the optimum model. Since the E20 Bällinge-Vårgårda dataset has the true bedrock surface scanned, it possible to compare the predicted

and true bedrock levels at exactly the same locations, thus quantifying the accuracies of the optimum model with the scanned true surface. Figure 8 shows the fitting between the measured and predicted bedrock levels of the validation dataset from Tvärförbindelsen. It can be observed that Ordinary Kriging (OK) has a significantly lower fitting degree than the AI modelling. The comparison was done with OK, because similar to AI, it is an inherently statistical technique using a weighted procedure and can work with non-normal data distribution aiming to reduce the error of prediction. Moreover, in the prediction process both OK and AI generate uncertainties. To check the applicability of OK for big data, different variogram functions integrated with different corresponding parameters such as sill, range, nugget and number of lags were examined. The result revealed that OK gave different interpolation results and through this tunning process the OK with highest fitting between measured and predicted data for the test dataset was selected.

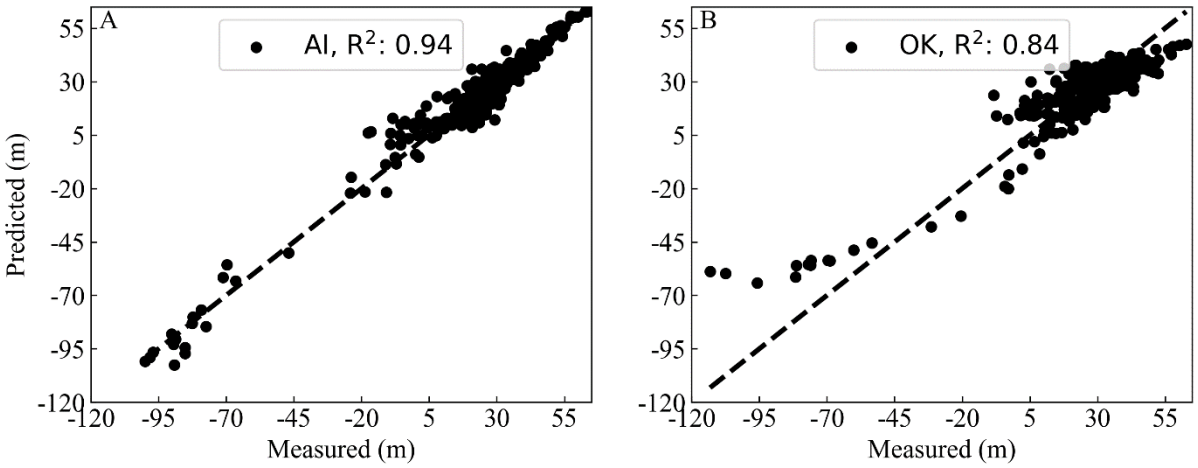


Figure 8. The fitting between measured and predicted values and a comparison of the predictability of AI (A) and OK (B) modelling concerning a validation dataset from Tvärföbindelsen.

Figure 9 shows a comparison between the true scanned bedrock surface and the predicted surface in 3D from the E20 Bällinge-Vårgårda dataset. The red points show the input data used by deep learning to predict the surface. The differences between the true and predicted values are shown in Figure 10. Since Kriging interpolation is one of the most commonly used interpolation methods in the fields of geology and hydrology, we compared the OK method with QR and ARDCW to see the performances of different methods. Figure 10 shows the difference (true – predicted values) in the comparison between different uncertainty analysis methods; for example, we see that OK has higher differences at more places than QR and ARDCW. The fitting between measured (true scanned bedrock) and predicted values (Figure 11) shows that QR and ARDCW have higher fitting degrees than OK. Table 1 shows the summary of the differences between predicted and scanned true values for 45667 points. The Total Absolute Difference is the sum of the absolute difference of each point, and the total absolute difference of QR and ARDCW are smaller than OK. The Average Difference/point is the total absolute difference divided by the total number of points (45667). The QR has the lowest value and OK has the highest value. The improvement from the OK column shows how much better the predicted bedrock surfaces of QR and ARDCW are for the tested dataset compared to OK.

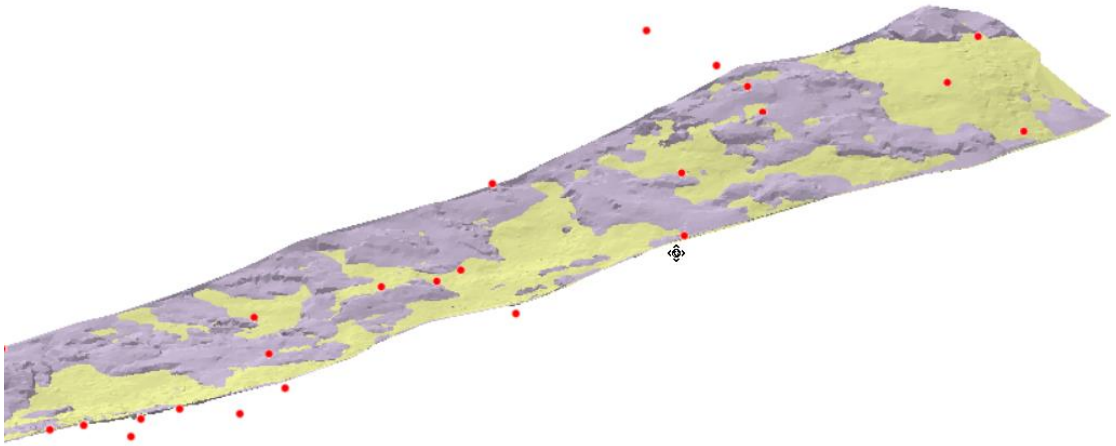


Figure 9. A 3D view of the scanned true bedrock surface (grey) and the predicted bedrock surface (yellow) from the measured data (red dots) in dataset E20.

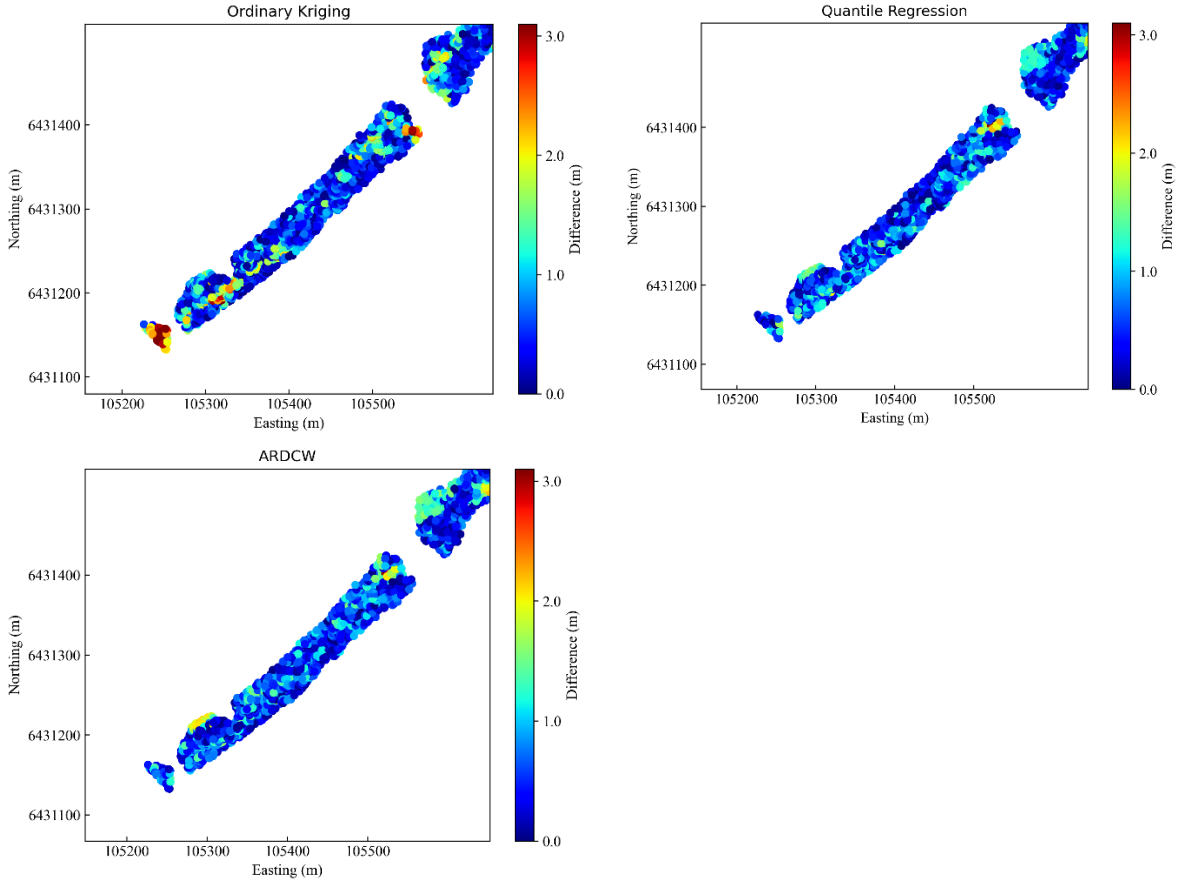


Figure 10: The difference plot between the predicted/estimated bedrock levels and the true bedrock levels.

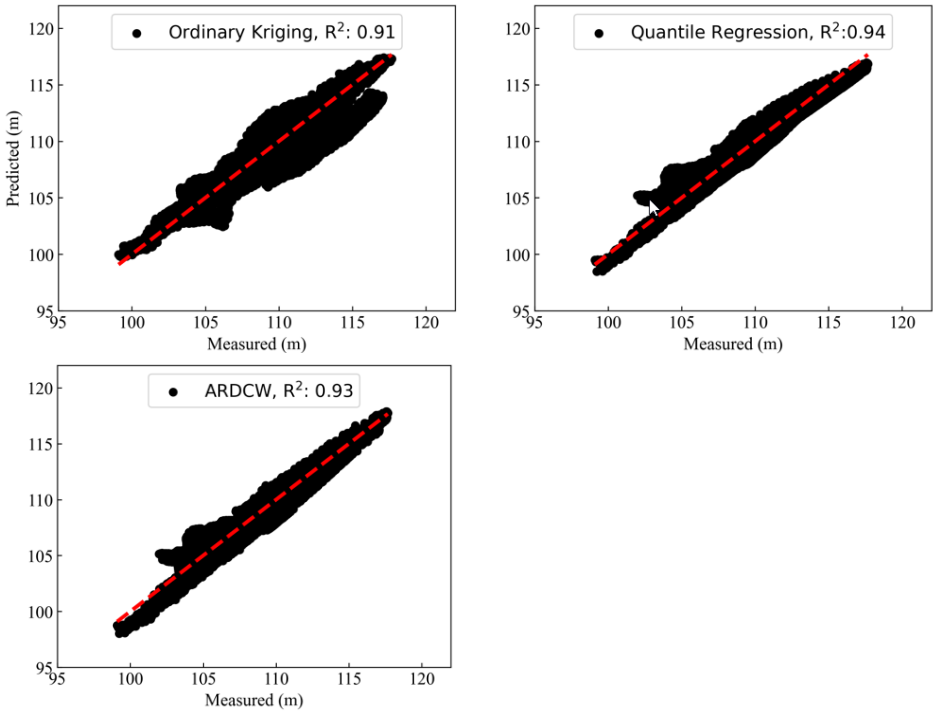


Figure 11. The fitting between measured and predicted values and a comparison of the predictability of AI and OK modelling on true scanned bedrock data from E20.

Table 1: Summary of differences between the predicted and true bedrock levels

Method	Total absolute difference (m)	Average difference/point (m)	Improvement from OK (%)
Ordinary Kriging	31470	0.69	0%
Quantile Regression	27787	0.61	12%
ARDCW	30907	0.67	3%

Figure 12 shows the histogram of uncertainty intervals estimated by different methods. For OK the intervals are between 0 to 3 meters with 73% of true values falling inside the interval (see Figure 5 for illustration). QR has slightly higher intervals, between 1.1 to 4.5 meters with 77% of true values falling inside the intervals, and ARDCW has uncertainty intervals between 0.5 to 4.5 meters with 77% of true values falling inside the intervals. The OK uncertainties are estimated from the OK variance. Many geostatisticians consider the OK variance an unreliable representation of the spatial interpolation error (Heuvelink and Pebesma, 2002). Goovaerts (1997) has pointed out that OK variance is independent of data values, and the errors are not dependent on the vertical variations of Z (bedrock level). This may be the reason why OK has smaller intervals than the other AI methods. The uncertainty intervals from deep learning such as QR and ARDCW start from non-zero values. This is mainly due to the regression modelling process where the predicted values are not usually the same as the input (observed) data. The size of the starting interval is dependent on the input dataset (e.g., observational errors) and the modelling method. We have observed zero starting intervals on other tested datasets.

4.2 GWT modelling and uncertainty quantification

The same technique used for bedrock surface modelling was applied for GWT modelling. In the dataset from Tvärförbindelsen, there is no true GWT that we could use to compare with the predicted groundwater surface. Therefore, the performance studying of deep learning modelling is based on the validation dataset we split from the available input data. The

validation data acts as unseen data for the model and provides an unbiased evaluation of the obtained model.
Figure 13 shows the 95% prediction intervals of the three methods together with the observed GWT for the validation dataset. The X-axis represents the number of validation data splits from the whole dataset used for the performance analysis of different uncertainty methods. Table 2 shows the compared different statistical criteria and the ranking of criteria. E_c is the coefficient of efficiency (Nash and Sutcliffe, 1970), *RMSE* means the Root Mean Square Error, *IA* is the index of agreement (Willmott, 1984), R^2 is the square differences between observed and predicted data, and $P_{PI}^{95\%}$ is the averaged 95% prediction interval. As seen in the Total Rank, the ARDCW method has the highest rank score.

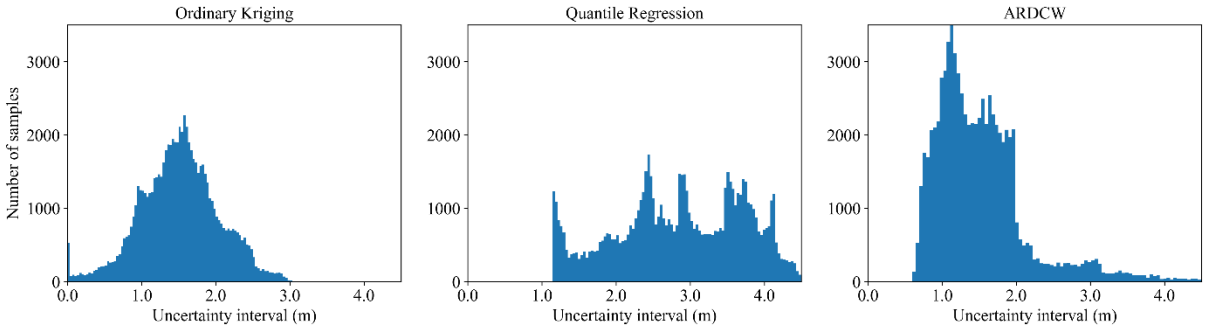


Figure 12: Uncertainty intervals, 85% prediction intervals for QR and ARDCW.

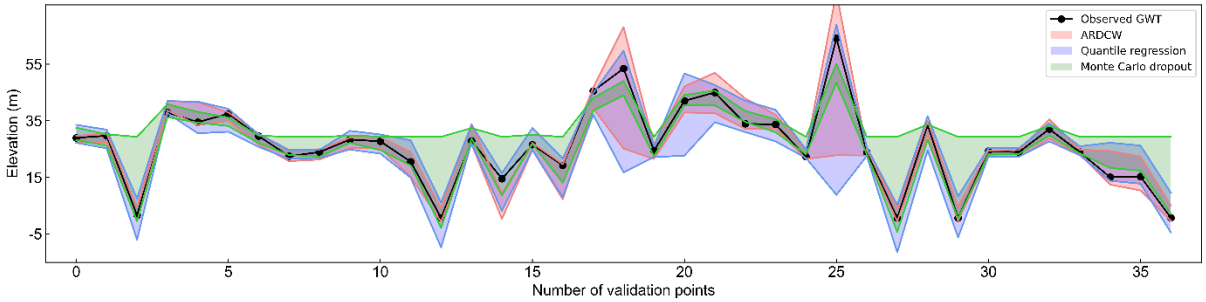


Figure 13. Compared $\boldsymbol{P_{CI}^{95\%}}$ of three models using validation dataset.

Table 2. Compared statistical criteria of the three uncertainty methods

	Performance criteria					**Ranking of criteria**						
Model	E_c	RMSE	IA	R^2	$P_{PI}^{95\%}$	E_c	RMSE	IA	R^2	$P_{PI}^{95\%}$	Total rank	Sort order
ARDCW	0.97	2.51	0.99	0.97	8.2	3	3	3	3	3	15	3
QR	0.96	2.99	0.99	0.96	11.6	2	2	3	2	1	10	2
MCD	0.94	3.5	0.98	0.94	9.4	1	1	1	1	2	6	1

Figure 14 shows the 95% uncertainty intervals estimated for the studied area (right). It provides the distribution of different uncertainty levels and gives an indication of the reason for high uncertainties, e.g., whether they result from the lack of observed data or from the sudden big changes in groundwater levels. The location of the boreholes and the levels of groundwater can be observed from the colour plot as shown in Figure 14 (left). It can be seen that in a few areas the uncertainties are ~ 50 meters. This is due to the lack of data points and sudden GWT changes (an up to 40 meter level difference between adjacent boreholes) by looking at the same areas in Figure 14 (left). This map can help in planning for future data collection and borehole drilling locations to reduce high uncertainties in estimating the groundwater surface.

Figure 15 shows a 3D view of the predicted groundwater surface together with the estimated upper and lower boundaries.

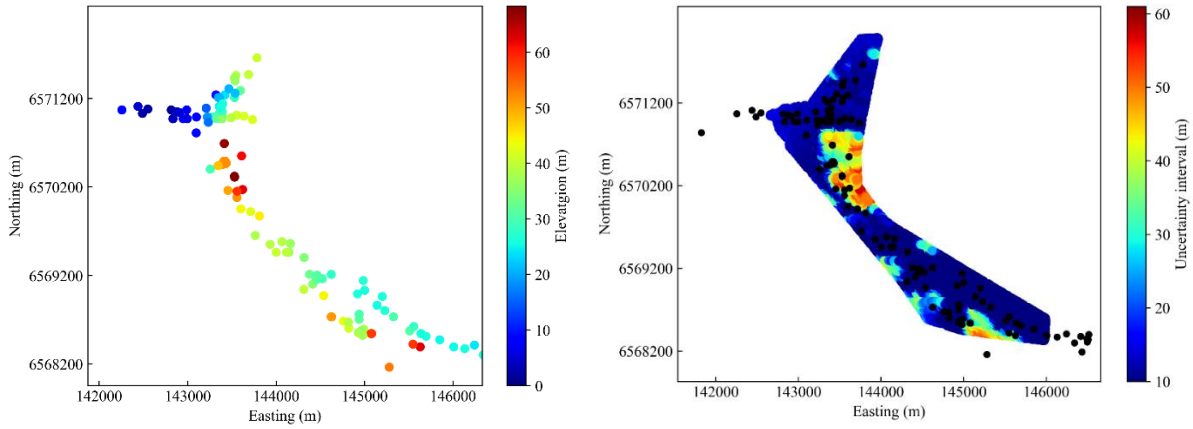


Figure 14. Colour map of the measured GWT data (left) and estimated uncertainty intervals using the measured GWT data (black dots) for a small part of the project area (right).

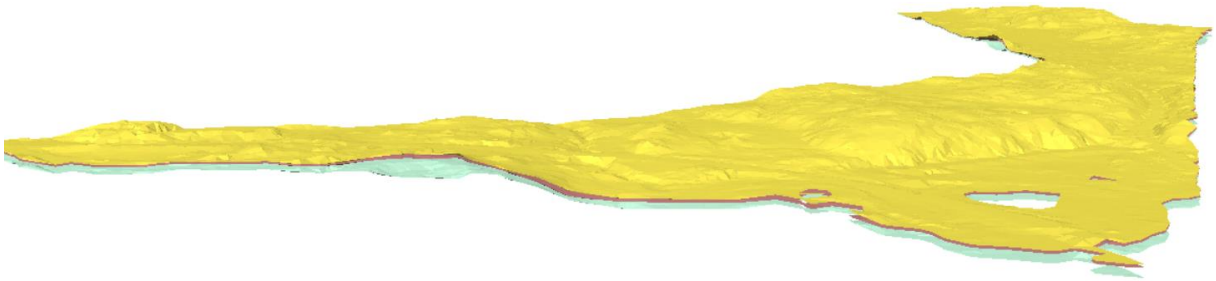


Figure 15. A 3D view of the estimated groundwater surface (brown, middle) for the one small part of the study area together with the upper (yellow) and lower (light blue) boundaries.

5 Implementation of the work

The result of this project was presented in terms of code and implemented/integrated with the GeoBIM system to create an automated process for 3D bedrock surface modelling. In the system, the calculations for bedrock surface modelling were made on the GeoBIM server, and the user received an email with the link to the final 3D bedrock surface model together with estimated uncertainties. This can facilitate all the geotechnicians, rock engineers and engineering geologists who use bedrock surface models for decision making and project planning. We need to point out that only those members of projects added into the GeoBIM system can use this function. Other users can find the theories of the study in our open access article (Shahri et al., 2021). We also have plans to make our Python code open access in GitHub so that it is available for everyone to use with their own data.

6 Conclusion

In this study, deep learning is used to map the bedrock and groundwater surfaces and estimate the uncertainties in field measured data. Moreover, a novel state-of-the-art ensembled-based deep learning approach for uncertainty analysis was proposed, developed and applied. This new approach is abbreviated as ARDCW, denoting the automated random deactivating connective weights. The comparison between deep learning and OK with respect to bedrock surface modelling shows that deep learning can generate a surface that is closer to the true surface. The results of uncertainty quantification show that deep learning can estimate more

accurate uncertainties that cover 4% more of the true values in the intervals than OK does. The comparison of different uncertainty quantification methods in the GWT modelling shows that the ARDCW method has the highest scores on all the selected statistical criteria and thus is the best method for uncertainty analysis concerning this dataset. It is important to point out that our method can build time dependent GWT models by just simply selecting data from different time intervals.

Due to approved robustness against changes and noise in the predictors, the AI models have gained more attractions than traditional statistical/probabilistic analyses. Moreover, AI models promisingly can provide efficient solutions in estimating the non-Gaussian, non-stationary and complex cases while for example ordinary kriging due to the used assumptions for nonlinear, and categorical data is not necessarily optimal, and the associated variance is often overly optimistic.

From this study it can be concluded that deep learning gives a good performance for recovering the variation of subsurface bedrock topography, and the predicted surface is more accurate than the geostatistical method. QR show 12% improvement and ARDCW shows 3% improvement compared with OK regarding the tested dataset. The uncertainty quantification using deep learning, which also shows superior performance and a better estimation of the uncertainties, covering more true values than the uncertainty quantification with Ordinary Kriging. The produced outcomes and interpreted results indicated that the developed AI models are feasible, cost-effective, economic and accurate enough tool to be applied for geospatial DTB and GWT surface predictions.

7 Referenser

[1] Agerberg, F., 2020, Data interpolation for groundwater modelling- How choice of interpolation method and sample size affect the modelling results. Master thesis, Luleå University of Technology.

[2] Carlsson, M. (2005). Management of geotechnical risks in infrastructure projects: an introductory study. Licentiate Thesis, division of soil and rock mechanics, KTH.

[3] Chang, J.R., & Chao, S.J. (2009). Applying group method of data handling (GMDH) method to predict depth to bedrock. Computing in Civil Engineering, ASCE, doi: 10.1061/41052(346)27.

[4] Charniak, E. (2019). Introduction to deep learning. The MIT press.

[5] Fournier, A., Mosegaard, K., Omre, H., Sambridge, M., & Tenorio, L., (2013). Assessing uncertainty in geophysical problems — Introduction. Geophysics, 78(3), 1MJ-Z75.

[6] Goovaerts, P. (1997). Geostatistics for natural resource evaluation. Oxford University Press.

[7] Hengl, T., Mendes de Jesus, J., Heuvelink, G.B.M., Ruiperez, G. M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M.N., Geng, X., Bauer-Marschallinger, B., Guevara, M.A., Vargas, R., MacMillan, R.A., Batjes, N.H., Leenaars, J.G.B., Ribeiro, E., Wheeler, I., Mantel, S., & Kempen, B. (2017). SoilGrids250m: Global gridded soil information based on machine learning. PLoS ONE, 12(2): e0169748., doi: 10.1371/journal.pone.0169748.

[8] Heuvelink, G., & Pebsema, E. (2002), Is the ordinary kriging variance a proper measure of interpolation error? Institute for Biodiversity and Ecosystem Dynamics, Universiteit van Amsterdam.

[9] Hood, S.B., Cracknell, M.J., Gazley, M.F., & Reading, A.F. (2019). Improved supervised classification of bedrock in areas of transported overburden: Applying domain expertise at Kerkasha, Eritrea. Applied Computing and Geosciences, 3–4, 100001, doi: 10.1016/j.acags.2019.100001.

[10] Hu, L., & Jiao, J.J. (2010). Modeling the influences of land reclamation on groundwater systems: A case study in Shekou peninsula, Shenzhen, China.

Engineering geology, 114(3–4),144–153. https://doi.org/ 10.1016/j.enggeo.2010.04.011
[11] Kitanidis, P. K. (1997). Introduction to geostatistics: Applications in Hydrogeology. Cambridge: Cambridge university press.
[12] Kitterød, N. (2017). Estimating unconsolidated sediment cover thickness by using the horizontal distance to a bedrock outcrop as secondary information. Hydrol. Earth Syst. Sci., 21, 4195–4211, doi: 10.5194/hess-21-4195-2017.
[13] Kitterod, N., & Leblois, E., 2019, Modelling Bedrock Topography. Earth Surface Dynamics, Discuss. [preprint], https://doi.org/10.5194/esurf-2019-57.
[14] Koc, K., Grugun, A., & Ozbek, M.E. (2020). Effects of geotechnnical risks on cost and schedule in infrastructure projects. Proceedings of International Structural Engineering and Construction, 7(2), doi.org/10.14455/ISEC.2020.7(2).CON-18.
[15] Koenker, R., & Hallock, K.F. (2001). Quantile regression and introduction. University of Illinois at Urbana-Champaign.
[16] Li, J., & Heap, A. D. (2014). Spatial interpolation methods applied in the environmental sciences: A review. Environmental Modelling & Software, 53, 173-189. doi:10.1016/j.envsoft.2013.12.00.
[17] Li, J., Wan, H., & Shang, S. (2020). Comparison of interpolation methods for mapping layered soil particle-size fractions and texture in an arid oasis. Catena, 190. doi:10.1016/j.catena.2020.104514.
[18] Machado, B., (2019), Artificial intelligence to model bedrock depth uncertainty. Degree project architecture, KTH Royal Institute of Technology.
[19] Marsland, S., 2018. Machine learning: An algorithmic perspective, Taylor & Francis Group, CRC Press.
[20] Nash, J.E., & Sutcliffe, J.V. (1970). River flow forecasting through conceptual models part I – A discussion of principles*. J Hydrol.,* 10, 282–290. https://doi.org/ 10.1016/0022– 1694(70)90255–6.
[21] Nordström, A. (2005). Dricksvatten för en hållbar utveckling. Lund: Studentlitteratur AB.
[22] Salvo, C.D., Mancini, M., Cavinato, G.P., Moscatelli, M., Simionato, M., Stigliano, F., Rea, R., & Rodi, A. (2020). A 3D geological model as a base for the development of a conceptual groundwater scheme in the area of the Colosseum (Rome, Italy). *Geosciences*, 10, 266. https://doi.org/10.3390/geosciences10070266.
[23] Samui, P, & Sitharam, T.G. (2011). Application of geostatistical models for estimating spatial variability of rock depth, Engineering, 3, 886-894, doi:10.4236/eng.2011.39108.
[24] Seoh, R. (2019). Qualitative analysis of Monte Carlo Dropout. arXiv:2007.01720 [stat.ML]
[25] Shahri, A.A., Shan, C., Zäll, E., & Larsson, S., (2021), Spatial distribution modelling of subsurface bedrock using a developed automated intelligence deep learning procedure: A case study in Sweden. Journal of Rock Mechanics and Geotechnical Engineering, https://doi.org/10.1016/j.jrmge.2021.07.006.
[26] Shahri, A.A., Larsson, S., & Renkel, C. (2020). Artificial intelligence models to generate visualized bedrock level: a case study in Sweden. Modeling Earth Systems and Environment, 6, 1509–1528. https://doi.org/ 10.1007/s40808–020–00767–0.
[27] Sveriges geologiska undersökning. (n.d.). Risker. Retrieved April 18, 2020, from https://www.sgu.se/samhallsplanering/risker/.
[28] Tang, Y., Zhou, J., Yang, P., Yan, J., & Zhou, N. (2017). *Groundwater engineering*. Springer. https://doi.org/ 10.1007/978–981–10–0669–2.
[29] Todd, D. K., & Mays, L. W., (2005), Groundwater Hydrology (3. ed.). Danvers: John Wiley & Sons Inc.
[30] Viswanathan, R., Jagan, J., Samui, P., & Porchelvan, P. (2014). Spatial variability of rock depth using simple kriging, ordinary kriging, RVM and MPMR. Geotech Geol Eng., 33(1):69-78, doi: 10.1007/s10706-014-9823-y.

[31] Willmott, C.J. (1984). On the evaluation of model performance in physical geography. In G.L. Gaile, C.J. Willmott (Eds.), *Spatial statistics and models*. *Theory and decision* (volume 40, pp: 443–460), Springer. https://doi.org/ 10.1007/978–94–017–3048–8_23

Bilaga A Conference Abstracts

A.1 Artificial intelligence-based models to predict the spatial bedrock levels for geoengineering application (extended abstract published)

3RD CONFERENCE OF THE ARABIAN JOURNAL OF GEOSCIENCES (CAJG)
2-5 NOVEMBER 2020, SOUSSE, TUNISIA

Springer
SPRINGER NATURE

Artificial intelligence-based models to predict the spatial bedrock levels for geo-engineering application

Chunling Shan[1,2], Abbas Abbaszadeh Shahri[2,3], Stefan Larsson[1] and Emma Zäll[2]

[1]KTH Royal Institute of Technology, 11428 Stockholm, Sweden
[2]Tyréns, 11886 Stockholm, Sweden
[3]Johan Lundberg AB, 75103 Uppsala, Sweden
chunling.shan@tyrens.se

Abstract. Delineate and mapping the bedrock and overlaid deposits due to complex spatial patterns, associated uncertainties and sparse data is a vital difficult task in geo-engineering applications. The improperly maps of spatial subsurface bedrocks can cause significant influence on costs and the risks of a project. In the past decades, different geophysical techniques and geomathematical models commonly have been used for subsurface imaging, but due to well recognized drawbacks such methods can neither the only nor the optimum solutions. This motives to examine modern computational artificial intelligence-based models (AIM) to overcome the deficiencies and underlying assumptions of previous methods. Moreover, approved abilities of AIM in solving the complex problems as well as capacity to work with imprecise and different types of data provide a powerful alternative.

Investigating the feasibility of AIM in prediction of 3D spatial distribution of subsurface bedrock in a large area in Stockholm, Sweden is the objective of this study. To have flexibility and access to more opportunities, the modeling process using soil-rock sounding borehole data comprising the geographical coordinates and ground surface elevation was carried out in both C++ and Python. To develop the codes, an algorithm using trial-error method based on randomized data into 65%, 25% and 15% for training, testing and validation was designed and compiled. In this process 6 training algorithms and 6 activation functions subjected to different internal characteristics were examined. The primary investigations using the calculated root mean square showed that the optimum model can have between 10 to 25 neurons. These neurons further were managed into numerous topologies in different hidden layers to capture the optimum models. All organized topologies were evaluated and ranked using different error metrics. According to ranked models, the 3-20-5-1 and 3-8-17-1 structures subjected to Quasi Newton and Nadam (Nesterov-accelerated Adaptive Moment) using tangent hyperbolic and PReLu (Parametric ReLU) were selected as optimums.

The optimum models then were used for prediction of bedrock levels and supported by the high resolution digital elevation surface data in the same study area. It was observed that in situation with sparse dataset, the developed models efficiently can provide much more accurate prediction than previously applied techniques such as geostatistical approaches. This implies that the developed AIM due to remarkable capacities and acceptable predictability level can decrease the residuals between the predicted and measured data.

Keyword: Sweden, bedrock, artificial intelligence, optimum model, spatial distribution

A.2 Uncertainty analysis of an optimum predictive neural network model in subsurface bedrock level modelling (extended abstract published)

MLRA2021
MACHINE LEARNING & RISK ASSESSMENT

Uncertainty analysis of an optimum predictive neural network model in subsurface bedrock level modeling

Abbas Abbaszadeh Shahri[1,2], Chunling Shan[1,3], Emma Zäll[1] and Second Author[3]

[1]Rock engineering, Tyréns, Peter Myndes Backe 16, 11886, Stockholm, Sweden.
[2]Johan Lundberg AB, Box 62, 75103, Uppsala, Sweden.
E-mail: abbas@johanlundberg.se, chunling.shan@tyrens.se, emma.zall@tyrens.se
[3]KTH Royal Institute of Technology, Brinellvägen 8, 10044, Stockholm, Sweden.
E-mail: stela@kth.se

Keywords: Artificial neural network; Geotechnical boreholes; Bedrock surface model; Uncertainty.

1 Introduction & Scope of Work

The outcome of predictive geo-engineering models include uncertainties. Therefore, formal frameworks are needed for the uncertainty quantification (*UQ*), to assess the reliability of the models and reduce hesitancy in both computations and real world applications. In prediction processes, the uncertainity is a combination of three main sources i.e. modeling (describing the real system), numerical (mathematical equations) and data meaurements. However, their influence on the predictions are not equal. Therefore, the *UQ* level is important for identifying limitations, increase precision and avoid unwarranted conclusions. Over the years, the computational efforts of *UQ* has been the subject of a prime concern in predictive geo-engineering models (Abbaszadeh Shahri et al., 2020; Helton and Sallaberry, 2017). Accordingly, future forecasts of such models can be significantly affected due to inherent uncertainties. In this study, the reliability of predicted bedrock levels for a large area in Stockholm, Sweden, was evaluated using different *UQ* indices. Accordingly, such *UQ* indices represent different ways of comparing the true and predicted value at the same point. The process was carried out using a developed optimum artificial neural network model on a dataset with 1967 geotechnical soil-rock soundings.

2 Methodology

Using *UQ* methods, the level of confidence for each measurement can be estimated. The *UQ* then allows proper judgments on the quality of the experiments and thus facilitate meaningful comparisons with other similar values or a theoretical prediction (Iman and Helton, 1988). However, finding an ideal *UQ* depends on available computational resources (Sacks et al., 1989). Statistically, a process is in control if most of its variation falls within a certain range. The confidence interval (*CI*) is a computed range of observed data that covers the true future populations of a predictive model, with a certain probability. It can also show the stability of the estimates, where the wider the *CI*, the more instable the estimate. Accordingly, the level of 95% means that 5% of the predicted value lies outside the *CI*. The prediction interval (*PI*) shows the certain probability of an estimation of future observation and is often used in regression analysis.

3 Research Outcomes

Referring to Figure1, the calculated error margins of *CI* and *PI* for training and validation datasets imply that most of the predictions fall within these limits. The confidence in experimental data then can be increased if the survey is repeated several times using the same method.

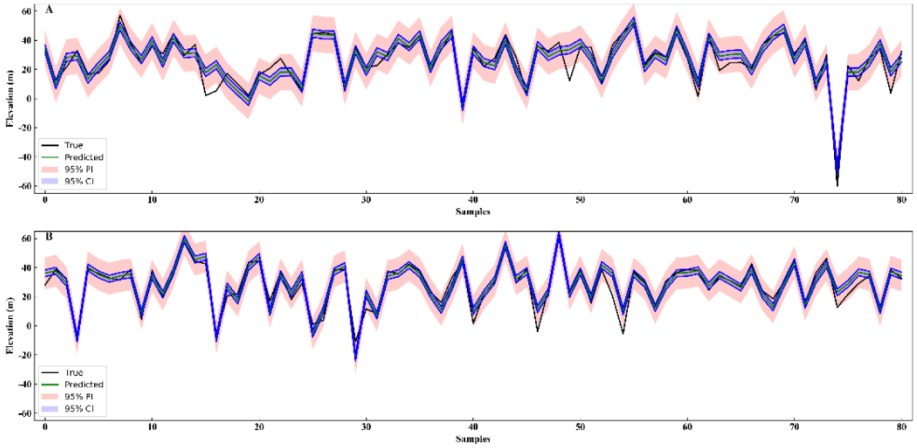


Figure 1. Estimated *CI* and *PI* for validation (A), and testing (B) dataset

The point(s) outside the limits indicates the presence of non-random variation which makes the process statistically unstable (Deng et al., 2012). In under control process, there is a probability of a point exceeding control limits i.e. false occurrence on average once every (1/probability) observations. To estimate the total uncertainty of a measurement, the *UQ* formaly is performed through statistical metrics. Analyzed results of the developed optimum neural network model and then the employed metrics for *UQ*, showed the importance of the standard deviation (σ) of each experiment for both the measured and predicted values, where the lower the σ, the lower the uncertainty. This implies on more confidence and thus higher reliability in the experiments. According to calculations, the values of ± 2.8 m and ±12 m were considered for *CI* and *PI* as the uncertainty and error bias in the measured data. Referring to achievements, validation step showed 20% and 1.1% progresses compared to the training data in *CI* and *PI* respectively.

References

Abbaszadeh Shahri A, Renkel C, Larsson S (2020) Artificial intelligence models to generate visualized bedrock level: a case study in Sweden. Model. Earth Syst. Environ., 6, 1509–1528.

Helton JC, Sallaberry CJ (2017) Treatment of uncertainty in performance assessments for the geological disposal of radioactive waste. 2nd Eds, Geological repository systems for safe disposal of spent nuclear fuels and radioactive waste, Woodhead Publishing Series in Energy, Elsevier, 499-527.

Sacks J, Welch WJ, Mitchell TJ, Wynn HP (1989) Design and analysis of computer experiments. Statistical Science, 4(4), 409–423.

Iman RL, Helton JC (1988) An investigation of uncertainty and sensitivity analysis techniques for computer models. Risk Analysis, 8(1), 71–90.

Deng H, Runger G, Tuv E (2012) System monitoring with real-time contrasts. Journal of Quality Technology, 44(1), 9–27.

Bilaga B Full Papers

B.1 Spatial distribution modelling of subsurface bedrock using a developed automated intelligence deep learning procedure: A case study in Sweden (article published)

Journal of Rock Mechanics and Geotechnical Engineering 13 (2021) 1300–1310

Contents lists available at ScienceDirect

Journal of Rock Mechanics and Geotechnical Engineering

journal homepage: www.jrmge.cn

Full Length Article

Spatial distribution modeling of subsurface bedrock using a developed automated intelligence deep learning procedure: A case study in Sweden

Abbas Abbaszadeh Shahri [a,b,*], Chunling Shan [a,c], Emma Zäll [a], Stefan Larsson [c]

[a] Division of Rock Engineering, Tyréns AB, Stockholm, Sweden
[b] Johan Lundberg AB, Uppsala, Sweden
[c] Division of Soil and Rock Mechanics, KTH Royal Institute of Technology, Stockholm, Sweden

ARTICLE INFO

Article history:
Received 6 April 2021
Received in revised form
20 May 2021
Accepted 17 July 2021
Available online 21 September 2021

Keywords:
Automated intelligence system
Predictive depth to bedrock (DTB) model
Three-dimensional (3D) spatial distribution

ABSTRACT

Due to associated uncertainties, modelling the spatial distribution of depth to bedrock (DTB) is an important and challenging concern in many geo-engineering applications. The association between DTB, the safety and economy of design structures implies that generating more precise predictive models can be of vital interest. In the present study, the challenge of applying an optimally predictive three-dimensional (3D) spatial DTB model for an area in Stockholm, Sweden was addressed using an automated intelligent computing design procedure. The process was developed and programmed in both C++ and Python to track their performance in specified tasks and also to cover a wide variety of different internal characteristics and libraries. In comparison to the ordinary Kriging (OK) geostatistical tool, the superiority of the developed automated intelligence system was demonstrated through the analysis of confusion matrices and the ranked accuracies of different statistical errors. The results showed that in the absence of measured data, the intelligence models as a flexible and efficient alternative approach can account for associated uncertainties, thus creating more accurate spatial 3D models and providing an appropriate prediction at any point in the subsurface of the study area.

© 2021 Institute of Rock and Soil Mechanics, Chinese Academy of Sciences. Production and hosting by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

In many countries, including Sweden, subsurface modelling is increasingly becoming a necessary part of three-dimensional (3D) urban planning. To create an informative and useful subsurface model, different data types need to be combined. Because of the dynamic nature of the subsurface and variation of implemented data density during the planning process, subsurface modelling techniques are not easily interoperable. Moreover, planners and construction experts are primarily looking for knowledge on the location of geological discontinuities, such as the surface of crystalline bedrock and the boundaries of soft sediments, as well as their geo-engineering properties. Therefore, depth to bedrock (DTB), measured as the thickness of the sediments above the bedrock, is of great interest in subsurface geo-engineering modelling and risk assessment (Sundell et al., 2015; Gomes et al., 2016; Wei et al., 2016; Ghaderi et al., 2019; Yan et al., 2020). Accordingly, the information on the spatial distribution of DTB is an important issue in both the design and construction phases. This is a concern in European countries due to the vast variety of geological conditions and the need to find a solution to various challenges associated with the urbanisation of densely populated cities, while meeting environmental regulations (Athanasopoulou et al., 2019). Therefore, visualised DTB models that include the interpretation of sparse geotechnical measurements are important tools for identifying solutions. Despite the presented essential knowledge on field development (e.g. Glasgow, Stockholm, Helsinki, and Oslo), for the cities on thick sequences such as Rotterdam and Vienna, the digitisation of DTB is not an issue (Schokker et al., 2017; Abbaszadeh Shahri et al., 2020). However, because of the associated uncertainties (Baecher, 1986), producing highly accurate DTB predictive models is a critical task that can have significant effects on the costs and risks of geo-engineering projects (Clarke et al., 2009; Mey et al., 2015).

* Corresponding author. Division of Rock Engineering, Tyréns AB, Stockholm, Sweden.
E-mail addresses: shahri@kth.se, shahri.abbas@bircham.edu (A. Abbaszadeh Shahri).

Peer review under responsibility of Institute of Rock and Soil Mechanics, Chinese Academy of Sciences.

https://doi.org/10.1016/j.jrmge.2021.07.006

1674-7755 © 2021 Institute of Rock and Soil Mechanics, Chinese Academy of Sciences. Production and hosting by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

The characterisation of DTB profiles is commonly interpreted through the sufficient sparse numbers of geotechnical soundings in and around a desired area. However, geotechnical investigations may suffer from certain limitations, for example, limited access to an entire area, costs of investigations and distance between the soundings. Consequently, with the increment of distance between two soundings, the uncertainty will increase abruptly, as many points are needed to estimate or remain totally unknown. To generate a continuous predictive DTB model, geophysical techniques (Abbott and Louie, 2000; Dowd and Pardo-Iguzquiza, 2005; Christensen et al., 2015; Nath et al., 2018), random fields (Uzielli et al., 2005; Li et al., 2015), geostatistical tools (Samui and Sitharam, 2011; Viswanathan et al., 2014; Kitterød, 2017), and variogram-based methods (Maus, 1999; MacCormack et al., 2018) as well as geomorphological-based models (Del Soldato et al., 2018) have been actively employed. In these methods, results from geotechnical soundings are interpolated to estimate the DTB in between the soundings that capitalises on the spatial structure and semivariance of the measured data (Goovaerts, 1997). Moreover, planar mesh generation, spatial interpolation and surface intersection are other generic widely used techniques in geological modelling (Mei, 2014). In addition to the abovementioned techniques, the falling weight deflectometer (Roesset et al., 1995), extracted attributes from satellite images (Kuriakose et al., 2009; Sun and Kim, 2017; Yan et al., 2020), topographic data (Gomes et al., 2016), and signal analyses (Lane et al., 2008; Setiawan et al., 2018; Du et al., 2019) have been highlighted as effective models.

In site investigation, the geophysical techniques can provide supplementary information on sparse observations (e.g. borings, test pits and outcrops). However, these techniques are limited to small-scaled surveys in ground-based techniques (Erkan, 2008). Moreover, geophysical data need to be correlated with information from direct geotechnical methods, as data are generally interpreted qualitatively, and useful results can only be obtained by experts familiar with the particular testing method (Pazzi et al., 2019; Christensen et al., 2015). Compared to geotechnical soundings, these testing methods can be complex and time-consuming because of the need for specialised equipment and experienced operators as well as logistical issues (Clayton and Smith, 2013).

Geostatistical tools are one of the most used interpolation technologies for DTB maps (Abbaszadeh Shahri et al., 2020). By enhancing the spatial distribution of data, these techniques offer convenient options for management and provide continuity that can reproduce the trend of DTB. This feature allows the user to be precise in interpretation; however, the success of a produced DTB map depends on the quality of available information on the study area and independent variables (Deutsch, 1996). Systematic sampling uses a fixed grid to assign values in a regular pattern. Cell change cannot be accounted for in this method, and interpolation thus estimates the centre of each unmeasured grid cell (Baskan et al., 2009). This limitation implies that the location of points may be problematic when using random sampling distribution, and the coverage of adjacent areas may not be supported. Therefore, spatial reconstruction of a given finite number of observations at different locations implies that measurements have been taken under measurement noise (Stein, 1999).

Random field theory is a mathematical definition using the Euler characteristic for smooth statistical maps that address the threshold problems in functional imaging (Brett et al., 2004). If datasets are limited in size, this method is not an appropriate alternative due to a complex training stage and computational potency (Fenton, 1999).

In recent years, artificial intelligence (AI) techniques have shown remarkable computational and learning capabilities in addressing geotechnical problems. As DTB modelling deals with various uncertainties (Gomes et al., 2017; Hood et al., 2019; Abbaszadeh Shahri et al., 2020), the subcategories of AI techniques are appropriate alternatives to overcome the limitation and simplifications of the illustrated methods (e.g. Chang and Chao, 2009; Hengl et al., 2017; Abbaszadeh Shahri et al., 2020). Furthermore, hybridising the AI techniques with metaheuristic algorithms can significantly optimise the model performance (Asheghi et al., 2019; Abbaszadeh Shahri et al., 2021). Different methods applied in DTB modelling are summarised in Table 1.

As illustrated, depending on the interpolation algorithm applied, different results can be observed in the produced geological DTB models. Therefore, it is not always clear which method can provide the most appropriate outcome. Accordingly, the resolution of complex 3D geological models can be increased and supplemented by ensuring accurate geospatial distribution of DTB. This study was motivated by the need to address such a challenge in a geo-engineering project in Stockholm, Sweden, where producing an adequately accurate quantitative model is of great interest. To find the optimum predictive DTB models, an automated AI training scheme was designed and developed, and then programmed using Python and C++. This allowed many different internal characteristics and optimiser to be tested in Python and C++. The proposed procedure was applied on 1968 datasets from soil–rock soundings in an urbanised area in Stockholm. Due to the use of automated programmes, the identified optimum models showed superior performance and more accurate spatial DTB compared to conventional ordinary Kriging (OK) technique. The results have an impact on the ability to reduce the number of boreholes and corresponding costs when using developed models.

2. Study area and data source

The study area encompasses a 20 km stretch of an ongoing highway project in Stockholm, Sweden. This area consists mainly of fine-to coarse-grained gneiss of sedimentary origin and medium-to coarse-grained metavolcanic rocks as well as occasional coarse-grained pegmatite passages. Sedimentary gneisses generally dominate in the area. According to the bedrock map provided by the Geological Survey of Sweden (SGU), the faults in the area include one with a SE-NW direction that is the result of structural

Table 1
A summary of applied techniques to predict the DTB.

Source	Used method	Study area
Chang and Chao (2009)	FWD, PNN	Texas, USA
Abbaszadeh Shahri et al. (2020)	ANN, RBF, GFFN, SVR	Stockholm, Sweden
Hengl et al. (2017)	GBT, RF, ANN	Worldwide data
Hood et al. (2019)	RF, RS, airborne geophysics	Kerkasha, Eritrea
Mey et al. (2015)	ANN	Rhone Valley-Alp, Switzerland
Yan et al. (2020)	RS, RF, GBT	China
Clarke et al. (2009)	ANN	Glacial and fluvial valleys
Viswanathan and Samui (2016)	GPR, least squares SVM, ELM	Chennai, India
Sitharam et al. (2008)	Geostatistic, ANN, SVM	Bangalore, India
De Carneiro et al. (2012)	Self-organising ANN, airborne geophysics	Brazilian Amazon
Pfaffhuber et al. (2019)	Geophysics, ANN	Norway
Del Soldato et al. (2018)	GIST	Mountain ranges, Italy

Note: ANN – Artificial neural network; SVM – Support vector machine; RBF – Radial basis function; SVR – Support vector regression; RF – Random forest; RS – Remote sensing; FWD – Falling weight deflectometer; PNN – Polynomial neural networks; GPR – Gussian process regression; ELM – Extreme learning machine; GBT – Gradient boosting tree; GFFN – Generalised feedforward neural network; GIST – Geomorphological index soil thickness.

deformation zone. The widths are decreased from 75–100 m in larger faults to 50 m in smaller faults. The plan for this highway in the NW-SE direction crosses the existing bedrock, where the road will be built as concrete tunnels in some sections. Among the executed geotechnical tests and acquired data, 1968 soil–rock soundings were compiled in the area (Fig. 1a). These soundings encompass a varied and complex set of data derived from subsurface explorations and in situ instrumentation. However, a lack of data needed to provide a consistent database is an ongoing challenge not only in this study but also in most geo-engineering applications. This limitation in the ability to improve the datasets was overcome by using a random data creator (RDC) (Abbaszadeh Shahri et al., 2020), an intelligent knowledge-based framework used to generate appropriate pseudo observations that can be used to compare, interpret, and describe the results. Accordingly, 62 new pseudo datasets were generated for the area (Fig. 1a, points with black +) to extend the region of influence for each soil–rock sounding and decrease the degree of variability in the extrapolation direction. These retrieved sparse data, which have been distributed alongside the planned road, can be used to supplement the collected DTB information from geotechnical soil–rock soundings, as this is the most common probing method used in Sweden that can be performed in both soil and rock. This method can provide good and accurate soil and rock interfaces, but there is uncertainty in the interpretation of bedrock levels when the top of the bedrock is cracked and brittle. The thickness of overlaid postglacial sediments varies from 0 to more than 100 m. An overview of the constructed digital elevation model (DEM) and geological setting of the area is shown in Fig. 1b and c, respectively.

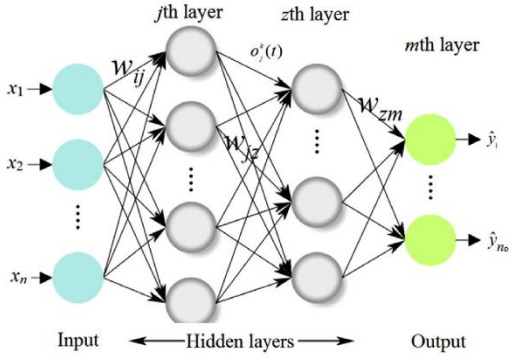


Fig. 2. Simple configuration of multilayered ANN structure.

3. Artificial neural network processing paradigm

ANNs, as connectionist computing systems of processing elements, are configured for specific applications through a learning process that aims to mimic and replicate the operation of the human brain. Recent developments in system analyses and the significant proven advantages over traditional modelling approaches have led to extensive use of ANN techniques. Properly tuned ANNs improve the diagnostic performance and modifications and thus

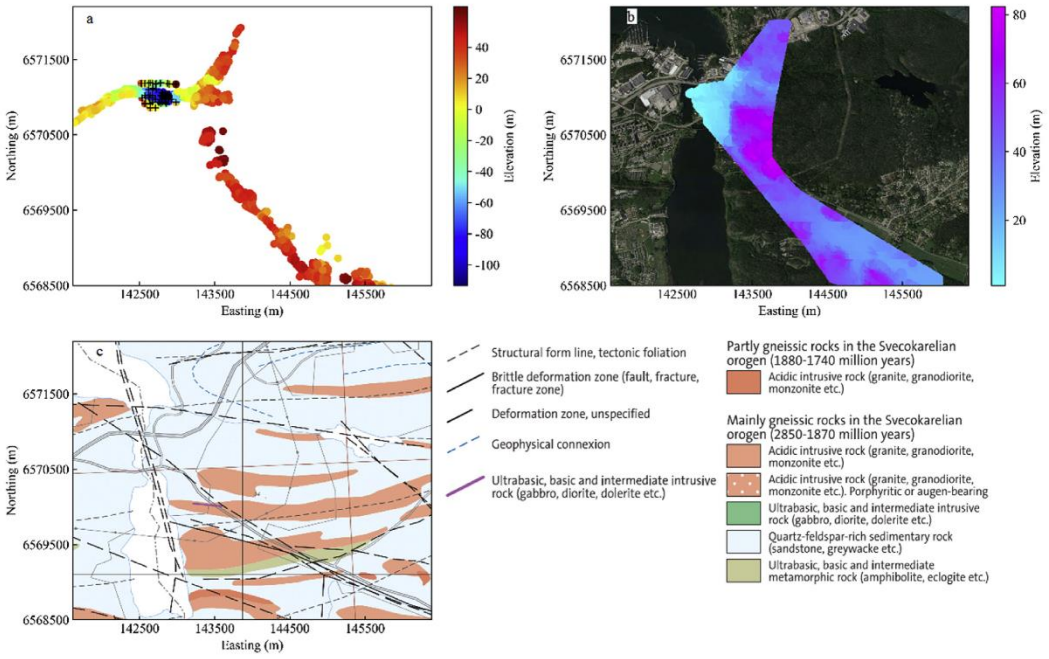


Fig. 1. Colour plot of the DTB measured through soil–rock soundings with the generated RDC data (black +) (a), the overview of overlaid DEM of study area and satellite map taken from Google earth (b), and the geological map of the study area from SGU (c).

are easily adapted to incorporate new data. The goal is to fit outputs with a linear function of nonlinear transformed inputs where any gradient optimisation method may be used.

As presented in Fig. 2, the received signal from the ith input (x_i) is associated with weights connected to the jth neuron (w_{ij}) and is passed through one or more hidden layers to be processed. The output of the jth neuron of the kth hidden layer in the tth iteration ($o_j^k(t)$), using activation function (f), is confined into a pre-defined range and then transferred as

$$o_j^k(t) = f\left(\sum_{i=1}^{n_{k-1}} w_{ij}^k x_i^{k-1}(t) + b_j^k\right) \quad (j \le n_k) \tag{1}$$

where n_k is the number of neurons in the kth layer, and b_j^k denotes the bias which shifts the summed signals received from the neuron.

The error of neuron j at the tth iteration, $e_j(t)$, and the corresponding error signal of the ith neuron in the kth hidden layers $\rho_i^k(t)$, as well as the total error for the entire network, $E(t)$, are then expressed by

$$e_j(t) = d_j(t) - o_j^k(t) \tag{2}$$

$$\rho_i^k(t) = \dot{f}\left(x_i^k(t)\right)\sum_j \rho_j^{k+1}(t) w_{ji}^{k+1}(t-1) \quad (k = 1, 2, \ldots, m-1) \tag{3}$$

$$E(t) = \frac{1}{2}\sum_{j=1}^{n} e_j^2 \tag{4}$$

where $d_j(t)$ denotes the desired output of neuron j at the tth iteration, and $\dot{f}(x_i^k(t))$ is the first derivative of $f(x_i^k(t))$ with respect to $x_i^k(t)$.

Accordingly, the weights and biases are updated to minimise the prediction error, $\varepsilon(t)$, as

$$\Delta w_{ij}^k(t) = \eta_w \rho_i^k(t) o_j^{k-1}(t) + \alpha_w \Delta w_{ij}^k(t-1) \tag{5}$$

$$\Delta b_i^k(t) = \eta_b \rho_i^k(t) + \alpha_b \Delta b_i^k(t-1) \tag{6}$$

$$\varepsilon(t) = y - \hat{y}(t) \tag{7}$$

where y is the actual output; α_w and α_b are the momentum constants that determine the influence of the past parameter changes on the current direction of movement in the parameter space, and α_w usually varies within [0.1, 1] interval and is used to avoid instability in the updating procedure; η_w and η_b represent the learning rates; and $\rho_i^k(t)$ is the error signal of the ith neuron in the kth layer, which is back-propagated in the network.

The outcome of the lth neuron in the mth output layer ($\hat{y}_l$) is then calculated using the updated weight by

$$\hat{y}_l(t) = \sum_{i=1}^{n_m-1} w_{ij}^m x_i^{m-1}(t) \quad (l \le n_o) \tag{8}$$

where n_o is the number of neurons in the output layer.

3.1. Developing optimum DTB predictive models

Advanced ANN techniques can be considered robust tools for DTB modelling. However, referring to dependency on the defined problem and a lack of a standardised method for configuration, the identification of an optimum model is a difficult and critical task (e.g. Vogl et al., 1988; Currry and Morgan, 2006; Krasnopolsky et al., 2018; Ghaderi et al., 2019; Abbaszadeh Shahri et al., 2021). During the training procedure, the model should not be trapped in local minima nor overfit. To overcome these problems, the regularisation and tuning of internal characteristics (e.g. training algorithm, number and arrangement of neurons, learning rate, activation function and architecture) play a significant rule (Abbaszadeh Shahri, 2016). Using different combinations of these parameters makes learning faster and prevents convergence in local minima (Abbaszadeh Shahri et al., 2021). Overfitting occurs when a model fits the data in the training set, while incurring a larger generalisation error (Tetko et al., 1995). Regularisation refers to the process of modifying a learning algorithm to prevent overfitting by fixing the number of parameters in the model (Girosi et al., 1995). Early stopping is used as a form of regularisation to control the number of iterations that can be run before the training algorithm begins to overfit. Therefore, in each iteration, early stopping improves the performance of the learning algorithm on data outside of the training set (Zhang and Yu, 2005; Yuan et al., 2007).

To find the optimum models, the presented method by Abbaszadeh Shahri et al. (2020) was updated using an automated identification process (Fig. 3). Using an iterative procedure, this process was then integrated with a constructive technique and

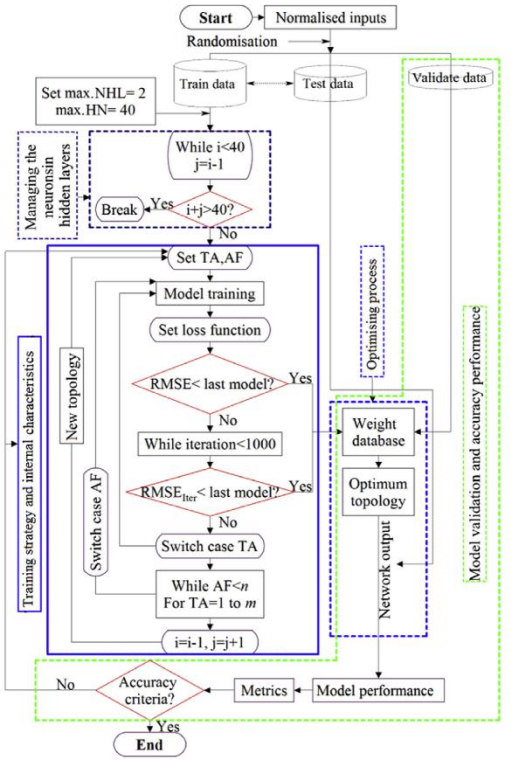


Fig. 3. The block diagram used to assess the optimum model. NHL – Number of hidden layers; HN – Hidden neurons; n – Number of used activation functions; m – Number of used training algorithms.

programmed with both Python and C++. This was done not only to capture the capabilities of both programming languages but also to increase the power of the models and avoid the problems of overfitting and getting stuck in local minima. A programming language is a specification, and the success of an application will therefore depend on making an appropriate choice. Python and C++ were selected due to their popularity, history, and access to libraries. Moreover, these programming languages are turing-complete (by design) from a theoretical standpoint, with quite similar semantics, even if their syntax is very different. Therefore, based on each programmed code, a wide variety of internal characteristics for both shallow and deep neural networks have been examined (Table 2). Deep neural network refers to models with multiple hidden layers that can be reused to compute the features of a combined structure with fewer weights (LeCun et al., 2015). This implies that after learning, deep structures can improve the generalisation to new examples (Kriegeskorte and Golan, 2019). However, the need for adequate computing power and data for learning are the main potential limitations of deep structures (Bengio, 2009).

Fig. 3 shows the block diagram of the proposed method that can automatically capture the optimum models using the characteristics in Table 2. To constrain the search space and save time, the learning rate was set to 0.7 with a step size domain within [0.001, 1]. Using three embedded switch cases in the programmes, the codes were automated to monitor all characterised training algorithms (TA) and activation functions (AF). The maximum numbers of hidden layers and neurons as user-defined parameters were set to 2 and 40, respectively. Therefore, the training strategy is followed by loops and switch cases, where the system starts with one hidden layer and checks the topology of 3-40-1 for all internal characteristics (Table 2). When the system switches to two hidden layers, the procedure automatically starts with 3-1-39-1 structure using the first TA and AF and is pursued to topology 3-39-1-1. Afterwards, it returns with the same TA, and the AF will switch to next case. This iterative process is repeated for all TAs, which are switched step by step to a variety of internal characteristics. Applying different step sizes for the learning rate in TAs, i.e. replacing the conjugate gradient descent (CGD) with adaptive momentum (AM) in step size of 0.001, allows the model to avoid the overfitting problem and maximises the chances of avoiding local minima (Abbaszadeh Shahri et al., 2020). A two-step termination criterion, including the root mean square error (RMSE) and number of iterations (set to 1000), was considered. Accordingly, the minimum RMSE and the maximum network coefficient of determination (R^2) were stored and ranked for all trained structures in three runs.

Table 2
Applied internal characteristics to capture the optimum models.

Programming language	Training algorithm (TA)	Activation function (AF)	Maximum number of hidden layers (user-defined)	Maximum number of used neurons (user-defined)
C++	QP, CGD, AM, QN, LM, SGD	Sig, HyT, Lin, Relu, SS	3	40
Python	AM, Nadam, SGD, RMSProp	Sig, HyT, Lin, PRelu	3	40

Note: QP – Quick propagation; CGD – Conjugate gradient descent; AM – Adaptive momentum; LM – Levenberg-Marqurdt; QN – Quasi Newton; SGD – Stochastic gradient descent; Sig – Sigmoid; Hyt – Hyperbolic tangent; Lin – Linear; Relu – Rectified linear unit; PRelu – Parametric rectified linear unit; SS – Softsign; Nadam – Nesterov-accelerated adaptive moment estimation; RMSProp – Root mean square propagation.

Due to differences in terms of syntax, simplicity, use and overall approach to programming, there is considerable debate over the performance of Python and C++ in specified tasks. As shown in Table 3 and the programmed procedure in Fig. 3, a distribution of 40 neurons in two hidden layers provides 78 different topologies. Therefore, considering the combination of employed AFs between layers, in each round of training, the automated system monitors approximately 2000 topologies with different internal characteristics. This implies that the optimum models are screened among numerous examined structures, even those with similar topologies but different internal characteristics. Accordingly, the variation of network RMSE using 40 neurons in different topologies, starting from 3-1-39-1 to 3-39-1-1, is reflected in Fig. 4. Summarised results (Table 3) show that the 3-28-12-1 and 3-25-15-1 topologies can be selected as optimal topologies. The differences in performance between these languages were expected, as distinctions are raised in terms of syntax, simplicity, use, and the overall approach to programming. This can technically be interpreted as the threading build of each employed language and procedure requirements to become machine code.

3.2. Outcomes of generated DTB models

The outcome and progression of predictive modelling are determined by the effectiveness of systemic feedback loops through structural changes that control whether individual models serve the required needs. Referring to Table 3, the predictability of the captured optimum models using tuned characteristics for both C++ and Python codes was plotted and presented. Fig. 5 shows the comparison between the fitness (Fig. 5a and b) and corresponding differences (Fig. 5c and d) for measured and predicted DTBs using training and testing datasets. According to the no free lunch theorem (Wolpert and Macready, 1997), biases are a fundamental property of the results generated in inductive learning systems, and the assumption of an intelligent model free of biases is not reasonable. In the search space, the achieved possible minimum cost function (RMSE) introduces the bias of predictions. Considering the designed procedure and examined different internal characteristics (Fig. 3), the reason for the observed biases can be referred to the implemented TAs and thus the trade-offs between accuracy, overfitting and overgeneralisation of each choice associated with the corresponding RMSE. Moreover, as the last layer only receives results generated in the previous layer, the detected biases state the differences in the mapping of fed data between the lower layer and its prediction. Traditionally, collected DTB field data are presented in two-dimensional (2D) digital versions of geological maps. However, assigning a vertical component in the areas without soil–rock sounding data in a way that provides a representative interpretation of the subsurface spatial DTB distribution is a challenging procedure (Abbaszadeh Shahri et al., 2020). Despite all the benefits of 2D mapping, there is a trend favouring in use of integrated 3D models with the ability to combine terrain data and aerial photos for geo-engineering applications. Such models provide a visual perspective of the study area, which enables more accurate interpretation through geological sequences. However, depending on the quality of the datasets used and the approach applied, the level of accuracy and the confidence of the model can vary in terms of their ability to prevent conflict with interpolation algorithms. In this respect, if adequate numbers of soil–rock soundings are not accessible, the pseudo data for unsampled locations can be estimated through the knowledge of experts or other methods, such as nearest neighbour and grid cells (Tacher et al., 2006; Abbaszadeh Shahri et al., 2020; Yan et al., 2020). These methods are effective and easy to use, but the

Table 3
Characteristic of optimum ANN-based structures.

Model	TA	R^2			Neuron	Topology	Activation function		$RMSE_{min}$		
		TR	TE	VL			Hidden	Output	TR	TE	VL
C++	CGD	0.96	0.96	0.94	40	3-28-12-1	Hyt	HyT	6	5.55	6.3
Python	RMSProp	0.92	0.93	0.9	40	3-25-15-1	PRelu	PRelu	8.06	6.75	7.84

Note: TR – Training data; TE – Testing data; VL – Validation data.

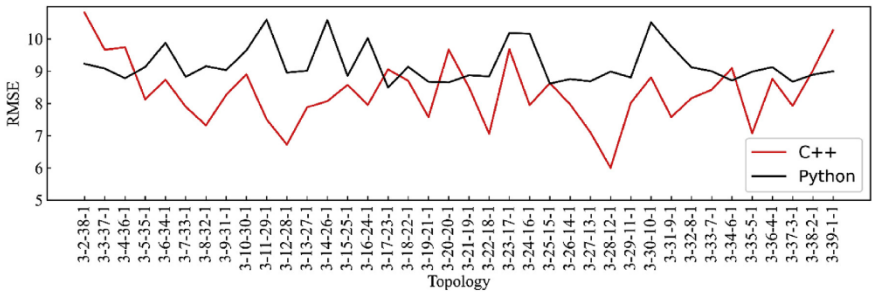


Fig. 4. Variation of network RMSE for developed optimum models using C++ and Python as a function of network topologies.

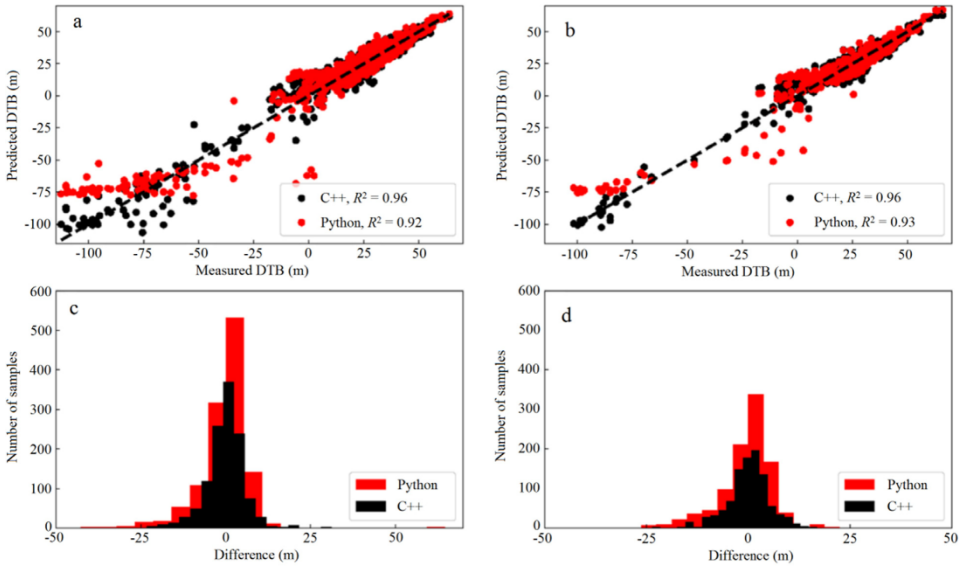


Fig. 5. Comparison of the predictability and corresponding residuals of identified optimum models subjected to developed codes using the training datasets (a, c) and testing datasets (b, d).

uncertainty in the generated pseudo observations in faraway points is increased because of the lack of information nearby, as is the case in all interpolation methods. Accordingly, the generated pseudo data play an important role in building knowledge of phenomena within a specific topic, and data synthesis is thus at the centre of the scientific enterprise in the software engineering discipline (Cruzes and Dyba, 2011). In this study, the drawbacks and concerns over insufficient data were addressed using the RDC approach (Abbaszadeh Shahri et al., 2020). This intelligent knowledge-based system is iteratively applied on test point coordinates to generate new shuffled synthesised DTBs, prescribing the number of soundings and statistical noise for the region between soundings. Nevertheless, the generated 3D models are built over limited number of neighbouring data in relatively small-scale

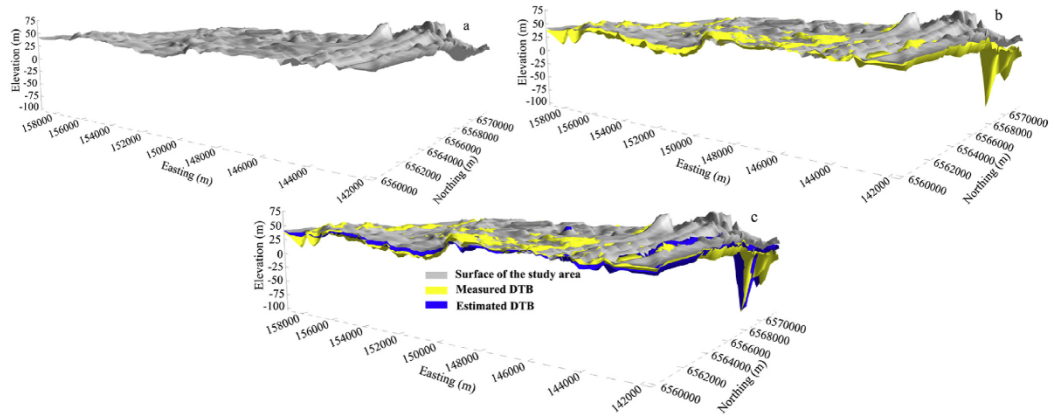


Fig. 6. Step-by-step addition of the 3D distributed spatial DTB predicted by optimum C++ topology: (a) Surface of the area, (b) measured DTB, and (c) estimated DTB.

areas and thus it is not always clear which procedures can provide the most appropriate DTB surfaces.

Fig. 6 shows the results of the creation of a visualised 3D model of the study area from applied data, depicting the retrieved outlines of the subsurface spatial distribution of DTB using the designed training system. The presented 3D model is computed directly from the soil–rock soundings, because the embedded automation procedure can be quickly regenerated. Accordingly, the rock outcrops in the area can be identified by integrating the generated ground surface (Fig. 6a) and spatial measured DTB (Fig. 6b). Such incorporation can provide a 3D subsurface model with high resolution and adequate predictive accuracy in geo-engineering projects (Fig. 6). This implies that the automated procedure provides more flexibility in the modelling process to be developed for future data. Therefore, it can be relevant to exploiting more comprehensive concepts on subsurface geological or petrophysical distributions. Accordingly, such models are a preferred tool for geo-engineers and decision planners in the observation and analysis of geo-environmental engineering issues within a project.

4. Validation and discussion of DTB models

Modelling the spatial distribution of subsurface DTB plays an important role in proper site characterisation. Integrating such DTB

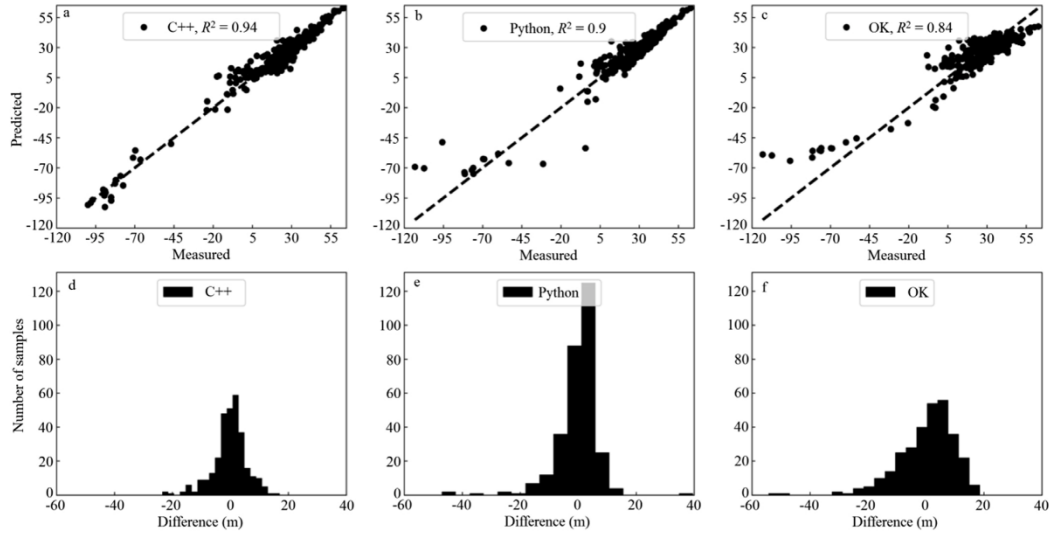


Fig. 7. Comparing the predictability of predicted values using (a) C++, (b) Python, and (c) OK, and the calculated residuals between measured and predicted DTBs achieved from (d) C++, (e) Python, and (f) OK.

Table 4
Established confusion matrices of applied models.

Model		Network output										Results		
	Target	<−113.19	[−113.19, −95.6]	[−95.6, −78]	[−78, −60.41]	[−60.41, −42.81]	[−42.81, −25.21]	[−25.21, −7.62]	[−7.62, 9.98]	[9.98, 27.58]	[27.58, 45.17]	[45.17, 62.77]	>62.77	Total	True	False
C++	<−113.19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	[−113.19, −95.6]	0	0	2	1	0	0	0	0	0	0	0	0	3	0	3
	[−95.6, −78]	0	0	2	0	0	0	0	0	0	0	0	0	2	2	0
	[−78, −60.41]	0	2	1	2	0	0	0	0	0	0	0	0	5	2	3
	[−60.41, −42.81]	0	0	0	2	0	0	0	0	0	0	0	0	2	2	0
	[−42.81, −25.21]	0	0	0	0	0	1	0	0	0	0	0	0	1	0	1
	[−25.21, −7.62]	0	0	0	0	1	0	0	1	0	0	0	0	2	0	2
	[−7.62, 9.98]	0	0	0	0	0	0	0	14	11	0	0	0	25	14	11
	[9.98, 27.58]	0	0	0	0	0	0	0	6	67	21	0	0	94	67	27
	[27.58, 45.17]	0	0	0	0	0	0	0	0	23	115	1	0	139	115	24
	[45.17, 62.77]	0	0	0	0	0	0	0	0	0	7	25	0	32	25	7
	>62.77	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Sum	0	2	5	3	2	1	1	21	101	143	26	0	305	227	78
Python		0	0	0	11	3	0	2	11	117	130	30	1	305	222	83
OK		0	0	0	2	10	2	4	9	107	168	3	0	305	174	131

models with geological and geomechanical information can provide scalable 3D framework for geo-engineering applications. However, 3D spatial DTB modelling in complex terrain must have high resolution data to provide an accurate characterisation of subsurface features and a realistic overall depiction to capture any proven or hypothesised subsurface connections. Moreover, due to observed conflict in the results of interpolation algorithms in the manipulation and handling of all requirements, it is not always clear which modelling tools can best reflect DTB surfaces. Here, the validity of the developed models is discussed through comparison with traditional OK and performance analyses using confusion matrix. The models were then ranked according to different statistical metrics.

4.1. Ordinary Kriging

Kriging (Krige, 1951) is one of the most commonly used probabilistic interpolation methods for unknown values of spatial and temporal variables (Dauphiné, 2017), which gives a least square estimate of data (Remy et al., 2011). In this algorithm, the distance-weighted incorporation with the spatial variability is followed to estimate the values of the unsampled locations (Miller et al., 2007). In OK as the most commonly applied Kriging method, the optimal weights for reducing the error variance are determined using the embedded semi-variogram to ensure an unbiased estimator and minimise the estimation variance (Wackernagel, 1995). Using OK, the DTB can be locally estimated based on the neighbourhood locations as

$$\gamma(h) = \frac{1}{2m(h)} \sum_{i=1}^{m(h)} [DTB(x_i) - DTB(x_i + h)]^2 \tag{9}$$

where $\gamma(h)$ is the semi-variogram, and $m(h)$ reflects the number of observation pairs of $DTB(x_i)$ and $DTB\ (x_i+h)$ samples at distance h in locations x_i and x_i+h, respectively. Further, the spatial estimation of DTB for unsampled location, $DTB\ (x_0)$, is then calculated through the linear combination of the observed values, $z_i = Z\ (x_i)$, and weights $w_i\ (x_0)$ $(i = 1, 2, \ldots, N)$:

$$DTB(x_0) = [w_1\ \ w_2 \ldots w_N] \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_N \end{bmatrix} = \sum_{i=1}^{n} w_i(x_0) Z(x_i) \tag{10}$$

where w_i denotes the weight values around the unsampled location. In OK as a linear unbiased estimator, the sum of all the weights is equal to 1.

Therefore, OK in skewed data can better represent estimated error variance than the Kriging (Yamamoto, 2005). Referring to the reasons given above, after a series of analyses, the predicted DTB subjected to training datasets using 12 lags due to better performance was selected and reflected in Fig. 7. Generally, geostatistical and AI techniques can be used as forecasting strategies of subsurface or geological characteristics. However, because of high heterogeneity of spatial distributions in the prediction process, the success of the geostatistical interpolation algorithm (Fig. 7c) was significantly lower than that of AI models (Fig. 7a and b). The development of such modelling process provides an extensive collection of visual data to describe 3D objects. This is an important aspect of the procedure designed in the study, where the 3D objects of each point of the study area can be described using geo-location vectors that serve as a search key in the database. Comparing different models to compensate the weaknesses of the applied techniques can assist in finding a robust tool across the data sources

Table 5
Compared CA, ME and improved progress of optimum models.

Model	CA (%)		ME (%)		Progress (%)	Improvement compared to other models (%)			Rank
	TE	VL	TE	VL		C++	Python	OK	
C++	0.74	0.75	0.26	0.25	1.3		2.7	24	1
Python	0.71	0.73	0.29	0.27	2.7	−2.7		21.9	2
OK	0.58	0.57	0.42	0.43	−1.7	22.6	−21.9		3

(e.g. Chew, 1989; Dickerson et al., 1997; Held, 2001; Domiter and Zalik, 2008; Mei et al., 2013). The differences between the measured DTB and those predicted by OK and ANN models are reflected in Fig. 7d–f.

4.2. Progress control using confusion matrix

Confusion matrix or error matrix (Stehman, 1997) is an intuitive visualised table layout to describe the performance of a model on a set of data. In this matrix, each row and column represent the predicted and actual classes, respectively. Therefore, each array $[a_{ij}]$ shows the number of true labelled instances in a categorised class and thus provides an easy platform to find mislabelled classes. It is also able to show the relations between the individual classified outputs and the true labelled inputs. In practice, a confusion matrix conceptualises the error probabilities of developed models in assigning the individual predicted outputs into the classified input. Accordingly, the best performance is that obtained with zero values, except on diagonal arrays, and thus the better the performance, the better the effectiveness. Using confusion matrix, the classification accuracy (CA) and misclassification error (ME) for the applied models can be quantified (Asheghi et al., 2019). As presented in Tables 4 and 5, the ANN models (C++ and Python) have more true predicted instances (227 and 222, respectively) than the OK (174). The performance of C++ code with 75% correct estimation showed 2.7% and 24% progresses in precious predicted DTB model than Python code and OK, respectively.

4.3. Ranking error metrics

Statistical error metrics are commonly used to evaluate the performance of models. Here, the models were assessed using mean absolute percentage error (MAPE), mean absolute deviation (MAD), RMSE, R^2 and calculated residuals (CR), and indices of agreement (IA). The MAPE is one of the most popular indices used to describe the accuracy and size of the forecasting error, while the MAD shows the variability of datasets using the average distance between each data point and the mean. Using the IA, the compatibility of modelled and observed values is investigated (Willmott, 1984), whereas residual represents a fitting deviation of the predicted value from measured one. Therefore, higher values of IA and R^2 as well as smaller MAPE, CR, MAD and RMSE can be interpreted as a higher predictability level. As shown in Table 6, C++ contributed the best total rank among the three methods. The reason for observed differences in the performance of the programming languages is related to the optimisation methods and initialised condition in the training procedure.

5. Concluding remarks

Due to the variation of subsurface bedrock topography, the production of a more accurate, generalised predictive model for unmeasured DTB areas is of great importance in geo-engineering projects. Such models can be developed and visualised using a learning scheme and finite number of datasets through the trained intelligence system platform. Furthermore, a robust 3D regional framework can reflect the potential subsurface risks associated with the spatial distribution of DTB in geo-engineering applications with a considerably more powerful geological understanding than traditional 2D maps and cross-sections. Moreover, combining the code tools and scientific approaches can assist in the creation of more comprehensive and useful 3D predictive models.

In this study, concerns associated with the generation of a 3D visualised subsurface predictive DTB model were addressed using an automated intelligence training system by means of C++ and Python computer programming environments. To enable more efficient learning, network models composed of different internal characteristics were examined to capture the optimum models.

The lack of data in a part of the study area was compensated using the RDC and 62 new pseudo datasets to extend the region of influence for each soil–rock sounding and decrease the degree of variability in the extrapolation direction. Topologies 3-28-12-1 and 3-25-15-1 were characterised as the optimum predictive DTB topologies for the 2028 data points. Referring to CA (Table 5), the model developed using C++ showed 2.7% and 24% progress in comparison to Python code and the OK technique, respectively. Subsequently, the ranked models using supplementary error indicators reflected almost superiority of the code developed in C++. Accordingly, the 3-28-12-1 topology trained by the CGD algorithm subjected to hyperbolic tangent (Hyt) activation function was selected as the most appropriate structure. The inability of OK to interpolate and handle outlier data was verified with the observed over/underestimated DTB values by using the randomised datasets. It was concluded that OK cannot be presumed to be a representative model for the entirety of the study area, while the developed intelligence models provide significant cost-effective and accurate enough tools in subsurface DTB geo-spatial prediction purposes.

In practice, the dedicated 3D predictive DTB model can present geospatial distribution and the boundary between the overlaid sediments from the hard rocks. This issue can play a significant role in the design phase for the city of Stockholm, which has many ongoing projects in underground openings and transport tunnels. From a geo-engineering point of view, DBT enables the modelling of induced vibration by tunnelling, landslide risk assessment and groundwater. This makes such a DTB model an indispensable tool for decision makers in urban development projects (e.g. building houses, roads, railways and bridges), where substantial land surface processes can be imposed.

Table 6
Results of statistical error criteria in evaluated model performance.

Model	Performance criteria						Ranking of criteria							
	MAPE	RMSE	IA	MAD	R^2	CR domain	MAPE	RMSE	IA	MAD	R^2	CR	Total rank	Sort order
C++	0.28	6.3	0.98	1.03	0.94	[−29, 28]	3	3	3	3	3	3	18	1
Python	0.41	7.84	0.97	1.21	0.9	[−47, 49]	2	2	2	2	2	1	11	2
OK	0.5	9.87	0.95	1.68	0.84	[−54, 19]	1	1	1	1	1	2	7	3

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

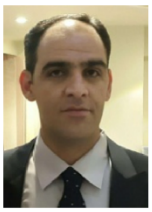
Acknowledgments

This research was funded through the support of the Swedish Transport Administration through Better Interactions in Geotechnics (BIG), the Rock engineering Research Foundation (BeFo) and Tyréns AB, for which the authors express their deepest gratitude. We also wish to thank our colleagues, who provided insight and expertise that greatly assisted us in our research.

References

Abbaszadeh Shahri, A., Renkel, C., Larsson, S., 2020. Artificial intelligence models to generate visualize bed rock level: a case study in Sweden. Model. Earth Syst. Environ. 6, 1509–1528.
Abbaszadeh Shahri, A., Asheghi, R., Khorsand Zak, M., 2021. A hybridized intelligence model to improve the predictability level of strength index parameters of rocks. Neural Comput. Appl. 33, 3841–3854.
Abbaszadeh Shahri, A., 2016. An optimized artificial neural network structure to predict clay sensitivity in a high landslide prone area using piezocone penetration test (CPTu) data: a case study in southwest of Sweden. Geotech. Geol. Eng. 34 (2), 745–758.
Abbott, R.E., Louie, J.N., 2000. Depth to bedrock using gravimetry in the Reno and Carson City, Nevada, area basins. Geophysics 65 (2), 340–350.
Asheghi, R., Abbaszadeh Shahri, A., Khorsand Zak, M., 2019. Prediction of strength index parameters of different rock types using hybrid multi output intelligence model. Arabian J. Sci. Eng. 44 (10), 8645–8659.
Athanasopoulou, A., Bezuijen, A., Bogusz, W., Bournas, D., Brandtner, M., Breunese, A., Burbaum, U., Dimova, S., Frank, R., Ganz, H., Grunicke, U., Jung, H., Lewandowska, A., Nuijten, G., Pecker, A., Psomas, S., Roessler, K., Sciotti, A., Sousa, M.L., Stille, H., Subrin, D., 2019. Standardisation Needs for the Design of Underground Structures. EUR 29633 EN. Publications Office of the European Union, Luxembourg. https://doi.org/10.2760/615209. JRC115352.
Baecher, G.B., 1986. Geotechnical error analysis. Transport. Res. Rec. 1105, 23–31.
Baskan, O., Erpul, G., Dengiz, O., 2009. Comparing the efficiency of ordinary kriging and cokriging to estimate the Atterberg limits spatially using some soil physical properties. Clay Miner. 44, 181–193.
Brett, M., Penny, W., Kiebel, S., 2004. An introduction to random field theory. In: Human Brain Function, second ed. Academic Press, Elsevier, USA, pp. 867–879.
Bengio, Y., 2009. Learning deep architectures for AI. Found. Trends Mach. Learn. 2 (1), 1–127.
Chang, J.R., Chao, S.J., 2009. Applying group method of data handling (GMDH) method to predict depth to bedrock. In: Caldas, C.H., O'Brien, W.J. (Eds.), Computing in Civil Engineering. American Society of Civil Engineers (ASCE). https://doi.org/10.1061/41052(346)27.
Chew, L.P., 1989. Constrained delaunay triangulations. Algorithmica 4, 97–108.
Christensen, C.W., Pfaffhuber, A.A., Anschütz, H., Smaavik, T.F., 2015. Combining airborne electromagnetic and geotechnical data for automated depth to bedrock tracking. J. Appl. Geophys. 119, 178–191.
Clarke, G.K.C., Berthier, E., Schoof, C.G., Jarosch, A.H., 2009. Neural networks applied to estimating subglacial topography and glacier volume. J. Clim. 22, 2146–2160.
Clayton, C.R.I., Smith, D.M., 2013. Advantages and disadvantages of some different methods of exploration. In: Effective Site Investigation. ICE Publishing, London, UK. https://doi.org/10.1680/esi.35058.071.
Cruzes, D.S., Dyba, T., 2011. Research synthesis in software engineering: a tertiary study. Inf. Software Technol. 53, 440–455.
Currry, B., Morgan, P.H., 2006. Model selection in neural networks: some difficulties. Eur. J. Oper. Res. 170 (2), 567–577.
Del Soldato, M., Pazzi, V., Segoni, S., De Vita, P., Tofani, V., Moretti, S., 2018. Spatial modeling of pyroclastic cover deposit thickness (depth to bedrock) in perivolcanic areas of Campania (southern Italy). Earth Surf. Process. Landforms 43 (9), 1757–1767.
Dauphiné, A., 2017. Models of basic structures: networks. In: Geographical Models with Mathematica. ISTE Press, Elsevier, London, UK, pp. 199–224. https://doi.org/10.1016/B978-1-78548-225-0.50011-7.
De Carneiro, C.C., Fraser, S.J., Crósta, A.P., Silva, A.M., de Barros, C.E.M., 2012. Semiautomated geologic mapping using self-organizing maps and airborne geophysics in the Brazilian Amazon. Geophysics 77 (4), K17–K24.
Deutsch, C.V., 1996. Correcting for negative weights in ordinary kriging. Comput. Geosci. 22 (7), 765–773.
Dickerson, M.T., Scot Drysdale, R.A., McElfresh, S.A., Welzl, E., 1997. Fast greedy triangulation algorithms. Comput. Geom. 8 (2), 67–86.
Domiter, V., Zalik, B., 2008. Sweep-line algorithm for constrained delaunay triangulation. Int. J. Geogr. Inf. Sci. 22 (4), 449–462.
Dowd, P.A., Pardo-Iguzquiza, E., 2005. Estimating the boundary surface between geologic formations from 3D seismic data using neural networks and geostatistics. Geophysics 70 (1), 1–11.
Du, Y., Xu, P., Ling, S., Tian, B., You, Z., Zhang, R., 2019. Determining the soil-bedrock interface and fracture-zone scope in the central urban area of the Jinan city, China, by using microtremor signals. J. Geophys. Eng. 16, 680–689.
Erkan, K., 2008. A Comparative Overview of Geophysical Methods. Report No. 488. The Ohio State University, Columbus, Ohio, USA.
Fenton, G.A., 1999. Random field modeling of CPT data. J. Geotech. Geoenviron. Eng. 125 (6), 486–498.
Ghaderi, A., Abbaszadeh Shahri, A., Larsson, S., 2019. An artificial neural network based model to predict spatial soil type distribution using piezocone penetration test data (CPTu). Bull. Eng. Geol. Environ. 78, 4579–4588.
Girosi, F., Jones, M., Poggio, T., 1995. Regularization theory and neural networks architectures. Neural Comput. 7 (2), 219–269.
Gomes, G.J.C., Vrugt, J.A., Vargas, E.A., 2016. Toward improved prediction of the bedrock depth underneath hillslopes: bayesian inference of the bottom-up control hypothesis using high-resolution topographic data. Water Resour. Res. 52, 3085–3112.
Gomes, G.J.C., Vrugt, J.A., Vargas, E.A., Camargo, J.T., Velloso, R.Q., van Genuchten, M.T., 2017. The role of uncertainty in bedrock depth and hydraulic properties on the stability of a variably-saturated slope. Comput. Geotech. 88, 222–241.
Goovaerts, P., 1997. Geostatistics for Natural Resources Evaluation, first ed. Oxford University Press, Oxford, UK.
Held, M., 2001. FIST: fast industrial-strength triangulation of polygons. Algorithmica 30 (4), 563–596.
Hengl, T., Mendes de Jesus, J., Heuvelink, G.B.M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M.N., Geng, X., Bauer-Marschallinger, B., Guevara, M.A., Vargas, R., MacMillan, R.A., Batjes, N.H., Leenaars, J.G.B., Ribeiro, E., Wheeler, I., Mantel, S., Kempen, B., 2017. SoilGrids250m: global gridded soil information based on machine learning. PloS One 12 (2), e0169748.
Hood, S.B., Cracknell, M.J., Gazley, M.F., Reading, A.F., 2019. Improved supervised classification of bedrock in areas of transported overburden: applying domain expertise at Kerkasha, Eritrea. Appl. Comput. Geosci. 3–4, 100001.
Kitterød, N., 2017. Estimating unconsolidated sediment cover thickness by using the horizontal distance to a bedrock outcrop as secondary information. Hydrol. Earth Syst. Sci. 21, 4195–4211.
Krasnopolsky, V., Nadiga, S., Mehra, A., Bayler, E., 2018. Adjusting neural network to a particular problem: neural network-based empirical biological model for chlorophyll concentration in the upper ocean. Appl. Comput. Intell. Soft Comput 7057363, 2018.
Krige, D.G., 1951. A statistical approach to some basic mine valuation problems on the Witwatersrand. J. S. Afr. Inst. Min. Metall 52 (6), 119–139.
Kriegeskorte, N., Golan, T., 2019. Neural network models and deep learning. Curr. Biol. 29 (7), R231–R236.
Kuriakose, S.L., Dekota, S., Rossiter, D.G., Jetten, V.G., 2009. Prediction of soil depth using environmental variables in an anthropogenic landscape, a case study in the Western Ghats of Kerala, India. Catena 79, 27–38.
Lane, J.W., White, E.A., Steele, G.V., Cannia, J.C., 2008. Estimation of bedrock depth using the horizontal-to-vertical (H/V) ambient-noise seismic method. In: Symposium on the Application of Geophysics to Engineering and Environmental Problems Proceedings. Society of Exploration Geophysicists, Tulsa, USA, pp. 490–502.
LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444.
Li, X.Y., Zhang, L.M., Li, J.H., 2015. Using conditioned random field to characterize the variability of geologic profiles. J. Geotech. Geoenviron. Eng. 142 (4), 04015096.
MacCormack, K., Arnaud, E., Parker, B.L., 2018. Using a multiple variogram approach to improve the accuracy of subsurface geological models. Can. J. Earth Sci. 55 (7), 786–801.
Maus, S., 1999. Variogram analysis of magnetic and gravity data. Geophysics 64 (3), 776–784.
Mei, G., Tipper, J.C., Xu, N., 2013. Ear-clipping based algorithms of generating high-quality polygon triangulation. In: Lu, W., Cai, G., Liu, W., Xing, W. (Eds.), Proceedings of the International Conference on Information Technology and Software Engineering, vol. 212. Springer, Berlin, Heidelberg, pp. 979–988.
Mei, G., 2014. Summary of several key techniques in 3D geological modelling. Sci. World J. https://doi.org/10.1155/2014/723832.
Mey, J., Scherler, D., Zeilinger, G., Strecker, M.R., 2015. Estimating the fill thickness and bedrock topography in intermontane valleys using artificial neural networks. J. Geophys. Res. Earth Surf. 120, 1301–1320.
Miller, J., Franklin, J., Aspinall, R., 2007. Incorporating spatial dependence in predictive vegetation models. Ecol. Model. 202 (3–4), 225–242.
Nath, R.R., Kumar, G., Sharma, M.L., Gupta, S.C., 2018. Estimation of bedrock depth for a part of Garhwal Himalayas using two different geophysical techniques. Geosci. Lett. 5, 9.
Pazzi, V., Morelli, S., Fanti, R., 2019. A review of the advantages and limitations of geophysical investigations in landslide studies. Int. J. Geophys. 2019, 2983087.
Pfaffhuber, A.A., Lysdahl, A.O., Christensen, C.W., Vöge, M., Kjennbakken, H., Mykland, J., 2019. Extraction of depth of bedrock from airborn electromagnetic data using artificial neural networks. In: 32nd Annual Symposium on the Application of Geophysics to Engineering and Environmental (SAGEEP 2019). Portland, Oregon, USA.

Remy, N., Boucher, A., Wu, J., 2011. Applied Geostatistics with SGeMS: A User's Guide. Cambridge University Press, Cambridge, UK.
Roesset, J.M., Stokoe, K.H.I.I., Seng, C.H., 1995. Determination of depth to bedrock from falling weight deflectometer test data. Transport. Res. Rec. 1504, 68–78.
Samui, P., Sitharam, T.G., 2011. Application of geostatistical models for estimating spatial variability of rock depth. Engineering 3, 886–894.
Schokker, J., Sandersen, P., de Beer, H., Eriksson, I., Kallio, H., Kearsey, T., Pfleiderer, S., Seither, A., 2017. 3D Urban Subsurface Modelling and Visualisation- a Review of Good Practices and Techniques to Ensure Optimal Use of Geological Information in Urban Planning. TU1206 COST Sub-urban WG2 Report (TU1206-WG2.3-004). COST (European Cooperation in Science and Technology), Brussels, Belgium.
Setiawan, B., Jaksa, M., Griffith, M., Love, D., 2018. Estimating bedrock depth in the case of regolith sites using ambient noise analysis. Eng. Geol. 243, 145–159.
Sitharam, T.G., Samui, P., Anbazhagan, P., 2008. Spatial variability of rock depth in Bangalore using geostatistical, neural network and support vector machine models. Geotech. Geol. Eng. 26, 503–517.
Stehman, S.V., 1997. Selecting and interpreting measures of thematic classification accuracy. Remote Sens. Environ. 62 (1), 77–89.
Stein, M.L., 1999. Interpolation of Spatial Data: Some Theory for Kriging. Springer-Verlag, New York, USA.
Sun, C.G., Kim, H.S., 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, Nat. Hazards Risk 8 (2), 1592–1621.
Sundell, J., Rosen, L., Norberg, T., Haaf, E., 2015. A probabilistic approach to soil layer and bedrock-level modelling for risk assessment of groundwater drawdown induced land subsidence. Eng. Geol. 203, 126–139.
Tacher, L., Pomian-Srzednicki, I., Parriaux, A., 2006. Geological uncertainties associated with 3-D subsurface models. Comput. Geosci. 32, 212–221.
Tetko, I.V., Livingstone, D.J., Luik, A.I., 1995. Neural network studies. 1. Comparison of overfitting and overtraining. J. Chem. Inf. Model. 35 (5), 826–833.
Uzielli, M., Vannucchi, G., Phoon, K.K., 2005. Random filed characterization of strees-normalised cone penetration testing parameters. Geotechnique 55 (1), 3–20.
Viswanathan, R., Samui, P., 2016. Determination of rock depth using artificial intelligence techniques. Geosci. Front. 7, 61–66.
Viswanathan, R., Jagan, J., Samui, P., Porchelvan, P., 2014. Spatial variability of rock depth using simple kriging, ordinary kriging, RVM and MPMR. Geotech. Geol. Eng. 33 (1), 69–78.
Vogl, T.P., Mangis, J.K., Rigler, A.K., Zink, W.T., Alkon, D.L., 1988. Accelerating the convergence of the back-propagation method. Biol. Cybern. 59, 257–263.
Wackernagel, H., 1995. Ordinary kriging. In: Multivariate Geostatistics. Springer, Berlin, Heidelberg, pp. 74–81.
Wei, S., Hengl, T., Mendes de Jesus, J., Hua, Y., Dai, Y., 2016. Mapping the global depth to bedrock for land surface modeling. J. Adv. Model. Earth Syst. 9 (1), 65–88.
Willmott, C.J., 1984. On the evaluation of model performance in physical geography. In: Gaile, G.L., Willmott, C.J. (Eds.), Spatial Statistics and Models. Springer, Dordrecht, Netherlands, pp. 443–460.
Wolpert, D.H., Macready, W.G., 1997. No free lunch theorems for optimization. IEEE Trans. Evol. Comput. 1 (1), 67–82.
Yan, F., Shangguan, W., Zhang, J., Hu, B., 2020. Depth-to-bedrock map of China at a spatial resolution of 100 meters. Sci. Data 7 (1), 2.
Yamamoto, J.K., 2005. Comparing ordinary kriging interpolation variance and indicator kriging conditional variance for assessing uncertainties at unsampled locations. In: Dessureault, S., Ganguli, R., Dwyer, V.J.K. (Eds.), Application of Computers and Operations Research in the Mineral Industry. A.A. Balkema, Leiden, Netherlands.
Yuan, Y., Rosasco, L., Caponnetto, A., 2007. On early stopping in gradient descent learning. Constr. Approx. 26 (2), 289–315.
Zhang, T., Yu, B., 2005. Boosting with early stopping: convergence and consistency. Ann. Stat. 33 (4), 1538–1579.



Abbas Abbaszadeh Shahri received his BSc (2000) and MSc (2002) degrees in Mining Engineering and PhD (2009) degree in Geophysics with focus on earthquake geotechnical engineering. From 2013 to 2015, he was postdoctoral at Uppsala University, Sweden. Since 2015, he has worked as a researcher in Division of Soil and Rock Mechanics, Faculty of Civil and Architectural Engineering, KTH Royal Institute of Technology, Sweden. His academic career as lecturer and then senior lecturer in Islamic Azad University (IAU) of Iran started at 2002. He benefits from several years of experience as geotechnical engineer in civil, construction and mining industries. Currently, he works in Division of Rock Engineering at Tyrens-Sweden in collaboration with KTH Royal Institute of Technology on developing new approaches for intelligent computer vision of subsurface models and uncertainty analysis. His research interests include developing soft computing and hybrid approaches, nonlinear site response analyses, geo-hazard modelling and computer vision application in geo-engineering purposes.

B.2 A new approach to uncertainty analysis using automated predictive deep learning in groundwater (accepted by Natural Resources Research)

Natural Resources Research

A new approach to uncertainty analysis using automated predictive deep learning in groundwater

--Manuscript Draft--

Manuscript Number:	NARR-D-21-01106
Full Title:	A new approach to uncertainty analysis using automated predictive deep learning in groundwater
Article Type:	Original Research
Keywords:	Uncertainty quantification; automated modelling, ARDCW, ground water, Sweden
Corresponding Author:	Abbas Abbaszadeh Shahri, Ph.D Islamic Azad University of Hamedan Hamedan, Hamedan IRAN, ISLAMIC REPUBLIC OF
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Islamic Azad University of Hamedan
Corresponding Author's Secondary Institution:	
First Author:	Abbas Abbaszadeh Shahri, Ph.D
First Author Secondary Information:	
Order of Authors:	Abbas Abbaszadeh Shahri, Ph.D Chunling Shan, PhD Stefan Larsson, PhD
Order of Authors Secondary Information:	
Funding Information:	Trafikverket (BIG A2019:17) — Not applicable BeFo (BeFo 415) — Not applicable
Abstract:	Uncertainty quantification (UQ) is an important benchmark to assess the performance of artificial intelligence (AI) and particularly deep learning ensembled-based models. However, the ability to estimate the UQ using current AI -based methods is not only limited in terms of computational resources but also requires changes to topology and optimization processes, as well as multiple performances to monitor the model instabilities. From both a geoengineering and societal perspective, the predictive groundwater table (GWT) model presents an important challenge, where a lack of assessed UQ limits the validity of findings and may undermine science-based decisions. To overcome and address these limitations, a novel ensemble, the automated random deactivating connective weights approach (ARDCW), was presented and applied on retrieved geographical locations of GWT data from a geoengineering project in Stockholm, Sweden. In this approach, the UQ was estimated via a combination of several derived ensembles from a fixed optimum topology subjected to randomly switched off weights, which allows predictability with one forward pass. The process was developed and programmed to provide trackable performance in a specific task and access to a wide variety of different internal characteristics and libraries. A comparison of performance with the Monte Carlo dropout (MCD) and quantile regression (QR) using both computer vision and control task metrics showed significant progress in the ARDCW . This approach does not require changes in the optimization process and can be applied to already trained topologies in a way that outperforms other models.

Powered by Editorial Manager® and ProduXion Manager® from Aries Systems Corporation

Manuscript with titel page

Click here to view linked References

Click here to access/download;Manuscript;uncertainity quantification-English edited-NARR(17-09-2021).docx

A new approach to uncertainty analysis using automated predictive deep learning in groundwater

Abbas Abbaszadeh Shahri
Division of rock engineering, Tyrens AB, Stockholm, Sweden
Johan Lundberg AB, Uppsala, Sweden
shahri@kth.se, orcid.org/0000-0002-3832-572X
Chunling Shan
Division of rock engineering, Tyrens AB, Stockholm, Sweden
Division of Soil and Rock Mechanics, KTH Royal Institute of Technology, Stockholm, Sweden
cshan@kth.se
Stefan Larsson
Division of Soil and Rock Mechanics, KTH Royal Institute of Technology, Stockholm, Sweden
Stefan.larsson@byv.kth.se, orcid.org/0000-0001-9615-4861

Abstract

Uncertainty quantification (*UQ*) is an important benchmark to assess the performance of artificial intelligence (*AI*) and particularly deep learning ensembled-based models. However, the ability to estimate the *UQ* using current *AI*-based methods is not only limited in terms of computational resources but also requires changes to topology and optimization processes, as well as multiple performances to monitor the model instabilities. From both a geoengineering and societal perspective, the predictive groundwater table (*GWT*) model presents an important challenge, where a lack of assessed *UQ* limits the validity of findings and may undermine science-based decisions. To overcome and address these limitations, a novel ensemble, the automated random deactivating connective weights approach (*ARDCW*), was presented and applied on retrieved geographical locations of *GWT* data from a geoengineering project in Stockholm, Sweden. In this approach, the *UQ* was estimated via a combination of several derived ensembles from a fixed optimum topology subjected to randomly switched off weights, which allows predictability with one forward pass. The process was developed and programmed to provide trackable performance in a specific task and access to a wide variety of different internal characteristics and libraries. A comparison of performance with the Monte Carlo dropout (*MCD*) and quantile regression (*QR*) using both computer vision and control task metrics showed significant progress in the *ARDCW*. This approach does not require changes in the optimization process and can be applied to already trained topologies in a way that outperforms other models.

Keywords: Uncertainty quantification; automated modelling, *ARDCW*, ground water, Sweden

Declarations

Funding: This work was supported by Rock engineering Research Foundation (BeFo) (Grant number BeFO 415), and Swedish Transport Administration through Better Interactions in Geotechnics (BIG) (Grant number BIG A2019:17).
Competing interests: The authors declare that they have no competing interests.
Conflicts of interest: The authors have no known conflicts of interest to declare associated with this publication.
Availability of data and material: Not applicable. The data belongs to Tyrens and will not be shared.
Code availability: Not applicable.
Authors' contributions: All authors whose names appear on the submission, made substantial contributions to the conception (All authors), methodology (Abbaszadeh Shahri), code development (Abbaszadeh Shahri, Shan), data acquisition (Shan, Larsson), analysis (Abbaszadeh Shahri, Shan), and result interpretation (All authors). The first draft was written by Abbaszadeh Shahri, supplemented by Shan, and critically for important intellectual content was commented and reviewed by Larsson and then English editing by native experts. All authors read and approved the final manuscript.
Acknowledgments: The authors would like to express their special gratitude to Swedish Transport Administration through Better Interactions in Geotechnics (BIG), the Rock engineering Research Foundation (BeFo) and Tyréns AB for supporting this research. We also wish to thank our colleagues, who provided insight and expertise that greatly assisted us in our research.
Compliance with Ethical Standards: Not applicable

Article Highlights

- A novel state-of-the-art ensemble automated deep learning-based uncertainty quantification approach was proposed.
- Analytical interpretations on developed groundwater predictive models in Stockholm, Sweden were dedicated.
- Significant progress in compare with Monte Carlo dropout and quantile regression was observed.

Highlights:

- A novel state-of-the-art ensemble automated deep learning-based uncertainty quantification approach was proposed.
- Analytical interpretations on developed groundwater predictive models in Stockholm, Sweden were dedicated.
- Significant progress in compare with Monte Carlo dropout and quantile regression was observed.

A new approach to uncertainty analysis using automated predictive deep learning in groundwater

Abbas Abbaszadeh Shahri[1,2], Chunling Shan[1,3], Stefan Larsson[3]

[1]Division of Rock Engineering, Tyréns AB, Stockholm, Sweden

[2]Johan Lundberg AB, Uppsala, Sweden

[3] Division of Soil and Rock Mechanics, KTH Royal Institute of Technology, Stockholm, Sweden

Abstract

Uncertainty quantification (*UQ*) is an important benchmark to assess the performance of artificial intelligence (*AI*) and particularly deep learning ensembled-based models. However, the ability to estimate the *UQ* using current *AI*-based methods is not only limited in terms of computational resources but also requires changes to topology and optimization processes, as well as multiple performances to monitor the model instabilities. From both a geoengineering and societal perspective, the predictive groundwater table (*GWT*) model presents an important challenge, where a lack of assessed *UQ* limits the validity of findings and may undermine science-based decisions. To overcome and address these limitations, a novel ensemble, the automated random deactivating connective weights approach (*ARDCW*), was presented and applied on retrieved geographical locations of *GWT* data from a geoengineering project in Stockholm, Sweden. In this approach, the *UQ* was estimated via a combination of several derived ensembles from a fixed optimum topology subjected to randomly switched off weights, which allows predictability with one forward pass. The process was developed and programmed to provide trackable performance in a specific task and access to a wide variety of different internal characteristics and libraries. A comparison of performance with the Monte Carlo dropout (*MCD*) and quantile regression (*QR*) using both computer vision and control task metrics showed significant progress in the *ARDCW*. This approach does not require changes in the optimization process and can be applied to already trained topologies in a way that outperforms other models.

Keywords: Uncertainty quantification; automated modelling, *ARDCW*, ground water, Sweden

1. Introduction

Groundwater table (*GWT*) is not only one of the Nation's most important natural resources but also a nonlinear time-dependent concern in many geoengineering projects (Hu & Jiao, 2010; Parry et al., 2014; Tang et al., 2017;

Salvo et al., 2020). Modelling of this concern is applied to conceptualize and integrate knowledge on the natural and engineering disciplines to address a range of issues in different spatial and temporal scales (Bizhanimanzar et al., 2019; Mohammadi, 2009; Yeh et al., 2015). However, due to a number of embedded nonlinear complexities in the system, modelling of *GWT* is an inherently difficult task. Moreover, the sources of uncertainties related to model parameters, underlying conceptual assumptions, structure of model, geological conditions, observed spatial data (i.e., lack of data on natural variability), and steady- state *GWT* should be considered (Beven & Binley, 1992; Hauser et al., 2017; Stedinger et al., 2008; Yeh et al., 2015). As a result, the computer code models have primarily been developed over the simplified assumptions using long term time series through different approaches such as statistical techniques, approximation methods and numerical analyses (Yan et al., 2018; Rushton, 2003). Therefore, evaluating the robustness and accuracy performance of a predictive *GWT* model using uncertainty quantification (*UQ*) analysis is often required. Nowadays, alternative modern computational artificial intelligent techniques (*AIT*) lead to systematic *UQ* analysis that can be effectively applied to generate a *GWT* model (Guillaume et al., 2016; Yin et al., 2021; Chen et al., 2020; Sahoo & Russo, 2017; Wunsch et al., 2020).

Uncertainty can simply be described as knowledge situations involving imperfect or unknown information (Gärdenfors & Sahlin, 1982). To quantify the uncertainty, three sources including physical variability of equipment, data, and model error should be considered (Barford, 1985; Kennedy & O'Hagan, 2001). In the modelling process, the *UQ* methods are described in the context of different factors such as input variabilities, assumptions and approximations, measurement errors, and sparse and imprecise data (Yan et al., 2015). In the literature, statistical techniques, sensitivity analyses, Taylor series approximation, Monte Carlo simulation, numerical estimation, and probability distributions are the most commonly used methods to assess the *UQ* (e.g. Cox & Baybutt, 1981; Glimm & Sharp 1999; Bárdossy & Fodor, 2001; Uusitalo et al., 2015; Asheghi et al., 2020; Elam & Rearden, 2017). However, when using these techniques for value-based judgements, all sources of uncertainty may not be quantifiable (Morgan & Henrion, 1990). The taxonomy of different methods applied to estimate the *UQ* are given in Figure1.

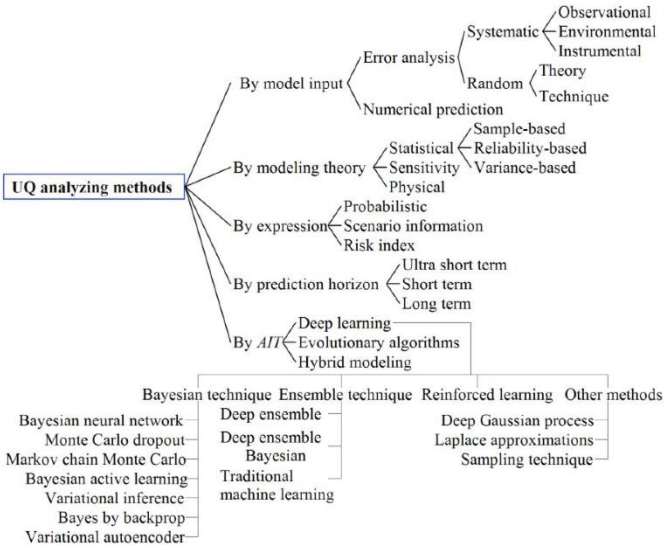


Figure 1. Taxonomy of the *UQ* methods

In statistical techniques, the *UQ* is usually evaluated through the estimated variance and determined confidence limits by assuming normally distributed error (e.g. Eisenhart et al., 1983; Cacuci & Ionescu-Bujor, 2004; Jiang et al., 2018; Zhang et al., 2020). Since the quantification is most often performed by statistical indices (e.g. mean, median, population quantiles) due to the limited sample size, estimations must be provided with the associated confidence intervals (Geffray et al., 2019). Therefore, due to underlying foundations, such as measurement errors, material properties, unknown design demand models, and stochastic environments, the applicability of statistical techniques is limited (Taper & Ponciano, 2016). Taylor series approximation can be used to determine theoretical error bounds but is very complex to derive and also tends to quickly increase the risk of error (Barrio et al., 2011). Monte Carlo is mainly applied to explain the probability density function but is a complex process and requires a large number of simulations and computational resources when the dataset is large. Therefore, it only provides statistical estimates of results, not exact figures (Atanassov & Dimov, 2008). In numerical *UQ* methods, the results, which are simulated by computer codes, inherently involve error or uncertainty. Such drawbacks can be addressed in part by defining the error magnitude or bound the error in a given simulation (Freitas, 2002). By using probability distributions, accurate information on the uncertainty can be achieved, but infinite extension along either side of the most probable region creates a sophisticated situation that is difficult to interpret (Kabir et al., 2018). Moreover, quantitative measures may bias the description of uncertainty towards the more computational components of the assessment (Bárdossy & Fodor, 2001). Thereby, the concepts and

77 methods used for *UQ* make it difficult to communicate the results effectively. This implies that the analysis of
78 the *UQ* through ensembling methods is one of the main benchmarks to assess the performance of complex
79 systems (Vrugt & Robinson, 2007; Zhang et al., 2021), especially in geoengineering and *GWT* predictive
80 models, which often suffer from inadequate experimental or data (e.g. Whitman, 2000; Huber, 2016; Chahbaz et
81 al., 2019; Sepulveda & Doherty, 2015; Li et al., 2018; Wu & Zeng, 2013). Furthermore, computing the
82 contribution of each error component on total uncertainty is difficult and leads to extreme weighting schemes
83 with applicability that is often questionable. Therefore, no unified applicable methodology exists for combining
84 uncertainties. Since these methods typically yield an estimate of the total variance of measurement values,
85 reliance on statistical variance can produce misleading results that are not readily applicable (e.g. Goodman,
86 1960; Yager, 1996; Farrance & Frenkel, 2012; Borgonovo, 2006).
87 Due to the revolution of different types of *AITs* in recent years, and deep neural networks (*DNNs*) in particular,
88 these mechanisms have shown impressive state-of-the-art performance on a wide variety of engineering tasks
89 dealing with challenging scientific data analysis and *UQ* problems (e.g. Vrugt & Robinson, 2007; Abbaszadeh
90 Shahri et al., 2020; Chahbaz et al., 2019; Hernandez & Lopez, 2020). However, despite the rapid emergence of
91 *AIT* and the subsequent *DNN*-based *UQ* analysis (e.g. Gal & Ghahramani, 2016; Krzywinski & Altman, 2013;
92 Zhu et al., 2019), the performance of these methods may suffer from a series of shortcomings, such as
93 complexities arising from topology and hyperparameter choices (Asheghi et al., 2020), computational cost for
94 multiple trainings, and weak performance due to model instabilities (Foong et al., 2019). Furthermore, since the
95 *DNNs* tend to produce overconfident predictions, accurate outcome, especially in out-of-domain data, is an
96 important issue (Loquerico et al., 2020; Klotz et al., 2021). Moreover, current *DNN*-based *UQ* methods, such as
97 Monte Carlo dropout (*MCD*) (Gal & Ghahramani, 2016) or quantile regression (*QR*) (Weerts et al., 2011),
98 behave differently at training and inference time due to implemented changes in topology and optimisation
99 processes.
100 When using the *MCD*, the *UQ* is estimated through a set of trained topologies, which are subjected to randomly
101 removed neurons, while in the *QR*, the *UQ* is captured through a conditional quantile of a dependent variable
102 without assuming any specific conditional distribution. However, these methods typically ignore prior
103 knowledge about the data and consequently tend to make assumptions that lead to oversimplification and thus
104 underestimate uncertainty (Waldmann, 2018). Such concerns imply that developing new schemes or approaches
105 to present robust *UQ* estimation is still a crucial challenge with a tremendous potential for application in
106 complex geoengineering problems, particularly for *GWT* pattern modelling.

In Sweden, the *GWT* is an important part of sustainable ecosystems and for human water consumption. However, the *GWT* pattern can be affected by geoengineering projects in combination with changes in the natural variability and the climate. Due to the significant benefits offered by an adequately accurate predictive model and the approved efficiency of *AIT* in providing more precise solution than many simulation processes, this paper aims to address and then examine a new systematic *UQ* analysis approach through the development of an automated procedure.

As a result, a novel state-of-the-art ensemble automated random deactivating connective weights approach (*ARDCW*) was proposed. In this framework, instead of dropping the neurons, the connective weights between layers of the identified optimum topology are randomly deactivated. To overcome the overfitting problem and avoid being trapped in local minima, the *ARDCW* uses several embedded internal and nested loops to monitor all the topologies based on different hyperparameters. This approach was then experimentally applied on 244 sets of *GWT* geolocation data in an urbanized area of Stockholm. Due to the use of the optimum topology, *ARDCW* led to a state-of-the-art performance of *UQ* analysis in which the produced multiple predictions can be interpreted in terms of average errors. In comparison to other methodologies such as *MCD* and *QR*, the evaluated and compared performance of *ARDCW* showed superior capability in the prediction of spatial *GWT*.

2. Study area and data source

The study area, which contained more than 300 investigated *GWT* wells, encompasses 20 km of an ongoing highway project in south Stockholm, Sweden. Among the monitored *GWT* data, 244 points were screened and compiled from continuous recorded intervals from 19 to 25 September 2020. The location of the study area with respect to the generated long-term map of spatial monitored *GWTs* for the entirety of Sweden is presented in Figure 2a, which was retrieved from Geological Survey of Sweden. As shown in Figure 2b, in the northwest of the study area, Mälaren Lake controls the *GWT* toward the downstream, while the coverage in the northeast consists of bedrock outcrop incorporated with intermediate soil-filled valleys. This implies that the topographical variation in the surface of the bedrock dominantly controls the groundwater flows. The *GWT* in the high terrain Masmo hills, located in the middle of the study area (Figure 2b), mostly follows the surface topography, but without connecting to any nearby aquifers. To the southwest (Figure 2b), homogenous bedrock and a uniform *GWT* pattern can be observed. However, in the southeast of study area, the bedrock is highly heterogenic without uniform *GWT* levels due to the influence of cracking on hydraulic connectivity. An integrated spatial distribution of ground water aquifers with the *GWT* data over the entire study area is presented in Figure 2c.

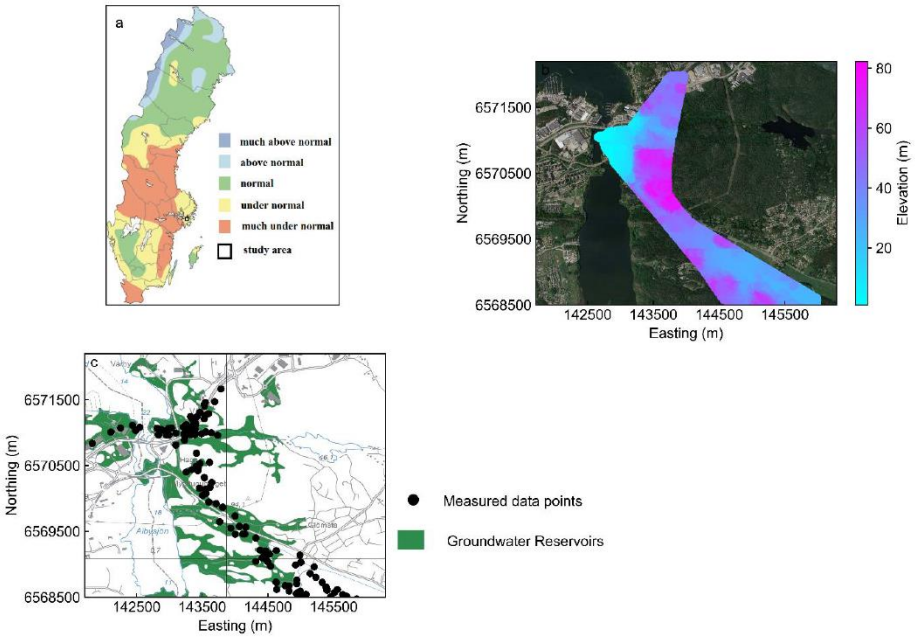


Figure 2: Distribution of *GWT* in Sweden (a), the overlaid digital elevation map and satellite image of the study area (b), the mapped aquifers and monitored *GWT* data in the area (c) (Figure a and c were retrieved from Geological Survey of Sweden)

3. A summary of *UQ* analysis using *AIT*

Estimating the *UQ* as one of the main challenges in *AIT* needs extensive exploration; however, currently, the focus is often on giving a best estimate as defined by a loss function. Referring to Figure 1, Bayesian techniques (MacKay, 1992) are the main *AIT*-based framework for *UQ* analyses. However, they are often computationally slow and difficult to train because they are traditionally formalized through parametric probability distributions of network activations and weights (Hernandez-Lobato & Adams, 2015). Sampling technique (Gal & Ghahramani, 2016) is another method used for *UQ* estimation, but due to the inability, explicitly modelling generates overconfident predictions (Hernandez & Lopez, 2020). Since an input with large noise provides wider model uncertainty than the same input with lower noise, the risk of underestimated uncertainties is increased because sampling-based methods generally disregard the relationship between data and model uncertainty (Kabir et al., 2018; Bárdossy & Fodor, 2001; Gal & Ghahramani, 2016).

Despite the significant performance of the Markov chain Monte Carlo (*MCMC*) in predicted *UQ*, it is mostly appropriate for small networks and thus computationally expensive for large *DNNs* (Quinonero-Candela et al., 2006). However, this shortcoming can be overcome using stochastic gradient *MCMC* (Chen et al., 2014), which only needs to estimate the gradient on small sets of mini batches. Since the *MCMC* requires a significant amount of time to converge on a desired distribution, the sufficient number of iterations is unknown (Neal, 2012). Therefore, to provide a relatively faster approximate Bayesian solution, different methods such as variational inferences, assumed density filtering, expectation propagation and the stochastic gradient Langevin diffusion technique have been introduced (Lakshminarayanan et al., 2017). However, the posterior *UQ* in the variational method is typically underestimated rather than overestimated in the expectation propagation technique.

QR is a type of powerful regression analysis tool that estimates the conditional median (or other quantiles) of the response predicted values, where the quantile itself is a parameter for the loss function. This implies that for each quantile, the training process should be analyzed through individual asymmetric weighting (Yang et al., 2016). Accordingly, the trained model, when subjected to different quantiles, generates different bounds that picks out the conditional quantiles away from the median, that is, the 95% prediction intervals can be found when subjected to the quantiles 0.025 and 0.975. The mean square error, as the most commonly used loss function, would then correspond to the Gaussian distribution. Accordingly, the mode (peak) of this distribution corresponds to the mean parameter, and *DNNs* predict the mean value of the output, which may have been noisy in the training set.

Furthermore, the dependency of the quality of prediction to computational complexity, degree of approximation and correctness of the prior distribution imply that the estimated *UQ* using these methods cannot be guaranteed to provide underlying beliefs (Hirschefeld et al., 2020).

The *UQ* can be estimated using ensemble-based methods, where the inputs are trained and passed through multiple networks (Figure3A). Greater accuracy can be achieved by using such methods than any individual model because the predictions vary across multiple runs of a set of models $M = \{M_1, M_2, ..., M_n\}$ rather than training a single M_i. However, the primary drawback of any ensemble-based *UQ* method is the increased training time, depending on the size of the ensembles. As presented in Figure 3, in addition to ensemble methods, the *UQ* can be estimated through mean-variance, distance-based strategies and union-based strategies (Hirschefeld et al., 2020; Angiulli & Fassetti, 2020; Zhang et al., 2020). The problem with training Bayesian neural networks for large data sets can be solved by using mean-variance estimation, where the output layer is modified and trained using a negative log likelihood loss to predict both the mean and variance of the interested input (Hirschefeld et

al., 2020). Through distance-based methods, the minimum distance between each interested prediction from its nearest neighbours in the training set is interpreted as uncertainty. The larger the distance between the training sets and predictions, the higher the error and thus the greater the uncertainty (Angiulli & Fassetti, 2020). However, significant sensitivity of the calculated distance to outliers is the major drawback of this *UQ* estimator. The union-based method (Huang et al., 2015) is a combined confidence estimator, where the output of the trained neural network is fed to another model to provide an ensemble model and thus calculate uncertainty into the task-specific latent space.

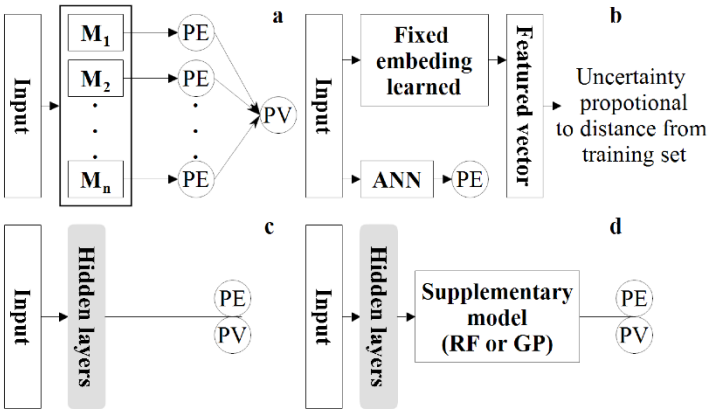


Figure 3. Simplified schematic of different strategies for *UQ* analysis: (a) ensemble-based, (b) distance-based, (c) mean variance and (d) union-based (*PE*: Predicted error; *PV*: Predicted variance)

The *MCD* (Gal & Ghahramani, 2016) is one of the most widely used *UQ* analysis methods and can even be implemented on an already trained model. In this procedure, any trained topology with dropouts is used to prevent overfitting and can thus be interpreted as an approximate inference of the posterior weights. Accordingly, the average of multiple predictions for the analysis of distributions is considered to calculate meaningful variance. In a fixed topology during the *MCD* process, all in/out connective weights and biases of dropped neurons are ignored. This implies that in each dropout, the data is trained using a different topology, while none of these connectives participate in the prediction process nor will they be updated. Therefore, the dependency of the estimated *UQ* on the dropped neuron and internal hyperparameters is the main drawback of *MCD*. Moreover, due to the training of several topologies, the produced approximations can be categorized in ensemble techniques, which significantly enable improved overall performance (Dietterich, 2000). In comparison to the Bayesian approach (MacKay, 1992), which attempts to average and thus find the single best

model (or parameters), ensembles combine the topologies to obtain a more powerful model. Thereby, ensembles potentially provide a complementary source for predictive *UQ* estimation. Referring to Figure 4, the *UQ* through the *MCD* is predicted over multiple networks at the test time, where a model with n neuron can provide a collection of 2^n possible topologies as an ensemble combination (Srivastava et al., 2014).

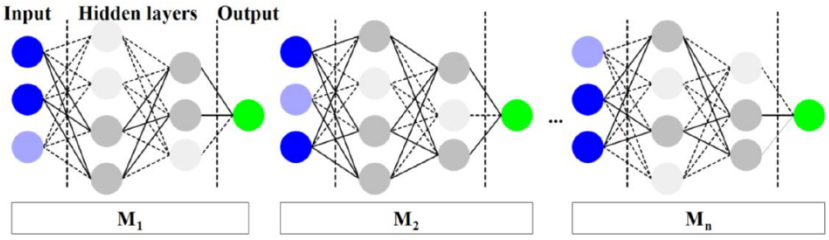


Figure 4. Overview of *MCD* modelling process mimicking ensemble-based method (highlighted and dash lines correspond to dropped neurons and weights)

4. Proposed *UQ* approach

In recent years, estimating the *UQ* using different *AITs* has increased, as the provided models can offer simple but precise solutions in many engineering simulation problems. A systematic *UQ* analysis can be applied by considering a combination of several components, including data, model structure and parameter values. This implies that the *UQ* as a function should be conceptualized in a concrete mathematical sense and thus be amenable for programming. Since noise will be produced in each random dropping, which will affect results, therefore, mathematically, the performance and predictability of the *DNN* structure should always be evaluated with the original optimum model. By using the dropout concept integrated with the optimum topology, the ensemble automated random deactivating connective weights (*ARDCW*) approach was proposed. The *ARDCW* is solely focused on randomly switched off weights, not neurons, where the remaining weights are forced to participate in learning processes and assist in decreasing the overfitting. Accordingly, this approach uses an optimum trained topology capable of performing a given task even when the weights are randomly sampled. This implies that the training of multiple different topologies is avoided, as the uncertainty will be estimated by changing the internal assigned weights of a fixed optimum model.

In a *DNN* topology, all the m neurons of the layer are associated with the probability p, otherwise is dropped (set to zero), with probability $1\text{-}\ p$ (Hinton et al., 2012). According to probability theory, the expected value of a random variable ($E[X]$) is a generalization of the weighted average with probabilities of p_i as:

$$E[X] = \sum_{i=1}^{n} x_i p_i \quad where \ \sum_{i=1}^{n} p_i = 1 \qquad (1)$$

Mathematically, for a constant random variable, $E[X=c]=c$. Therefore, $E[X]$ with equiprobable outcomes $\{c_1, c_2, \ldots, c_n\}$ expresses the arithmetic mean of the terms with the probabilities $P(X=c_i)$. Using this concept, the impact of dropped weights ($\hat{w}_{ij}$) in *UQ* can be described with mean field theory (Kadanoff, 2009) as:

$$\hat{w}_{ij} = \begin{cases} w_{ij} & with\ p(c) \\ 0 & otherwise \end{cases} \quad (2)$$

Where $P(c)$ denotes the probability of keeping a weight, whereby the implemented neurons are kept with the probability p and otherwise will drop (0) with probability $1\text{-}p$. This implies that the *UQ* analysis can be generalized by dropping the connective weights with probability $1\text{-}p$ rather than the neurons (Wan et al., 2013). Consequently, in a topology with h hidden layers, $l \in \{1, 2, \ldots, h\}$, where $Z^{(l)}$ and $y^{(l)}$ denote the vector of inputs into and output from layer l, the feed-forward process is described as:

$$z_i^{(l+1)} = w_i^{(l+1)} y^l + b_i^{(l+1)} \quad (3)$$

$$y_i^{(l+1)} = f\left(z_i^{(l+1)}\right) \quad (4)$$

Referring to neural network topology, the activity of dropout weight of the i^{th} neuron in h^{th} hidden layer (a_i^h), and consequently the corresponding output (O_i^h), can be expressed by:

$$a_i^h = \sum_{l<h} \sum_j w_{ij}^{hl} \delta_{ij}^l a_j^l \qquad with\ a_j^0 = X_j \quad (5)$$

$$O_i^h = f\left(a_i^h\right) = f\left(\sum_{X<h} \sum_j w_{ij}^{hl} \delta_{ij}^l a_j^l\right) \qquad with\ O_j^0 = X_j \quad (6)$$

Where; δ_{ij}^l is a gating 0-1 Bernoulli variable with $P(\delta_{ij}^l=1)= p_j^l$.

Therefore, in dropout, the output of layer l using element-wise product (*) is then presented through:

$$z_i^{(l+1)} = w_i^{(l+1)}\left(p_i^l * y^l\right) + b_i^{(l+1)} \quad (7)$$

Accordingly, the expectation of the activity of all neurons for a fixed input vector taken over all possible δ_{ij}^l variables, and thus the possible ensemble models, can be achieved through:

$$E(a_i^h) = \sum_{l<h} \sum_j w_{ij}^{hl} p_j^l E(a_j^l) \qquad for\ h > 0 \quad (8)$$

The dropout is not applied on a fixed input that produces a constant output in the l^{th} layer (O_j^l), thus, the dropout on the h^{th} hidden layer is then presented with:

$$Var\left(a_i^h\right) = \sum_{l<h} \left(w_{ij}^{hl}\right)^2 \left(O_j^l\right)^2 p_j^l (1 - p_j^l) \quad (9)$$

Where; Var is the variance.

Therefore, the expectation of dropout gradient as a random variable is the regularized ensemble error associated with all possible models (E_{ens}) and is expressed by:

$$E\left(\frac{\partial E_D}{\partial w_i}\right) = \frac{\overbrace{\partial\left(0.5\left(t-\sum_{i=1}^n p_i w_i X_i\right)^2\right)}^{regularized\ E_{ens}}}{\partial w_i} + w_i p_i (1-p_i) X_i^2 \frac{\overbrace{\partial\left(0.5\left(t-\sum_{i=1}^n p_i w_i X_i\right)^2\right)}^{regularized\ E_{ens}}}{\partial w_i} + w_i X_i^2 Var(\delta_i) \quad (10)$$

$$E[\text{dropout gradient}] = E_{ens} + 0.5 \sum_{i=1}^{n} w_i^2 X_i^2 Var(\delta_i) \quad (11)$$

Where; E_D denotes the error of the network with dropout.

Referring to Eqs. 8 and 11, additional layers corresponding to real-valued output for error or variance can be attached to the optimum topology. This implies producing a multi-objective model, where the output of error and variance are also predicted using the same fixed optimum *DNN* with an added layer(s).

Referring to presented block procedure of the proposed *ARDCW* (Figure 5), it does not require any changes in optimum topology and thus can even be used for already trained models. In this process, the optimum topology was captured through the proposed automated strategy by Abbaszadeh Shahri et al., (2021a). The performance of the optimum model is then monitored through the randomly sampled deactivated weights using an automated dropout procedure for the achieved database. Since all deactivations occur for the one optimum topology, the *ARDCW* captures different models and as shown in Figure 3a, it can be interpreted as an ensemble-based Bayesian approximation of the Gaussian process probabilistic model. Accordingly, the inadvertent tuning on the model is addressed through a regularization strategy, which needs a robust implementation that accounts for floating-point stability and reproducibility in geoengineering-related problems. The minimum number of sampled weights in this study is one, while the maximum number of desired droppings was set at 50% of the whole connective weights. However, selecting this rate for dropout is flexible and user defined, but the dependency on the topology and network type (shallow or deep) that overfits the training data should be considered. Moreover, the greater the number of deactivations, the more examined ensembles, which thus requires more analysis time.

As an advantage, in each examination, the model is trained using a constant topology but with a different set of randomly deactivated weights in the layers. Therefore, the outcome can be considered as an averaging ensemble of many different models trained on one batch of data only. The *ARDCW* can be regularized to calculate the mean and variance of randomized data and can then predict unlabelled data. Furthermore, using the optimum model will minimize the problem of overfitting, while the overall performance of the *DNN* topology due to randomly switched off weights in each layer becomes less sensitive.

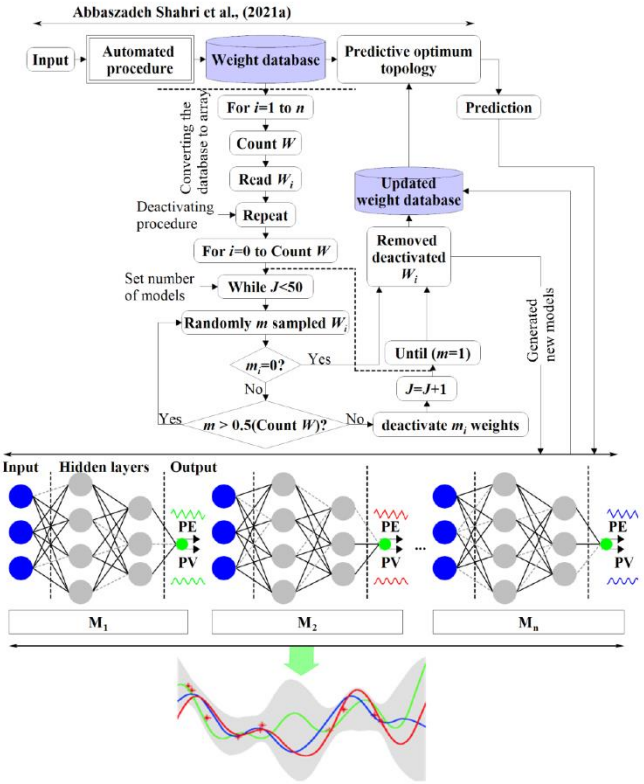


Figure 5. Block diagram of the proposed *ARDCW* approach for *UQ* modelling

5. Result analysis of experimental application

According to Figure 5, when applying 65% of the compiled *GWT* data, a 3-32-8 topology subjected to hyperbolic tangent activation function trained with quasi-Newton algorithm was considered as the optimum, which was then retrained using two extra predicted outputs, including the error and variance. Considering the y and $\hat{y}$ as the target and model output, respectively, the error and variance are predicted through $(y\text{-}\hat{y})$ and $(y-\hat{y})^2$. Referring to Figure 4, a series of models were retrained using the updated dropout weights database (Figure 5) subjected to the fixed defined custom loss function in optimum topology for each mini batch. To visualize the range of predicted distribution, the results of the randomly deactivated weight using K-fold validation for a series of dropout test sets are presented in Figures 6 and 7. Accordingly, Figure 6a shows the variation of the predicted *GWT* from all the K-fold validations using *ARDCW*, where the uncertainties of each point can then be estimated from these outcomes. Subsequently, the errors between the observed and predicted *GWT* in each K-fold

validation are reflected in Figure 6b, which shows high error values and thus high uncertainties at data points. Figure 6c depicts the variance of the errors as described where uncertainties are high. According to these outcomes, *PE* and *PV* in the analyzed *UQ*, when subjected to different random dropouts are more preferred due to lower variance estimators.

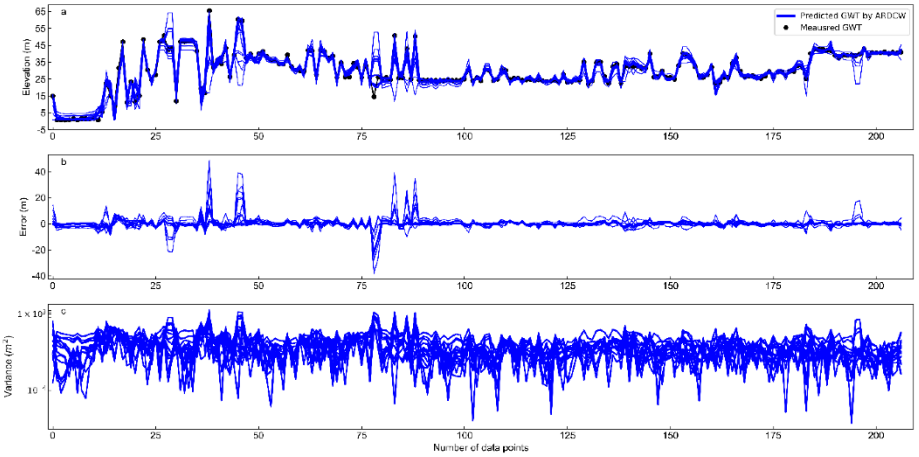


Figure 6. Predictability of the optimum model subjected to applied dropouts in training and testing stages for (a) *GWT* data, (b) mean and (c) variance

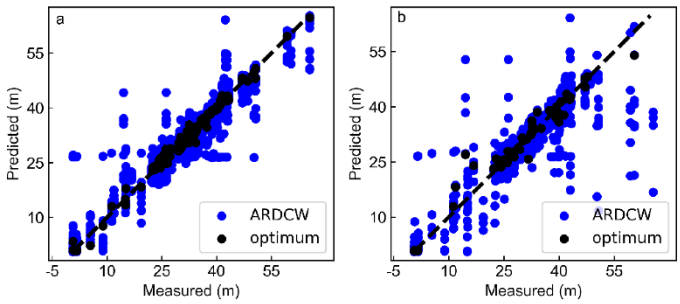


Figure 7. Comparison of the distribution of optimum and dropout models for (a) training and (b) testing data sets

Proper experimental measurement associated with an estimated level of confidence not only allows a scientific hypothesis to be confirmed or refuted, but also facilitates the judgments on the quality of the data and leads to meaningful comparisons with other similar values or predictive models. Accordingly, 95% prediction interval

(*PI*) shows the certain probability of an estimation of future observation and is often used in regression analysis (Figure 8).

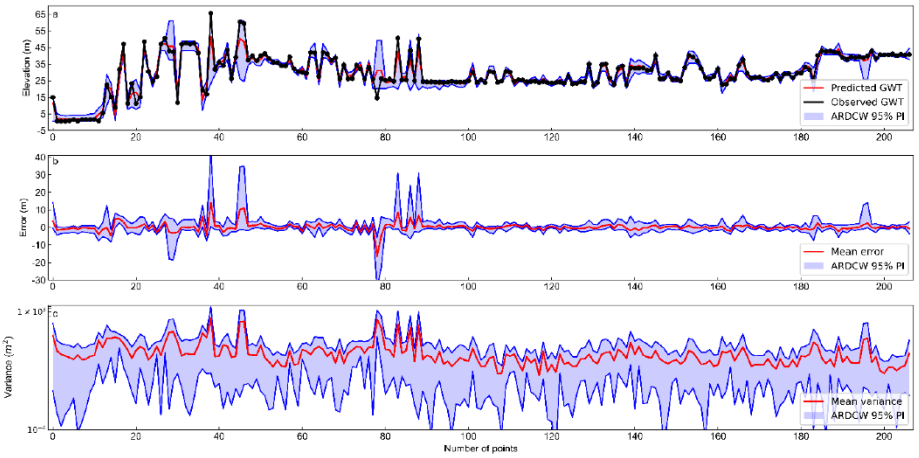
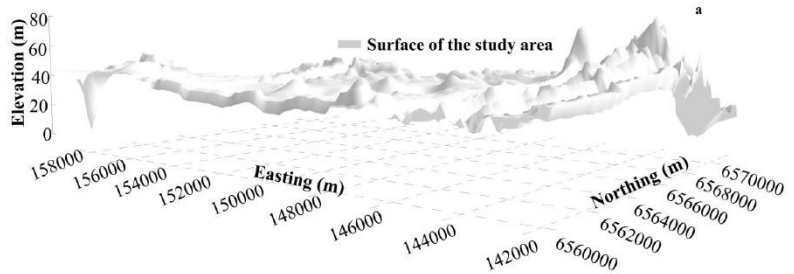
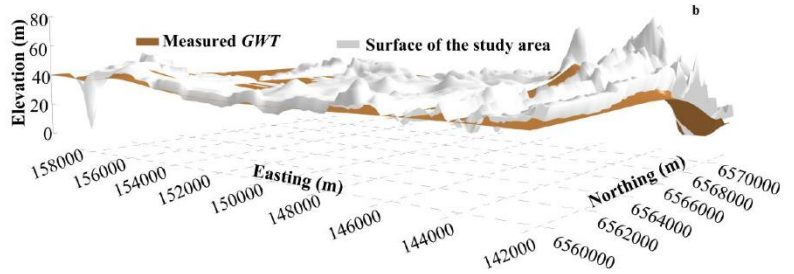


Figure 8. Calculated *PI* using the optimum model for (a) *GWT* data, (b) predicted mean and (c) predicted variance

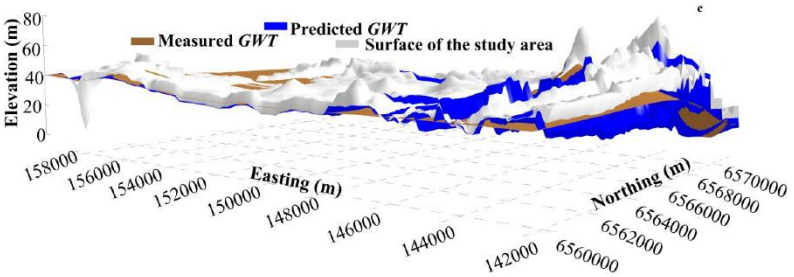
Due to the dynamic nature of areas without data, creating a 3D visual perspective depicting the predictive spatial distribution of the *GWT* levels is a challenging task, but this can provide more utility in interpreting the subsurface characterization (Abbaszadeh Shahri et al., 2020, 2021b). However, the level of accuracy and the confidence of the model can vary depending on the quality of available data and the approach. This is the reason why pseudo data can therefore play an important role in building knowledge for unsampled locations (Tacher et al., 2006; Abbaszadeh Shahri et al., 2020; Cruzes & Dyba, 2011). Figure 9 shows the step-by-step creation of the 3D model of the study area depicting the retrieved outlines of the uncertainties. Referring to the estimated *UQ*, more comprehensive concepts on the spatial *GWT* pattern can be realized to avoid the relevant risk of facing aquifers during geoengineering projects. Due to the ease of updating with new data, the flexibility of such models provides a preferred tool for geo-engineers and decision planners in the observation and analysis of geo-environmental engineering issues within a project.



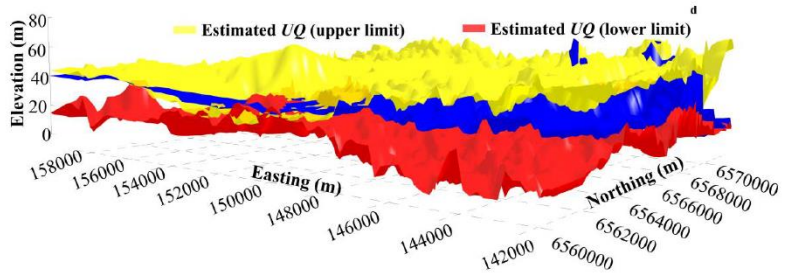
334



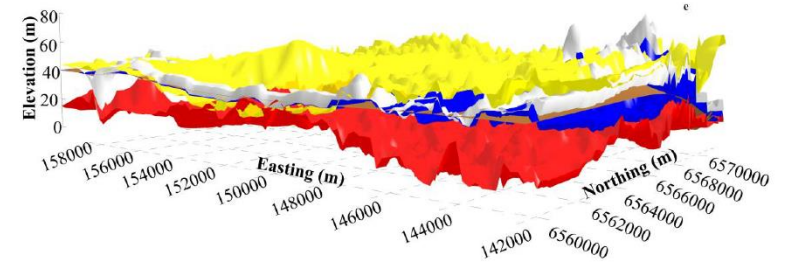
335



336



337



338

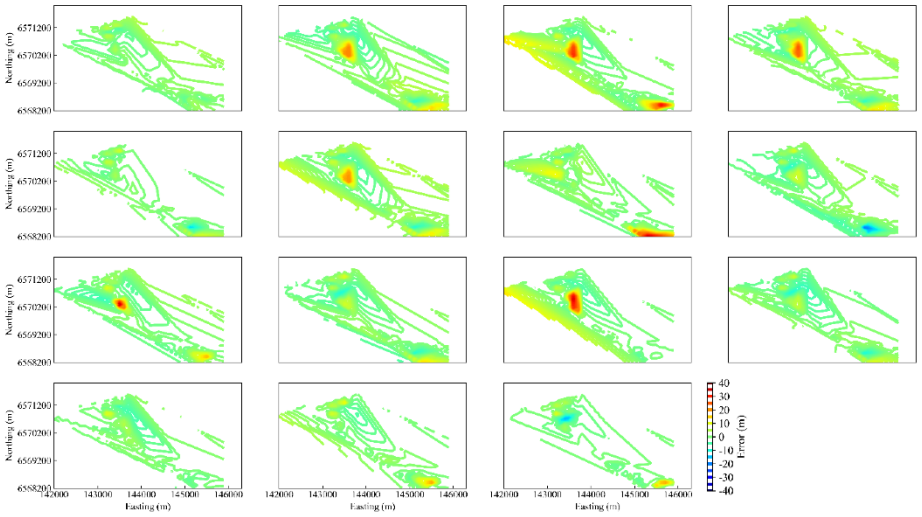
Figure 9. Incorporating the results of predictive optimum topology and the proposed *ARDCW* in visualizing the estimated uncertainty: (a) surface of the area, (b) measured *GWT*, (c) predicted *GWT*, (d) upper and lower limit of estimated *UQ* and (e) the overlaid supplementary perspective for the entire study area

6. Discussion and validation

Predictive *AIT*-based models are indispensable tools in geoengineering applications, not only to ensure the function in the labelled variables but also to estimate the *UQ* in unlabelled data. In recent years, several studies have been dedicated to the importance of *UQ* in *GWT* modelling and decision-making in different scales (Middlemis & Peeters, 2018; WWAP, 2012). However, from geoengineering point of view, each parameter of a geologic object can be a source of uncertainty, and thus in practice, the number of involved sources can be innumerable. To evaluate the robustness and accuracy of the predictive spatial pattern of *GWT*, the *UQ* analysis is often required. Since standard *AIT*-based *UQ* procedures are still limited, developing an efficient method that uses *DNN* predictive models in particular is of great interest for industrial-scale and real-world applications. In the current paper, the presented *UQ* approach, *ARDCW*, is discussed and validated through contour maps of predicted error in different dropouts, accuracy metrics, success rate, and through comparison with *MCD* and *QR* methods.

6.1. Effects of dropout on estimated *UQ*

This concern was monitored through the comparison of the variation of captured distributions in post and prior predictions with respect to the optimum topology. These variations are referred to in the *ARDCW* by using the saved updated database for each examined model (Figure 5). The comparison of different post distributions and the corresponding residual contour map of the area are given in Figure 10, which helps to practically find a function to fit variational objective data with the minimum penalty with respect to the prior distribution in the optimum model. This implies that in case of a lack of observations, the *UQ* can be interpreted through the interpolation of unlabelled predictions.



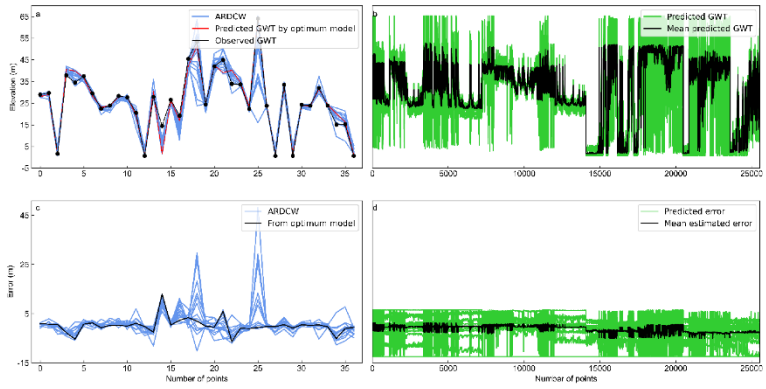
365

366 **Figure 10.** Contour map of the residual for different results of *ARDCW*

367

368 The uncertainty is generally evaluated from data and models due to the presence of noise and an imbalanced
369 training data distribution (Loquerico & Segu, 2020). In *AIT*-based models, the source of uncertainty occurs when
370 the feed data are mismatched, where the smaller differences between the observed *GWT* and those predicted by
371 *ARDCW* show a higher degree of safety in the prediction process (Figures 11a and b). Subsequently, the higher
372 differences can be interpreted as higher uncertainties in the prediction. Accordingly, the performance of the
373 optimum and examined models using *ARDCW* using *GWT*, as well as error for the measured and unlabelled
374 prediction datasets, are given in Figure 11 (c and d).

375



376

377 **Figure 11.** Variation of predicted *GWT* and error for validation (a and c) and unlabelled data (b and d)

6.2. Comparing different *UQ* models

6.2.1. Statistical metrics

The coefficient of efficiency (*Ec*) (Nash & Sutcliffe, 1970) is one of the most widely used metrics to evaluate the performance of hydrologic models:

$$E_c = 1.0 - \frac{\sum_{i=1}^{N}(O_i - P_i)^2}{\sum_{i=1}^{N}(O_i - \bar{O})^2} \qquad \bar{O}: observed\ mean \tag{12}$$

Where; O_i and P_i denote the observed and predicted data. $E_c \epsilon$ (-∞, 1.0] interval in which the value of 1.0 expresses the perfect fit. E_c=0 reflects large variability in the observed data, while E_c<0 indicates that the $\bar{O}$ would have been a better predictor than the model. Further, E_c physically is the ratio of mean square error (*MSE*):

$$MSE = N^{-1}\sum_{i=1}^{N}(O_i - P_i)^2 \tag{13}$$

Therefore, E_c represents an improvement over the coefficient of determination for model evaluation purposes in that it is sensitive to differences in the observed and model simulated means and variance (Leavesley et al., 1983; Wilcox et al., 1990).

To cover the lack and insensitivity of E_c and R^2 in considering the calculated square differences between the observed and predicted means and variances (Legates & McCabe, 1999), the index of agreement (Willmott, 1984) as the ratio of mean square and potential errors is defined as:

$$IA = 1 - \frac{\sum_{i=1}^{N}(O_i - P_i)^2}{\sum_{i=1}^{N}(|P_i - \bar{O}| - |P_i - \bar{O}|)^2} \tag{14}$$

The percentage of observed *GWT* (N_{GWT}) bracketed by *CI* in the level of 95% ($P_{CI}^{95\%}$) is defined by:

$$P_{CI}^{95\%} = \frac{N_{GWT}}{N} \times 100 \tag{15}$$

Where; *N* shows the total number of *GWT* data. The properly modelled *UQ* is justified based on the closeness to 100%.

Referring to Jin et al., (2010) the quality of the predicted *UQ* can be assessed using the average relative interval length of the *CI* (*ARIL*):

$$ARIL_{CI}^{\%95} = \frac{1}{n}\sum \frac{UCL_{pred,i}^{GWT} - LCL_{pred,i}^{GWT}}{GWT_i^{obs}} \tag{16}$$

Where; $UCL_{pred,i}^{GWT}$ and $LCL_{pred,i}^{GWT}$ express the calculated upper and lower *CI* of the i^{th} predicted *GWT*, and *n* is the number of total observations; thus, a lower $ARIL_{CI}^{\%95}$ value represents better performance.

A comparison of these indexes in evaluating the predicted *UQ* using three different methods including *ARDCW*, *QR* and *MCD* is given in Figure 11 and Table 1. Referring to $P_{CI}^{95\%}$, although both *ARDCW* and *QR* showed competitive coverage of the true *GWT* levels (95%) with respect to *MCD* (68%), the ranking score of *ARDCW* was expressed to be superior to other methods. The closeness statistics of the *ARDCW* and *QR* methods can be interpreted as similar properties of posterior distribution of predicted *GWT* for these methods.

Table 1. Comparison and ranking the *UQ* models through statistical metrics

Model	Performance criteria							Ranking of criteria								
	E_c	*RMSE*	*IA*	R^2	$P_{CI}^{95\%}$	$ARIL_{CI}^{\%95}$	$P_{PI}^{95\%}$	E_c	*RMSE*	*IA*	R^2	$P_{CI}^{95\%}$	$ARIL_{CI}^{\%95}$	$P_{PI}^{95\%}$	Total rank	Sort order
ARDCW	0.97	2.51	0.99	0.97	10.66	3.52	0.95	3	3	3	3	1	2	3	18	1
QR	0.96	2.99	0.99	0.96	10.44	2.73	0.95	2	2	3	2	2	3	3	17	2
MCD	0.94	3.5	0.98	0.94	9.5	5.76	0.68	1	1	1	1	3	1	1	9	3

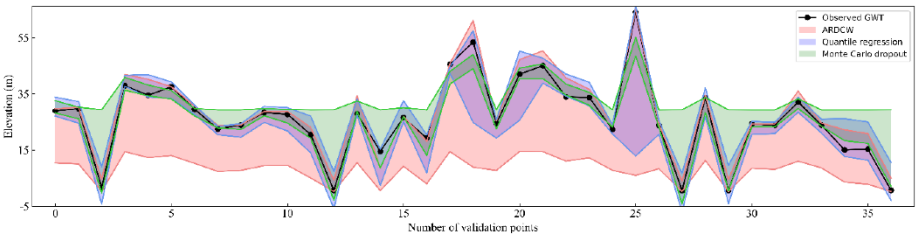


Figure 12. Compared $P_{CI}^{95\%}$ of three models using validation dataset

6.2.2. Success rate using confusion matrix

The layout of the confusion matrix can intuitively describe the performance of a model on a set of data using the true labelled instances (Asheghi et al., 2019). This concept practically conceptualizes the error probabilities of developed models in assigning the individual predicted outputs into the classified input. The best performance is with zero values, except on diagonal arrays. Referring to conducted confusion matrix (Table 2), *ARDCW* and *QR* methods provide more true predicted instances (33 and 29 out of 37) than the *MCD*. As presented in Table 3, the performance of *ARDCW* in terms of classification accuracy and misclassification error at an 89% correct classification showed 22% and 14% improvement in estimated *UQ* than the *MCD* and *QR*, respectively.

425 **Table 2.** Comparison of the success rate of different *UQ* methods for *GWT* data

	ARDCW												Results		
Target	**[<0.62]**	**[0.62, 6.98]**	**[6.98, 13.34]**	**[13.34, 19.70]**	**[19.70, 26.06]**	**[26.06, 32.42]**	**[32.42, 38.78]**	**[38.78, 45.14]**	**[45.14, 51.50]**	**[51.50, 57.86]**	**[57.86, 64.22]**	**[>64.12]**	**Total**	**True**	**False**
[<0.62]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
[0.62-6.98]	0	5	0	0	0	0	0	0	0	0	0	0	5	5	0
[6.98-13.34]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
[13.34-19.70]	0	1	0	3	0	0	0	0	0	0	0	0	4	3	1
[19.70-26.06]	0	0	0	1	8	0	0	0	0	0	0	0	9	8	1
[26.06-32.42]	0	0	0	0	0	8	0	0	0	0	0	0	8	8	0
[32.42-38.78]	0	0	0	0	0	0	5	1	0	0	0	0	6	5	1
[38.78-45.14]	0	0	0	0	0	0	0	2	0	0	0	0	2	2	0
[45.14-51.50]	0	0	0	0	0	0	0	1	0	0	0	0	1	0	1
[51.50-57.86]	0	0	0	0	0	0	0	0	0	1	0	0	1	1	0
[57.86-64.22]	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0
[>64.12]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Note	0	6	0	4	8	8	5	4	0	1	1	0	37	33	4
	MCD												Results		
Note	0	1	4	2	11	9	5	3	1	1	0	0	37	27	10
	QR												Results		
Note	2	3	1	3	10	7	5	4	1	1	0	0	37	29	8

426

427 **Table 3.** Improvement of *ARDCW* in estimating the *UQ* compared to *MCD* and *QR*

Model	**Classification accuracy (%)**		**Misclassification error (%)**		**Progress (%)**	**Improvement compared to other models (%)**			
	Test	**Validate**	**Test**	**Validate**		***MCD***	***ARDCW***	***QR***	**Rank**
MCD	0.69	0.73	0.31	0.27	0.04	---	-22	-6.8	3
ARDCW	0.77	0.89	0.23	0.11	0.12	22	---	14	1
QR	0.73	0.78	0.27	0.22	0.05	6.8	-14	---	2

428

429 ### 6.2.3. Uncertainty intervals in studied area

430 Sufficiently accurate modelling procedures are essential tools in subsurface geoengineering disciplines to reflect
431 fidelity between the real and counterparts. However, despite the enhanced computational power, different
432 sources such as modelling errors in describing the real system, numerical errors generated by mathematical

equations, and data errors caused by uncertainties can cause the appearance of differences. A critical part of prediction is an assessment of how much a predicted value will fluctuate due to noise or variations in the data. Figure 13 shows the 95% uncertainty intervals estimated with *ARDCW* for the studied area. It can provide the distribution of different uncertainty levels and gives an indication of the reason for high uncertainties, for example, if uncertainty arises from the lack of observed data or from sudden big changes in the groundwater levels. This map can also help in the planning of future data collection and determining where to drill more groundwater boreholes to reduce high uncertainties in estimating the groundwater surface.

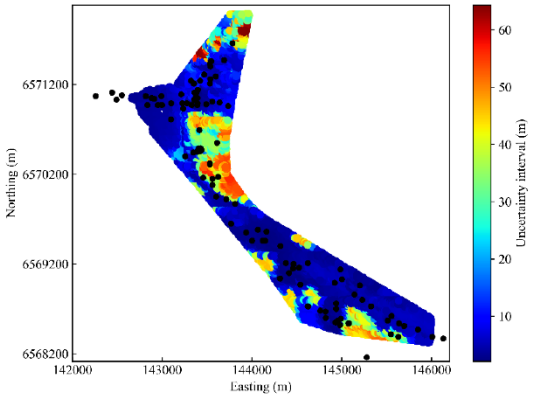


Figure 13. Estimated uncertainty intervals using applied *GWT* data (black dots) in the entirety of the study area

6.2.4. Comparison of *CI* and *PI*

Due to the dynamic nature and time-dependent behaviour, the exact modelling of *GWT* problems is a complex issue to resolve. The time dependent *GWT* data points over an interval can benefit from outlier removal to present an overall perspective of variation in the study area as well as a better analysis and understanding and thus uncover patterns in datasets for future forecasting (Shumway, 1988; Keogh & Kasetty, 2003). However, the problem regarding the generalization from a single study, difficulty in obtaining appropriate measures, and problems with accurately identifying the correct model to represent the data should be considered.

Therefore, the bias and discrepancy of the applied models can be interpreted through the comparison of the estimated *UQ*. Statistically, a predictive model is stable and under control if most of the predictions fall within the range of the confidence interval (*CI*). This range refers to the long-term success rate of the method in capturing the predicted output, where the wider the *CI*, the greater the instability. The *CI* level of 95% reflects a range of values where 5% can contain the false mean of the population. Due to the lack of knowledge,

observation error (variability of experimental measurements) and the underlying physics in *GWT* problems, such a comparison (Figure 14) can show how accurately a mathematical model describes the true system in a real-life situation. Mathematically, the variation in predicted *UQ* comes from the adjusted model parameters and, correspondingly, the implemented algorithm as well as feed input variables whose exact values are unknown to experimentalists or cannot be inferred by statistical methods (Kennedy & O'Hagan, 2001). Therefore, the dimensionality of an engineering problem would cause variability in its performance, whereby the implemented algorithm can provide discrete numerical uncertainty.

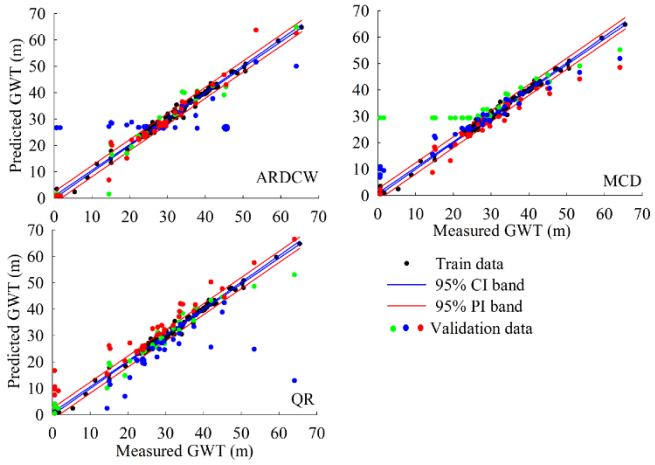


Figure 14. Comparative plots of the applied method in *UQ* analysis at a level of 95% for *CI* and *PI*

Concluding remarks

AI-based models are becoming a standard computational and decision-making tool in civil and construction industries. However, due to inherent complexities, computational costs and poor performance, most of the *AI*-based *UQ* methods rarely make the leap from research to production. Therefore, in this context the need for a validated coding procedure able to meet the regulatory requirements was highlighted in the current paper. In order to gain insight into the predictability level of the complex *AIT*-based predictive subsurface geoengineering model, an evaluation of the estimated *UQ* using simulation codes with iterative analyses is required. Furthermore, since a model or method can be adjusted using a wide range of internal characteristics, subsequently presenting reliable but robust *UQ* estimates presents a significant challenge. Therefore, if *AI* systems could reliably identify unlabelled data, they could be deployed with a greater degree of certainty. This

476 concern is especially vital for geoengineering applications, where the expectation is to achieve the same
477 distribution for observations and the training data.
478 In response to greater demand for *DNN*-based *UQ* methods, the *ARDCW,* as a new state-of-the-art approach that
479 applies randomly weighted components that dropout at test time, was introduced and developed. Using *ARDCW,*
480 the estimated *UQ* was visualized using the average results of many ensembled predictive models derived from a
481 demonstrated learned optimum automated *DNN* topology through *K*-fold validation and randomly deactivated
482 weights. The issue of underestimate uncertainty, which is usually observed in most *AIT*-based *UQ* methods,
483 originates from changes in model and optimization process. In this point of view, the *ARDCW* approach does not
484 require any changes in the topology and can be applied on already trained models. Since the *ARDCW* is built on
485 the weight database, it does not have restrictions in application on time-dependent data such as *GWT* with
486 sequences of seasonal fluctuations in monitored data over different intervals of the year. However, presenting a
487 seasonal time dependent *GWT* model was beyond the scope of this paper.
488 An experimental application of the proposed *ARDCW* was applied on 244 measured *GWTs* from a
489 geoengineering project in Stockholm, Sweden. The 3-32-8-1 predictive *DNN* topology, as the characterized
490 optimum model showed significant competence in comparison with *MCD* and *QR*. The results showed that due
491 to the use of different topologies, the *MCD* can behave differently at training and inference time, while for
492 assured dropout creating a custom layer with predefined training parameters for regular dropout should be
493 employed.
494 Using the *ARDCW,* the estimated *UQ* boundaries with a 95% *CI* were visualized and presented on the generated
495 3D predictive spatial pattern of subsurface *GWT* data. This cost-effective and adequately accurate tool can reflect
496 the potential risks associated with the distributed spatial aquifers, thus preventing water inrush, or facing
497 underground geo-structures.
498 The validity of most analyzed *GWTs* with respect to both engineering and societal challenges is limited in the
499 findings as it is often not included with the uncertainty assessment. This implies that uncertainty should be
500 estimated in many previously presented *GWT* models, which are intended for efficient decision support. Using
501 the created 3D predictive model, the geospatial distribution of *GWT* and corresponding variations within
502 subsurface geo-formations can be pursued. This concern can significantly influence the design of underground
503 transport infrastructures for the city of Stockholm, where, in geo-engineering perspective the drainage of the
504 *GWT* through tunnelling can cause ground settlements in the surrounding buildings. This phenomenon can be
505 particularly problematic in weak soil/rock formations and manifest as a defective injection process. Furthermore,

506 the created model can decrease the risk of water inrush from excavation. By adding more geological data, the
507 ability to recognize swelling clayey material which is a major threat in tunnelling projects can also be achieved.
508 This implies that the presented *GWT* model is an indispensable tool for decision makers in urban development
509 projects (e.g. building detached homes, roads, railways and bridges), where substantial land surface processes
510 can be imposed.
511

512 ## Declarations:

513 **Disclosure of potential conflicts of interest:** The authors have no known conflicts of interest to declare
514 associated with this publication.
515 **Declaration of competing interest:** The authors confirm that there has been no significant financial support for
516 this work that could have influenced its outcome.
517 **Funding:** This work was supported by Rock engineering Research Foundation (BeFo) (Grant number BeFO
518 415), and Swedish Transport Administration through Better Interactions in Geotechnics (BIG) (Grant number
519 BIG A2019:17).
520

521 ## Acknowledgments

522 The authors would like to express their special gratitude to Swedish Transport Administration through Better
523 Interactions in Geotechnics (BIG), the Rock engineering Research Foundation (BeFo) and Tyréns AB for
524 supporting this research. We also wish to thank our colleagues, who provided insight and expertise that greatly
525 assisted us in our research.
526

527 ## References:

528 1. Abbaszadeh, S. A., Shan, C., Zäll, E., & Larsson, S. (2021a). Spatial distribution modelling of
529 subsurface bedrock using a developed automated intelligence deep learning procedure: A case study in
530 Sweden. *Journal of Rock Mechanics and Geotechnical Engineering*, 13(6)
531 2. Abbaszadeh, S. A., Kheiri, A., & Hamzeh, A. (2021b). Subsurface topographic modelling using
532 geospatial and data driven algorithm. *ISPRS Int. J. Geo–Inf.*, 10(5), 341. https://doi.org/
533 10.3390/ijgi10050341

3. Abbaszadeh, S. A., Larsson, S., & Renkel, C. (2020). Artificial intelligence models to generate visualized bedrock level: a case study in Sweden. *Modeling Earth Systems and Environment*, 6, 1509–1528. https://doi.org/ 10.1007/s40808–020–00767–0
4. Angiulli, F., & Fassetti, F. (2020). Uncertain distance–based outlier detection with arbitrarily shaped data objects. *Journal of Intelligent Information Systems*, 57, 1–24. https://doi.org/ 10.1007/s10844–020–00624–7
5. Asheghi, R., Hosseini, S.A., Saneie, M., & Abbaszadeh S. A. (2020). Updating the neural network sediment load models using different sensitivity analysis methods: a regional application. *Journal of Hydroinformatics*, 22(3), 562–577. https://doi.org/ 10.2166/hydro.2020.098
6. Asheghi, R., Abbaszadeh, S. A., & Khorsand, Z. M. (2019). Prediction of uniaxial compressive strength of different quarried rocks using metaheuristic algorithm. *Arabian Journal for Science and Engineering*, 44, 8645–8659. https://doi.org/ 10.1007/s13369–019–04046–8
7. Atanassov, E., & Dimov, I.T. (2008). What Monte Carlo models can do and cannot do efficiently?. *Applied Mathematical Modelling*, 32(8):1477–1500. https://doi.org/ 10.1016/j.apm.2007.04.010
8. Bárdossy, G., & Fodor, J. (2001). Traditional and new ways to handle uncertainty in geology. *Natural Resources Research*, 10, 179–187. https://doi.org/ 10.1023/A:1012513107364
9. Barford, N.C. (1985). *Experimental measurements: precision, error, and truth*. Wiley–Blackwell.
10. Barrio, R., Rodriguez, M., Abad, A., & Blesa, F. (2011). Breaking limits: The Taylor series method. *Applied Mathematics and Computation*. 217(20):7940–7954. https://doi.org/ 10.1016/j.amc.2011.02.080
11. Bergström, S. (1976). *Development and application of a conceptual runoff model for Scandinavian catchments*. Sweden Meteorological and Hydrological Institute
12. Beven, K., & Binley, A. (1992). The future of distributed models: model calibration and uncertainty prediction. *Hydrol. Process.*, 6, 279–298. https://doi.org/ 10.1002/hyp.3360060305
13. Bizhanimanzar, M., Leconte, R., & Nuth, M. (2019). Modeling of shallow water table dynamics using conceptual and physically based integrated surface–water–groundwater hydrologic models. *Hydrology Earth System Science*, 23(5), 2245–2260. https://doi.org/ 10.5194/hess–23–2245–2019
14. Borgonovo, E. (2006). Measuring uncertainty importance: Investigation and comparison of alternative approaches. *Risk Analysis*, 26(5), 1349–1361. https://doi.org/ 10.1111/j.1539–6924.2006.00806.x

15. Cacuci, D.G., & Ionescu–Bujor, M. (2004). A comparative review of sensitivity and uncertainty analysis of large–scale systems–II: Statistical methods. *Nuclear Science and Engineering*, 147(3), 204–217. https://doi.org/ 10.13182/04–54CR
16. Chahbaz, R., Sadek, S., & Najjar, S. (2019). Uncertainty quantification of the bond stress– displacement relationship of shoring anchors in different geologic units. *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, 13(4), 276–283. https://doi.org/ 10.1080/17499518.2019.1612527
17. Chen, C., He, W., Zhou, H., Xue, Y., & Zhu, M. (2020). A comparative study among machine learning and numerical models for simulating groundwater dynamics in the Heihe River Basin, northwestern China. *Scientific Reports*, 10, 3904. https://doi.org/10.1038/s41598–020–60698–9
18. Chen, T., Fox, E., & Guestrin, C. (2014). Stochastic gradient hamiltonian monte carlo. In *Proc international conference on machine learning*, 1683–1691
19. Cox, D.C., & Baybutt, P. (1981). Methods for uncertainty analysis: a comparative survey. *Risk Analysis*, 1(4), 251–258. https://doi.org/ 10.1111/j.1539–6924.1981.tb01425.x
20. Cruzes, D.S., & Dyba, T. (2011). Research synthesis in software engineering: A tertiary study. *Information and Software Technology*, 53, 440–455. https://doi.org/10.1016/j.infsof.2011.01.004
21. Dietterich, T.G. (2000). Ensemble methods in machine learning. In *Multiple classifier systems, MCS 2000, Lecture notes in computer science,* 1857, 1–15. https://doi.org/ 10.1007/3–540–45014–9_1
22. Eisenhart, C., Ku, H.H., & Cole, R. (1983). Expression of the Uncertainties of Final Measurement Results: reprints. *U.S. Department of Commerce/ National Bureau of Standards,* NBS Special Pub. 644, Washington DC, USA.
23. Elam, K.R., & Rearden, B.T. (2017). Use of sensitivity and uncertainty analysis to select benchmark experiments for the validation of computer codes and data. *Nuclear Science and Engineering*. 145(2):196–212. https://doi.org/ 10.13182/NSE03–A2376
24. Farrance, I., & Frenkel, R. (2012). Uncertainty of measurement: A review of the rules for calculating uncertainty components through functional relationships. *Clin Biochem Rev.*, 33(2), 49–75
25. Foong, A.K., Burt, D.R., Li, Y., & Turner, R.E. (2019). Pathologies of factorized gaussian and MC dropout posteriors in Bayesian neural networks. In *Proc. 4th workshop on Bayesian Deep Learning (NeurIPS 2019)*, abs/1909.00719, Vancouver, Canada

592 26. Freitas, C. (2002). The issue of numerical uncertainty. *Applied Mathematical Modeling*. 26(2), 237–
593 248. https://doi.org/ 10.1016/S0307–904X(01)00058–0
594 27. Gal, Y., & Ghahramani, Z. (2016). Dropout as a Bayesian approximation: representing model
595 uncertainty in deep learning. In *Balcan MF and Weinberger KQ (eds), Proc 33nd International*
596 *Conference on Machine Learning, ICML*, 48, 1050–1059, NY, USA
597 28. Geffray, C., Gerschenfeld, A., Kudinov, P., Mickus, I., Jeltsov, M., Kööp, K., Grishchenko, D., &
598 Pointer, D. (2019). Verification and validation and uncertainty quantification. *Thermal Hydraulics*
599 *Aspects of Liquid Metal Cooled Nuclear Reactors*, 383–405. https://doi.org/ 10.1016/B978–0–08–
600 101980–1.00008–9
601 29. Glimm, J., & Sharp, D.H. (1999). Prediction and the quantification of uncertainty. *Physica D*, 133(1–4),
602 152–170. https://doi.org/ 10.1016/S0167–2789(99)00103–7
603 30. Goodman, L. (1960). On the exact variance of products. *Journal of the American Statistical*
604 *Association*, 55(292), 708–713. https://doi.org/10.2307/2281592
605 31. Guillaume, J.H.A., Hunt, R.J., Comunian, A., Blakers, R.S., & Fu, B. (2016). Methods for exploring
606 uncertainty in groundwater management predictions. *Integrated groundwater management*, 711–737.
607 https://doi.org/10.1007/978–3–319–23576–9_28
608 32. Gärdenfors, P., & Sahlin, N.E. (1982). Unreliable probabilities, risk taking, and decision making.
609 *Synthese,* 53(3), 361–386. https://doi.org/10.1007/BF00486156
610 33. Huang, W., Zhao, D., Sun, F., Liu, H., & Chang, E. (2015). Scalable Gaussian process regression using
611 deep neural networks. In *Proc 24th International Conference on Artificial Intelligence*, IJCAI'15, 3576–
612 3582
613 34. Hauser, J., Wellmann, F., & Trefry, M. (2017). Water table uncertainties due to uncertainties in
614 structure and properties of an unconfined aquifer. *Groundwater*, 56(2):251–265. https://doi.org/
615 10.1111/gwat.12577
616 35. Hernandez–Lobato, J.M., & Adams, R. (2015). Probabilistic backpropagation for scalable learning of
617 bayesian neural networks. In *Proc 32nd International Conference on Machine Learning*, ICML, 37,
618 1861–1869
619 36. Hernandez, S., & Lopez, J.L. (2020). Uncertainty quantification for plant disease detection using
620 Bayesian deep learning. *Applied Soft Computing*, 96, 106597.
621 https://doi.org/10.1016/j.asoc.2020.106597

622 37. HInton, E., Srivastava, N., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2012). Improving neural
623 networks by preventing co–adaptation of feature detectors. arXiv:1207.0580.
624 38. Hirschefeld, L., Swanson, K., Yang, K., Barzilay, R., & Coley, C.W. (2020). Uncertainty quantification
625 using neural networks for molecular property prediction. *Journal of Chemical Information and*
626 *modeling.* 60, 3770–3780. https://doi.org/ 10.1021/acs.jcim.0c00502
627 39. Hu, L., & Jiao, J.J. (2010). Modeling the influences of land reclamation on groundwater systems: A
628 case study in Shekou peninsula, Shenzhen, China. *Engineering geology*, 114(3–4),144–153.
629 https://doi.org/ 10.1016/j.enggeo.2010.04.011
630 40. Huber, M. (2016). Reducing forecast uncertainty by using observations in geotechnical engineering.
631 *Probabilistic Engineering Mechanics*, 45, 212–219. https://doi.org/10.1016/j.probengmech.2016.02.002
632 41. Jiang, C., Zheng, J., & Han, X. (2018). Probability–interval hybrid uncertainty analysis for structures
633 with both aleatory and epistemic uncertainties: a review. *Struct Multidisc Optim.,* 57, 2485–2502.
634 https://doi.org/ 10.1007/s00158–017–1864–4
635 42. Jin, X., Xu, C.Y., Zhang, Q., & Singh, V.P. (2010). Parameter and modeling uncertainty simulated by
636 GLUE and a formal Bayesian method for a conceptual hydrological model. *Journal of Hydrology*, 383,
637 147–155. https://doi.org/ 10.1016/j.jhydrol.2009.12.028
638 43. Kabir, H.D., Khosravi, A., Hosen, M.A., & Nahavandi, S. (2018). Neural network–based uncertainty
639 quantification: a survey of methodologies and applications. *IEEE Access*., 6, 36218 – 36234.
640 https://doi.org/10.1109/ACCESS.2018.2836917
641 44. Kadanoff, L.P. (2009). More is the same, phase transitions and mean field theories. *Journal of*
642 *Statistical Physics,* 137(5–6), 777–797. https://doi.org/10.1007/s10955–009–9814–1
643 45. Kennedy, M.C., & O'Hagan, A. (2001). Bayesian calibration of computer models. *Journal of the Royal*
644 *Statistical Society, Series B (Statistical Methodology)*, 63(3), 425–464
645 46. Keogh, E., & Kasetty, S. (2003). On the need for time series data mining benchmarks: A survey and
646 empirical demonstration. *Data Mining and Knowledge Discovery*, 7, 349–371. https://doi.org/
647 10.1023/A:1024988512476
648 47. Klotz, D., Kratzert, F., Gauch, M., Sampson, A.K., Brandstetter, J., Klambauer, G., Hochreiter, S., &
649 Nearing, G. (2021). Uncertainty estimation with deep learning for rainfall–runoff modeling. *Hydrology*
650 *and Earth System Science Discussion*, https://doi.org/ 10.5194/hess–2021–154

48. Krzywinski, M., & Altman, N. (2013). Points of significance: importance of being uncertain. *Nature methods*, 10(9), 809–810. https://doi.org/ 10.1038/nmeth.2613

49. Lakshminarayanan, B., Pritzel, A., & Blundell, C. (2017). Simple and scalable predictive uncertainty estimation using deep ensembles. In *Proc 31st International Conference on Neural Information Processing Systems*, NIPS'17, 6405–6416

50. Leavesley, G.H., Lichty, R.W., Troutman, B.M., & Saindon, L.G. (1983). Precipitation-runoff modeling system user's manual. *Water-Resources Investigations Report* 83–4238, USGS, Water division, USA. https://doi.org/ 10.3133/wri834238

51. Legates, D.R., & McCabe, G.J. (1999). Evaluating the use of "goodness–of–fit" measures in hydrologic and hydroclimatic model validation. *Water resource Research*, 35(1), 233–241. https://doi.org/ 10.1029/1998WR900018

52. Li, F., Zhu, J., Deng, X., Zhao, Y., & Li, S. (2018). Assessment and uncertainty analysis of groundwater risk. *Environmental Research*, 160, 140–151. https://doi.org/10.1016/j.envres.2017.09.030

53. Loquerico, A., Segu, M., & Scaramuzza, D. (2020). A general framework for uncertainty estimation in deep learning. *IEEE Robotics and Automation Letters*, 5(2), 3153–3160. https://doi.org/10.1109/LRA.2020.297468

54. MacKay, D.J.C. (1992). A practical Bayesian framework for backpropagation networks. *Neural Computation*, 4(3), 448–472. https://doi.org/ 10.1162/neco.1992.4.3.448

55. Middlemis, H., & Peeters, L.J.M. (2018). Uncertainty analysis–Guidance for groundwater modelling within a risk management framework. *Report of Information guidelines explonatory note*, IESC on Coal Seam Gas and Large Coal Mining Development, Department of the Environment and Energy, Commonwealth of Australia, 77P. https://doi.org/10.13140/RG.2.2.34589.36323

56. Mohammadi, K. (2009). Groundwater table estimation using MODFLOW and artificial neural networks. In R.J. Abrahart, L.M. See, & D.P. Solomatine (Eds.), *Practical hydroinformatics* (volume 68, pp.1287–138), Springer. https://doi.org/ 10.1007/978–3–540–79881–1_10

57. Morgan, M.G., & Henrion, M. (1990). *Uncertainty: a guide to dealing with uncertainty in quantitative risk and policy analysis*. Cambridge University Press

58. Nash, J.E., & Sutcliffe, J.V. (1970). River flow forecasting through conceptual models part I – A discussion of principles. *J Hydrol.,* 10, 282–290. https://doi.org/ 10.1016/0022– 1694(70)90255–6

59. Neal, R.M. (2012). Bayesian learning for neural networks. *Lecture notes in statistics*, 118, Springer. https://doi.org/ 10.1007/978–1–4612–0745–0

60. Parry, S., Baynes, F.J., Culshaw, M.G., Eggers, M., Keaton, J.R., Lentfer, K., Novotny, J., & Paul, D. (2014). Engineering geological models: an introduction: IAEG commission 25, *Bull Eng Geol Environ.*, 73, 689–706. https://doi.org/ 10.1007/s10064–014–0614–8

61. Quinonero–Candela, J., Rasmussen, C.E., Sinz, F., Bousquet, O., Schölkopf, & B. (2006). Evaluating predictive uncertainty challenge. *Lecture notes in computer science*, 3944, 1–27, Springer. https://doi.org/ 10.1007/11736790_1

62. Rushton, K.R. (2003). *Groundwater hydrology: conceptual and computational models*. Wiley.

63. Sahoo, S., Russo, & T.A. (2017). Machine learning algorithms for modeling groundwater level changes in agricultural regions of the U.S.. *Water Resource Research*, 53(5), 3878–3895. https://doi.org/ 10.1002/2016WR019933

64. Salvo, C.D., Mancini, M., Cavinato, G.P., Moscatelli, M., Simionato, M., Stigliano, F., Rea, R., & Rodi, A. (2020). A 3D geological model as a base for the development of a conceptual groundwater scheme in the area of the Colosseum (Rome, Italy). *Geosciences*, 10, 266. https://doi.org/10.3390/geosciences10070266

65. Sepulveda, N., & Doherty, J. (2015). Uncertainty analysis of a groundwater flow model in east- central Florida. *Groundwater*, 53(3), 464–474. https://doi.org/ 10.1111/gwat.12232

66. Shumway, R.H. (1988). *Applied statistical time series analysis*. Englewood Cliffs

67. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(56), 1929–1958

68. Stedinger, J.R., Vogel, R.M., Lee, S.U., & Batchelder, R. (2008). Appraisal of the generalized likelihood uncertainty estimation (GLUE) method. *Water Resour. Res.*, 44(12), W00B06. https://doi.org/ 10.1029/2008WR006822.

69. Tacher, L., Pomian–Srzednicki, I., & Parriaux, A. (2006). Geological uncertainties associated with 3D subsurface models. *Computers and Geosciences*, 32, 212–221. https://doi.org/10.1016/j.cageo.2005.06.010

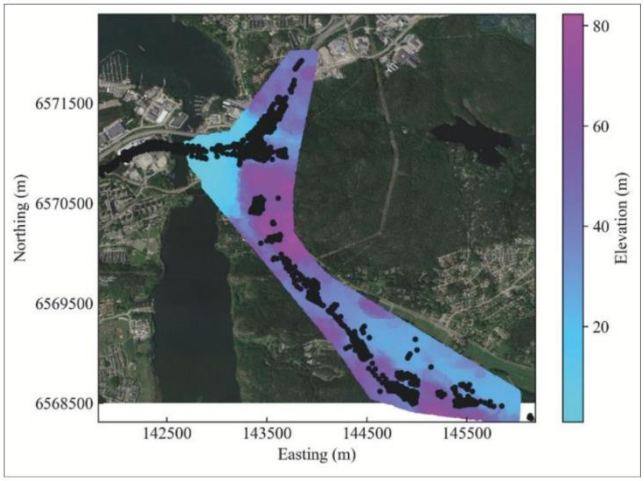
70. Tang, Y., Zhou, J., Yang, P., Yan, J., & Zhou, N. (2017). *Groundwater engineering*. Springer. https://doi.org/ 10.1007/978–981–10–0669–2

71. Taper, M.L., & Ponciano, J.M. (2016). Evidential statistics as a statistical modern synthesis to support 21st century science. *Popul. Ecol.*, 58, 9–29. https://doi.org/ 10.1007/s10144–015–0533–y
72. Uusitalo. L., Lehikoinen, A., Helle, I., & Myberg, K. (2015). An overview of methods to evaluate uncertainty of deterministic models in decision support. *Environmental Modelling Software*, 63, 24–31. https://doi.org/ 10.1016/j.envsoft.2014.09.017
73. Vrugt, J.A., & Robinson, B.A. (2007). Treatment of uncertainty using ensemble methods: comparison of sequential data assimilation and Bayesian model averaging. *Water Resource Research*, 43(1), W01411. https://doi.org/ 10.1029/2005WR004838
74. Waldmann, E. (2018). Quantile regression: A short story on how and why. Statistical Modelling, 18(3–4), 203–218. https://doi.org/ 10.1177/1471082X18759142
75. Wan, L., Zeilar, M., Zhang, S., LeCun, Y., & Fergus, R. (2013). Regularization of neural networks using DropConnect. In *Proc of ICML'13, International Conference on Machine Learning (PMLR),* 28(3), 1058–1066
76. Weerts, A.H., Winsemius, H.C., & Verkade, J.S. (2011). Estimation of predictive hydrological uncertainty using quantile regression: examples from the National Flood Forecasting System (England and Wales). *Hydrol. Earth Syst. Sci.*, 15(1), 255–265. https://doi.org/ 10.5194/hess–15–255–2011
77. Whitman, R.V. (2000). Organizing and evaluating the uncertainty in geotechnical engineering. *J Geotech Geoenviron Eng.*, 126(7), 583–593. https://doi.org/ 10.1061/(ASCE)1090–0241(2000)126:7(583)
78. Wilcox, B.P., Rawls, W.J., Brakensiek, D.L., & Wight, J.R. (1990). Predicting runoff from rangeland catchments: A comparison of two models. *Water Resour. Res.*, 26(10), 2401–2410. https://doi.org/ 10.1029/WR026i010p02401
79. Willmott, C.J. (1984). On the evaluation of model performance in physical geography. In G.L. Gaile, C.J. Willmott (Eds.), *Spatial statistics and models*. *Theory and decision* (volume 40, pp: 443–460), Springer. https://doi.org/ 10.1007/978–94–017–3048–8_23
80. Wu, J., & Zeng, X. (2013). Review of the uncertainty analysis of groundwater numerical simulation. *Chin Sci Bull.*, 58, 3044–3052. https://doi.org/ 10.1007/s11434–013–5950–8
81. Wunsch, A., Liesch, T., & Broda, S. (2020). Groundwater level forecasting with artificial neural networks: A comparison of LSTM, CNN and NARX. *Hydrology and Earth System Science Discussion.* https://doi.org/10.5194/hess–2020–552.

740 82. WWAP (World Water Assessment Programme) (2012). Managing water under uncertainty and risk.
741 *The United Nations World Water Development Report*,1, 417, UNESCO, Paris
742 83. Yager, R.R. (1996). On the inclusion of variance in decision making under uncertainty. *International*
743 *Journal of Uncertainty, Fuzziness and Knowledge–Based Systems*, 04(05), 401–419. https://doi.org/
744 10.1142/S0218488596000238
745 84. Yan, J., Liu, Y., Han, S., Wang, Y., & Feng, S. (2015). Reviews on uncertainty analysis of wind power
746 forecasting. *Renewable and Sustainable Energy Reviews*, 52, 1322–1330.
747 https://doi.org/10.1016/j.rser.2015.07.197
748 85. Liu, Y., Han, S., Wang, Y., & Feng, S. (2015). Reviews on uncertainty analysis of wind power
749 forecasting. *Renewable and Sustainable Energy Reviews*, 52, 1322–1330. https://doi.org/
750 10.1016/j.rser.2015.07.197
751 86. Yan, S., Yu, S., Wu, Y., Pan, D., & Dong, J. (2018). Understanding groundwater table using s statistical
752 model. *Water Science and Engineering*, 11(1), 1–7. https://doi.org/ 10.1016/j.wse.2018.03.00
753 87. Yang, Y., Wang, H.X., & He, X. (2016). Posterior inference in Bayesian quantile regression with
754 asymmetric Laplace likelihood. *International Statistical Review*, 84(3), 327–344.
755 https://doi.org/10.1111/insr.12114
756 88. Yeh, T., Mao, D., Zha, Y., Wen, J., Wan, L., Hsu, K., & Lee, C. (2015). Uniqueness, scale, and
757 resolution issues in groundwater model parameter identification. *Water Science and Engineering*,
758 8(3), 175–194. https://doi.org/ 10.1016/j.wse.2015.08.002
759 89. Yin, J., Medellin–Azuara, J., Escriva–Bou, A., & Liu, Z. (2021). Bayesian machine learning
760 ensemble approach to quantify model uncertainty in predicting ground water storage change. *Science*
761 *of The Total Environment*, 769, 144715. https://doi.org/ 10.1016/j.scitotenv.2020.144715
762 90. Zhang, J., Yin, J., & Wang, R. (2020). Basic framework and main methods of uncertainty
763 quantification. *Materia Problems in Engineering*, 2020. https://doi.org/ 10.1155/2020/6068203
764 91. Zhang, X., Xiao, H., Gomez, T., & Coutier–Delgosha, O. (2021). Evaluation of ensemble methods for
765 quantifying uncertainties in steady–state CFD applications with small ensemble sizes. *Computer &*
766 *Fluids*, 203, 104530. https://doi.org/ 10.1016/j.compfluid.2020.104530
767 92. Zhu, Y., Zabaras, N., Koutsourelakis, & P.S., Perdikaris, P. (2019). Physics–constrained deep learning
768 for high–dimensional surrogate modeling and uncertainty quantification without labelled data. *Journal*
769 *of Computational Physics*, 394, 56–81. https://doi.org/ 10.1016/j.jcp.2019.05.024

B.3 Visualisering av bergtopografi med artificiell intelligens (published)

Visualisering av bergtopografi med artificiell intelligens



Figur 1: Utvärderat djup till berggrunden (DTB) från Jb-sonderingar på projekt Tvärförbindelsen (svarta punkter, överlagda på den konstruerade digitala höjdmodellen och satellitbild från Google earth.

Rumslig fördelning av djup till berggrunden (DTB) är en viktig och utmanande fråga i många geotekniska tillämpningar. På grund av att DTB associeras med säkerhet och ekonomi i jord- och bergkonstruktioner kan generering av mer precisa modeller vara av avgörande betydelse. Med hjälp av resultat från Jb-sonderingar för ett infrastrukturprojekt i Stockholm har vi skapat en optimerad visualiserad 3D -prediktiv DTB-modell via en automatiserad artificiell intelligent datormetod (AI) och jämfört den med den metod som ofta används, det geostatistiska verktyget Ordinary Kriging (OK).

Modellering av bergtopografi – En geoteknisk utmaning

Underjordisk modellering blir alltmer en nödvändig del av tredimensionell (3D) stadsplanering. För att skapa en informativ och användbar modell under jordytan måste olika typer av data kombineras. Planerare och ingenjörer letar efter kunskap om platsen för geologiska diskontinuiteter och gränserna för lösa sediment samt deras geotekniska egenskaper. Därför är rumslig fördelning av djup till berggrunden (DTB, depth to bedrock) av stort intresse för geoteknisk modellering och riskbedömning [1, 4]. Som ett generellt intresse är visualiserade DTB-modeller viktiga verktyg för att beskriva komplexa geologiska förhållandena. Detta innebär att det är en utmanande uppgift att generera noggranna modeller som har inverkan på kostnaderna och riskerna i geotekniska projekt [3].

Geotekniska sonderingar ger endast punktbaserade data. Detta innebär att osäkerheten i prediktiva modeller är stora på grund av flera svårigheter såsom begränsad tillgång till ett helt område, kostnader för undersökningar och avstånd mellan sonderingar. Traditionellt använder sådana förutsägelser ofta olika geostatistiska interpoleringstekniker mellan sonderingspunkterna. Resultat från geofysiska undersökningar kan också korreleras med geoteknisk information, men jämfört med geotekniska sonderingar kan dessa data inte bara vara komplexa och tidskrävande utan också ha logistiska problem. De bör därför utföras av experter som är bekanta med testmetoden. Dessa problem innebär att kvalitén hos en genererad 3D DTB – modell beror på kvaliteten på tillgänglig information om området, den tillämpade metoden samt korrelerade variabler. Därför är det inte alltid klart vilken metod som ger det mest ändamålsenliga och korrekta resultatet.

Geoteknisk artificiell intelligens

Under de senaste åren har artificiell intelligens (AI) visat på anmärkningsvärda beräknings- och inlärningsmöjligheter för att hantera olika typer av komplexa problem. Eftersom DTB-modellering behandlar olika osäkerheter är AI-tekniker lämpliga alternativ för att övervinna begränsningen och förenklingar hos de traditionella geostatistiska metoderna som Kriging [2].

Med hänvisning till efterfrågan på högupplösta 3D-underjordsmodeller var vi motiverade att ta itu med denna utmaning med hjälp av ett utvecklat automatiserat AI-förfarande. Vi valde projektet


Abbas Abbaszadeh Shahri
Tyréns


Chunling Shan
Tyréns / KTH


Stefan Larsson
KTH

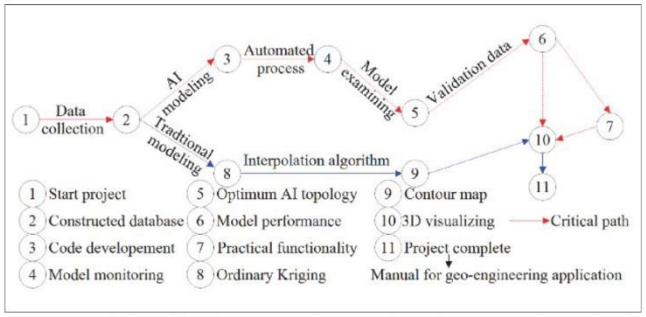
Tvärförbindelsen i Stockholm som inkluderar 1968 jord-bergsonderingar (Jb-sonderingar). Med hjälp av det utvecklade automatiserade schemat visade den identifierade optimala modellen överlägsen prestanda och mer precis DTB-modell jämfört med traditionell OK-teknik. Resultaten påverkar möjligheten att optimera antalet borrpunkter när man använder utvecklade AI-modeller.

Det studerade området

Det studerade området omfattar en cirka 5 km lång sträcka av det 20 km pågående NW-SE-motorvägsprojektet Tvärförbindelsen i Stockholm. Profilen för denna motorväg korsar den befintliga berggrunden som huvudsakligen består av sedimentära gnejser och metavulkaniska bergarter samt pegmatitpassager. I denna studie sammanställdes 1968 stycken Jb-sonderingar (*figur 1*) i området och randomiserades till data som delades upp i 65, 20 och 15 procent för att generera träning-, testning- och validering av data. Inom en del av det studerade området varierar berggrundsnivåerna dramatiskt och vi stötte därför på en utmaning på grund av begränsat antal sonderingar. Denna utmaning övervanns med hjälp av ett intelligent kunskapsbaserat ramverk [1]. En översikt över den konstruerade digitala höjdmodellen (DEM, digital elevation model) för området visas i *figur 1* och detta område med DEM användes för bergrundsmodellering.

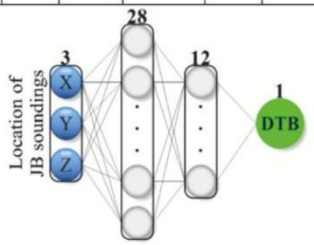
Utveckling av automatiserade metoder för prediktiv AI

Ett neuralt nätverk är ett artificiellt system som består av virtuella abstraktioner av neuronceller. Därför är AI ett försök att simulera en mänsklig hjärnstruktur där datorn kommer att konfigurera på ett sådant sätt att man kan lära sig, replikera procedurer och på så sätt fatta beslut. Detta innebär att efterlikning av den mänskliga hjärnan där AI-tekniker används för att lösa komplexa beräkningsproblem. Den senaste utvecklingen av systemanalyser och de betydande påvisade fördelarna jämfört med traditionella modelleringsmetoder har lett till ökad användning av olika AI -tekniker. För att få den optimala modellen utvecklades därför en automatiserad inlärningsprocess genom en iterativ procedur. Denna metod som använder flera inbäddade interna kapslade loopar kan automatiskt övervaka en mängd olika hyperparametrar för både grunda och djupa neurala nätverk. Den förenklade layouten av processen presenteras i *figur 2*. För mer läsning angående AI rekommenderas [5].



Figur 2: *Layout för den exekverade automatiserade AI -proceduren vid generering av den visualiserade DTB –modellen.*

Optimerad *AI*	R^2			$RMSE_{min}$		
	träning	testning	validering	träning	testning	validering
	0,96	0,96	0,94	6	5,55	6,3



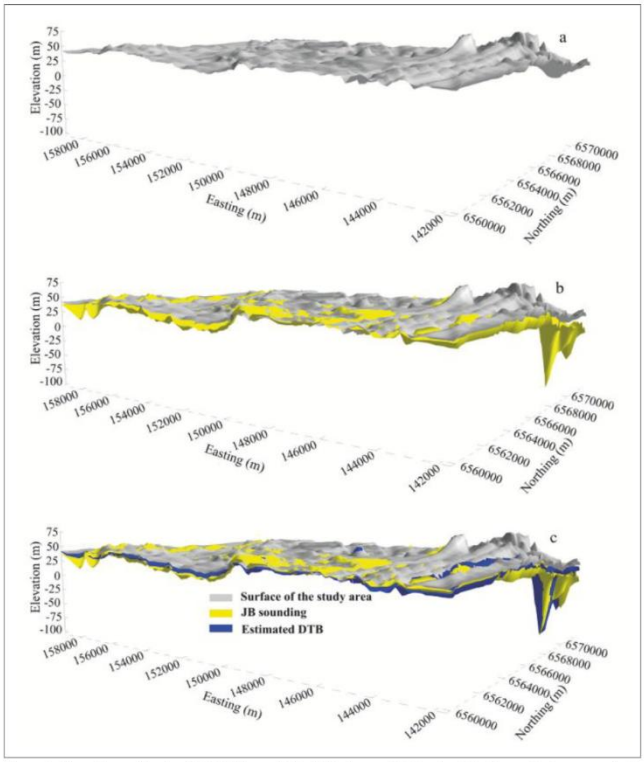
Figur 3: *Resultaten för optimal topologi.*

Resultatet av en automatiserad procedur med olika kombinationer av hyperparametrar visade det minsta felet i en modell med fyra lager med topologi 3-28-12-1 (*figur 3*). I denna optimala topologi visar 3 antalet indatavariabler (rumsliga koordinater för JB-sonderingar), 28 och 12 anger antalet neuroner i två dolda lager, och 1 uttrycker antalet utdata (den förutspådda DBT).

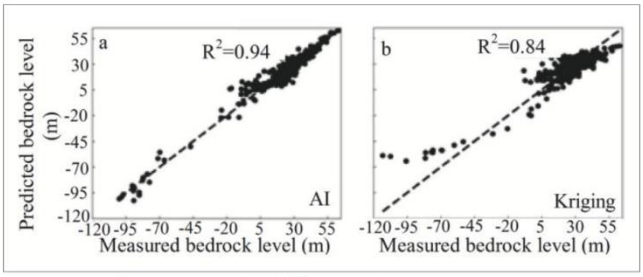
Genererad och visualiserad DTB-modell

Det finns en trend som gynnar användningen av 3D-modeller med möjlighet att kombinera terrängdata och flygfoton. Sådana modeller ger ett visuellt perspektiv på det aktuella området, vilket möjliggör mer precisa tolkningar. Beroende på kvaliteten på de data som används och tillvägagångssättet kan dock noggrannheten och modellens korrekthet variera när det gäller deras förmåga att förhindra konflikt med olika interpoleringsalgoritmer. *Figur 4* visar resultaten av den automatiserade processen för skapandet av den visualiserade 3D -modellen av studieområdet med hjälp av Jb –sonderingarna.

Jämförelsen mellan AI och OK visade på mer precisa uppskattningar med AI och därmed högre förutsägbarhet än OK. I *figur 5* presenteras determinationskoefficienten (R^2) och skillnader mellan de förutsagda och uppmätta värdena. Eftersom högre värden på R^2 och mindre skillnader ger AI-modellen med R^2=0,94 en betydligt bättre förutsägbarhet än OK med R^2= 0,84. *Tabell 1* visar summering av skillnader mellan de förväntade och verkliga bergytenivåer. Den genomsnittliga skillnaden/mätpunkt är 7,4 m för OK och 4,3 m för AI, vilket indikerar att det förutsagda bergytenivåerna ligger närmare det verkliga bergnivån för AI modellering. Det visar att AI har 42 procent förbättring än OK genom att titta på skillnaderna mellan förutsagda och verkliga värden.



Figur 4: *Skapad rumslig visuell 3D DTB-modell (a) Markytan, (b) Resultat från Jb-sonderingarna och (c) Predikterad DTB.*



Figur 5: *Förutsagd kontra uppmätt DTB-modell för AI respektive OK.*

Tabell 1: *Summering av skillnader mellan de förväntade och uppmätta bergytenivåer.*

Metod	Average difference/point (m)	Improvement from OK (%)
OK	7,4	0
AI	4,3	42 %

Slutord

På grund av variationen i bergtopografin är det av stor vikt att ta fram precisa 3D rumsliga prediktiva DTB-modeller för att identifiera de potentiella utmaningarna och riskerna i samband med infrastrukturprojekt. Denna utmaning i den aktuella studien hanterades med hjälp av ett automatiserat AI-system som visade en signifikant förbättring jämfört med traditionell geostatistisk OK-teknik. AI har potential att medföra bättre förutsättningar för industrin eftersom den kan hantera stora mängder data (big data), presentera mer precisa resultat och även kunna uppdateras i realtid. Sådana egenskaper tillhandahåller ett kostnadseffektivt verktyg för geo-ingenjörer i sitt dagliga arbete. Denna fråga kan spela en betydande roll vid design av till exempel erforderlig pålning och jordförstärkning. De tolkade resultaten indikerar att utvecklade AI-modeller är genomförbara, kostnadseffektiva och mer noggranna verktyg för att användas för rumsliga förutsägelser av DTB. ■

Finansiärerna

Projektet har finansierats av Trafikverket genom Branschsamverkan i grunden (BIG), Stiftelsen bergteknisk forskning (BeFo), Tyréns och KTH och har utförts som ett samarbete mellan Tyréns och KTH.

References

[1] Abbaszadeh Shahri, A., Renkel, C., Larsson, S., 2020. *Artificial intelligence models to generate visualize bed rock level: A case study in Sweden.* Model. Earth Syst. Environ. 6, 1509–1528.

[2] Abbaszadeh Shahri, A., Shan, C., Zäll, E., Larsson, S., 2021 *Spatial distribution modelling of subsurface bedrock using a developed automated intelligence deep learning procedure: A case study in Sweden.* J. Rock Mech. Geot. Eng. 13(6), 1300-1310.

[3] Mey, J., Scherler, D., Zeilinger, G., Strecker, M.R., 2015. *Estimating the fill thickness and bedrock topography in intermontane valleys using artificial neural networks.* J. Geophys. Res. Earth Surf. 120, 1301–1320.

[4] Sundell, J., Rosen, L., Norberg, T., Haaf, E., 2015. *A probabilistic approach to soil layer and bedrock-level modelling for risk assessment of groundwater drawdown induced land subsidence.* Eng. Geol. 203, 126–139.

[5] Marsland, S., 2018. *Machine learning: An algorithmic perspective*, Taylor & Francis Group, CRC Press.

Bygg & teknik på Grundläggningsdagen!

Kom gärna förbi vår monter på Stockholmsmässan 17 mars.

Välkomna!

B.4 3D modelling and uncertainty analysis of subsurface bedrock using a hybrid automated deep learning approach-A case study in Sweden (first draft)

Abbas Abbaszadeh Shahri[1,2]
Shan Chunling[1,3]
Stefan Larsson[3]
[1]Division of Rock Engineering, Tyréns AB, Stockholm, Sweden
[2]Johan Lundberg AB, Uppsala, Sweden
[3]Division of Soil and Rock Mechanics, KTH Royal Institute of Technology, Stockholm, Sweden

Abstract
The interest in creating high-resolution 3D subsurface geo-models through multisource retrieved data i.e., borehole, geophysical techniques, geological maps, and rock properties for emergency responses is progressively increasing. However, dedicating accurate, meaningful, and thus interpretable 3D views from integrated such heterogeneous nonlinear data require developing new methodology for convenient post-modeling analysis. Therefore, in the current paper a hybrid automated deep learning-based approach for 3D modeling of subsurface geological bedrock using multisource data is proposed. The uncertainty was quantified using a state-of-the-art novel ensemble randomly automated deactivating process implanted on the stored weight database. The applicability of automated process in capturing the optimum topology with emphasizing on processing flow for description and analysis of 3D subsurface geo-modelling is then validated through the laser scanned bedrock level data in Sweden. In compared to intelligent quantile regression and traditional geostatistical interpolation algorithms, the accuracy of proposed hybrid approach showed high competence for visualize and post-analyze the 3D subsurface model. Due to using integrated multi-source data, the presented approach and created 3D model can be a representative reconcile of the geoengineering context.
Keywords: 3D subsurface bedrock, deep learning, automated process, uncertainty, Sweden

1. Introduction
In geoengineering practice, each drilled borehole is a type of geological diagram that can be used for stratigraphical description such as stratum thickness or lithology characteristics. Accordingly, the retrieved information from a series of boreholes in a specified area can lead to analysis and decision-making of subsurface conditions and mapping for various underground projects. Furthermore, the collected borehole data due to diverse sources need to be unified according to defined standards in each country. However, due to inability of borehole data in explaining the laterally distribution (Mayoraz et al., 1992) as well as significant increase of the volume of the datasets even if after unification, the result of geological model often conflicts with geological knowledge that requires for continually updating in the later stage (Zhang et al., 2020). Three-dimensional (3D) modeling mathematically expresses the digital coordinate-based representation of any surface of an object through specialized software or developed codes. Accordingly, the capacity of digital views of objects practically have been implemented in many site characterizations through subsurface formations and corresponding associated features in geoengineering applications (Houlding, 1994; Hack et al., 2006; Dong et al., 2015; Abbaszadeh Shahri et al., 2021a). Therefore, spatial distribution and maybe interrelationships of subsurface objects can intuitively be reflected in 3D perspective based on unified borehole and other surveyed datasets. This implies on great advantages of 3D geo-models in subsurface monitoring and thus corresponding associated attributed information for underground utilization and urban environment. Such computer models for the subsurface applications can transmit the represented objective to the visualizer with superior level of detail tied with in compare with 2D. However, 3D geospatial information always has been a challenge due to a variety of data, resolutions and required details (Lee and Zlatanova, 2007). In geoengineering practices, extension of a specified geological unit (e.g., rocks, lithological strata, mineralized area, hazard and risk, foundation design, tunnel routing, building, and planning) commonly is determined through borehole data on 2D cross-sections. To conceptualize the spatial patterns of subsurface characteristics, 2D geo-models need to be developed in 3D (Anderson et al., 2015; Mallet, 2008) complemented using supplementary test data, where the more information,

the more accurate inference result (Günther, 2003; Perrin et al., 2005; Do Couto et al., 2014; Wu et al., 2015; Thornton et al., 2018; Abbaszadeh Shahri et al., 2020, 2021a, b). Nowadays, with the advent of huge of amount of produced geo data such 3D models can be used as an integrated carrier of geoengineering and environmental big data.

Interpretation of sparse laterally distributed parameter in a geo-model from vertically constrained borehole data is a difficult obstacle (Mayoraz et al., 1992). Therefore, integrating the different qualified borehole data with numerical algorithms (Apel, 2006; Mallet, 1992) or cognitive interpretative approaches (Caumon et al., 2009; Schaaf et al., 2020; Ghaderi et al., 2019; Zhan et al., 2020) commonly are used to interpolate between observational points for the incorporation of expert geological knowledge. In this view, 3D interpolation of planar meshes (Xu and Tian, 2009), surface mesh generation (Frank et al., 2007) and triangulation algorithms (Chew, 1989) are considered as the most appropriate alternatives to handle discrete bore hole data (Abbaszadeh Shahri et al., 2020). However, applicability of these methods due to several challenging interactive human-computer operations as well as time consuming process and maybe inapplicability in handling big data is still essential in the current geoengineering modeling methods (e.g. Turner, 2006; Niu et al., 2017; Kumar et al., 2019; Lakshmanan, 2012).

To tackle the difficulty of big data (Chen and Lin, 2014), application of skilled artificial intelligence techniques (*AIT*) in terms of shallow and deep neural learning networks (S*NLNs* and *DNLNs*), machine learning (*ML*), hybrid models and evolutionary algorithms have been emerged as a powerful tool in across all geoengineering problems to describe physically meaningful relationships within geoscientific data (e.g., Karpatne et al., 2018; Spina, 2019; Abbaszadeh Shahri et al., 2021a, b).The *DNLNs* due to characterized features such as creating transferable solutions and learnability from high-level data attributes (LeCun et al., 2015; Schmidhuber, 2015) are one of the most used systems for modeling (Toms et al., 2020; Abbaszadeh Shahri et al., 2020 and 2021a). Accordingly, the generated 3D visualized models using *DNLN* due to analytical capabilities can dedicate higher resolution, more flexible and vigorous tool than *GIS* and *CAD* systems (Mallet, 1992). This implies that in the current digital era developing scientific conceptual and 3D quantitative perspective of subsurface features for geoengineering applications through *DNLN*-based techniques are prime of concerns. However, there is no straight-forward answer on amount of actual enough data for training the system. The reason is referred to two-folded task of the *DNLN*, as it needs to learn about the domain through the training algorithm from scratch and thus many parameters to tune (Abbaszadeh Shahri et al., 202, 2021a). Therefore, the robustness of *DNLN*-based predictive models should be evaluated in terms of different accuracy performance criteria and uncertainty quantification (*UQ*) analysis (Morgan and Henrion, 1990; Yan et al., 2015).

Europe due the dense urbanization and variety of geological conditions needs for exploring sustainable use and management information of the subsurface in urban planning and development (Mielby et al., 2016). To overcome on challenges associated with sustainable cities and environmental regulations, the appropriate use of the subsurface in response to technology and socioeconomic demands have been highlighted (Athanasopoulou et al., 2019; Hooimeijer and Maring, 2018). From geoengineering point of view, depth to bedrock (*DTB*) corresponding to the thickness of the sediments above the bedrock is a crucial factor for the proper subsurface utilizing. In Sweden due to pursuing the *EU* rules and abundant of created infrastructures (e.g. transport tunnels, roads, railways), the *DTB* is an important concern, where geotechnical knowledge on it can provide critical insights into the influence of the stability of the structures and transport of contaminants through gradient on bedrock surfaces (Abbaszadeh Shahri et al., 2021b). Therefore, producing accurate 3D visualized spatial *DTB* model not only facilitates the interpretation of sparse geotechnical measurements but also provide momentous tool for identifying the optimum solutions and risk assessment (e.g., Gomes et al., 2016; Wei et al., 2016; Ghaderi et al., 2019; Abbaszadeh Shahri et al., 2021a, b; Yan et al., 2020). However, in geoengineering projects producing high resolution 3D spatial *DTB* predictive model due to incorporated uncertainties with geotechnical-based limitations (e.g., access to an entire area, cost, distance between the soundings) not only requires for different combination of data types but also is a critical task (Gomes et al., 2016; Wei et al., 2016; Abbaszasdeh Shahri et al., 2020, 2021a).

Traditionally, the produced geoengineering models reflecting the *DTB* have been generated by means of geostatistical techniques. These models due to the lack of computational system adopted for big geo-data and limited graphical power in complex problems often have been performed in less complex and small-scale areas. This point shows why developing modern computational *AIT*-based scheme for producing a more accurate and

high-resolution 3D visualized subsurface model is highly motivated. Such demand with the advent of *AIT* technology and code developing can provide powerful new application to the requirements of the geo-engineers to address the modeling obstacles.

The interest in developing 3D models (buildings or undergrounds) and geospatial data analysis in emergency management systems for field workers and decision makers is progressively increasing. Accordingly, in the current paper a developed automated *DNLN* procedure for generating the 3D spatial predictive of *DTB* for Sweden as a necessary part of 3D urban planning is addressed. This model was created using geolocation of 644 datasets of different soil-rock soundings and geophysical investigations. In addition of using the randomized validation data, a number of 75715 scanned rock surface datasets also were separately fed for predictability assessment. Detailed analyses in compare with quantile regression (*QR*) (Bremnes, 2004) and geostatistical ordinary kriging (*OK*) (Wackernagel, 1995) showed that the proposed automating approach and corresponding optimum *DNLN* topology properly can dedicate a high-resolution 3D spatial predictive subsurface *DTB* map. The analyzed *UQ* also indicated competitive performance in using the optimum topology led to superior performance in compare with *OK* and *QR*.

2. Study area and data source

The study area is located in the south-west of Sweden, close to Gothenburg city. The geotechnical data used in this study is collected along highway E20 between Bälinge and Vårgårda. Swedish transport administration is planning to build a new highway close to the old E20 by considering the currently problems with speed, traffic safety and environment influences. In the western part close to Bälinge, the proposed highway crosses rock outcrops, moraine, glacial clay, glacial deposits, and swell sediments. The bedrock along the planned road consists mainly of gneiss and it has a higher content of dark minerals such as amphibole and biotite. Pegmatite, a granitic coursed-grained rock is found along almost the whole stretching of road route. The bedrock is general heterogeneous and complex compound. But at the end of the road route the bedrock becomes more homogenous and consists mostly of the so-called Vårgårda granite (Åsander, 2015). The fracture zones in the gneiss in this area is generally in east-west direction. However, close to Bälinge the fracture zones are in a northwest-southeast direction and close to Vårgårda they are more in northeast-southwest direction. The fracture planes are general dipping to the south which can possibly cause stability problem if rock cuttings are performed (Åsander, 2015).

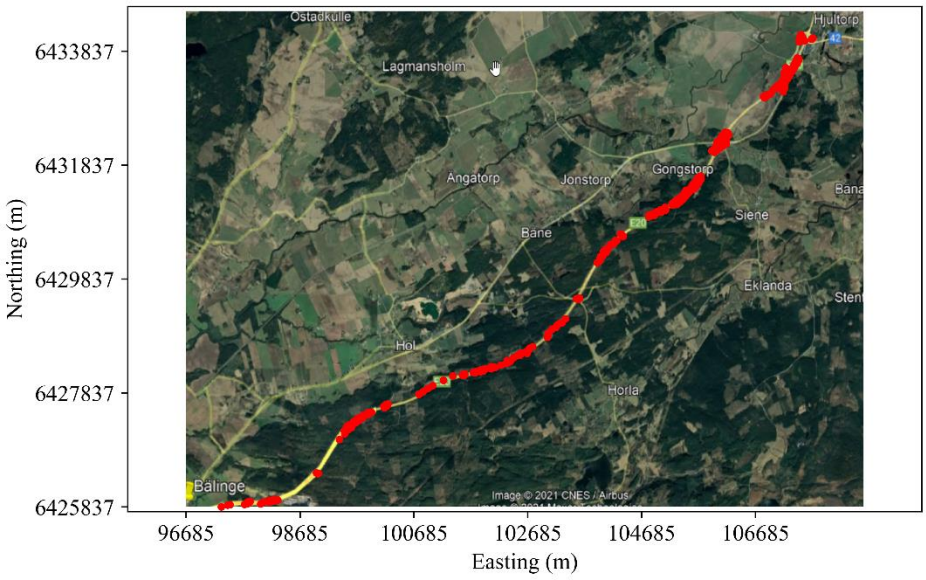


Figure 1. Location of project area in Stockholm with the location of acquired geotechnical data (in red) superimposed on google earth image

3. Model development process

3.1. Overview of *DNLN* structure

The concept of *AITs* is a simplified imitative learnable layout of human brain structure that aims to increase the computational power through the embedded connective processing elements. Referring to approved advantages of *AI*-based systems over traditional modelling approaches, they are largely growing in computer vision of

geoengineering applications. *DNLN* is a subcategory of *AIT* that without being explicitly will scan the data to search combinable features for faster learning. This ability implies on great performance of *DNLN* with unstructured data (different formats such as texts, pictures, pdf, …) as well as exploring new complex features that human can miss and thus more capacity over *ML* to execute feature engineering (LeCun et al., 2015). Such process significantly will save the time. Accordingly, the analysis of *UQ* in *DNLN* can be described in the context of input variabilities, assumptions and approximations, measurement errors as well as sparse and imprecise data (Yan et al., 2015). However, all sources of uncertainty using these techniques for value-based judgements may not be quantifiable (Morgan and Henrion, 1990).

As presented in Figure2, in a fully connected configuration of *DNLN* the output of the j^{th} neuron in the k^{th} hidden layer at the t^{th} iteration, $o_j^k(t)$, subjected to activation function, *f*, is defined as:

$$o_j^k(t) = f\left(\sum_{i=1}^{n_k-1} w_{ij}^k x_i^{k-1}(t) + b_j^k\right) \quad (j \le n_k) \tag{1}$$

Where; x_i is the i^{th} signal input with a connective weight of w_{ij} to the j^{th} neuron in the k^{th} hidden layer. n_k shows the number of neurons in the k^{th} layer, and b_j^k denotes the bias which shifts the summed signals received from the neuron.

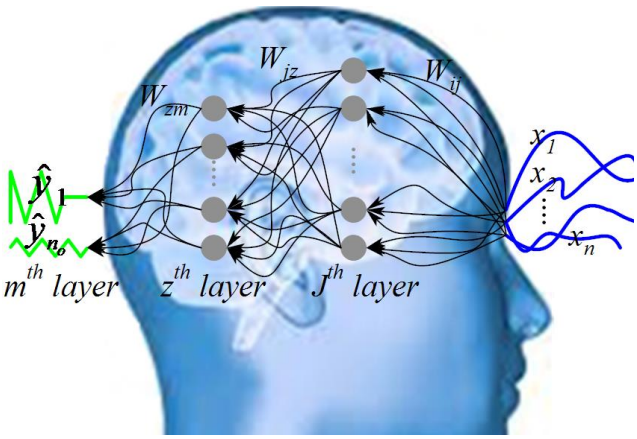


Figure 2. Simple configuration of *DNLN* architecture

The weights and biases are then updated using corresponding error of the i^{th} neuron in the *k*th hidden layers, $\rho_i^k(t)$, to minimize the prediction error by:

$$\Delta w_{ij}^k(t) = \eta_w \left[\dot{f}\left(x_i^k(t)\right)\sum_j \rho_j^{k+1}(t) w_{ji}^{k+1}(t-1)\right] o_j^{k-1}(t) + \alpha_w \Delta w_{ij}^k(t-1) \quad (k = 1, 2, \dots, m-1) \tag{2}$$

$$\Delta b_i^k(t) = \eta_b[\dot{f}\left(x_i^k(t)\right)\sum_j \rho_j^{k+1}(t) w_{ji}^{k+1}(t-1)] + \alpha_b \Delta b_i^k(t-1) \tag{3}$$

Where; $\dot{f}\left(x_i^k(t)\right)$ is the first derivative of $f\left(x_i^k(t)\right)$ with respect to $x_i^k(t)$. α_w and α_b express the momentum constants that determine the influence of the past parameter changes on the current direction of movement in the parameter space, and α_w usually varies within [0.1, 1] intervals and is used to avoid instability in the updating procedure. η_w and η_b represent the learning rates.

the total error for the entire network, *E*(*t*), prediction error, $\varepsilon(t)$, and the outcome of the *l*th neuron in the *m*th output layer, $\hat{y}_l$, is then calculated using the updated weight by:

$$E(t) = \frac{1}{2}\sum_{j=1}^{n}\left(d_j(t) - o_j^k(t)\right)^2 \tag{4}$$

$$\hat{y_l}(t) = \sum_{i=1}^{n_m-1} w_{ij}^m x_i^{m-1}(t) \quad (l \le n_o) \tag{5}$$

$$\varepsilon(t) = y - \hat{y}(t) \tag{6}$$

Where; $d_j(t)$ denotes the desired output of neuron *j* at the t^{th} iteration. *y* and n_o represent the actual output and the number of neurons in the output layer respectively.

3.2. Configuring the hybrid automated model monitoring

In the current paper as presented in Figure 3, a hybrid automated approach comprising two blocks and several embedded inner nested loops accompanying with switch cases for subsurface 3D modelling was developed. Using block A (Figure3) the procedure monitors variety of combined internal hyper parameters as presented in Table (1). This was executed because in *DNLN* the need for adequate computing power and data for learning should be considered (Bengio, 2009). Such integrated components then allow system for monitoring different

combinations of internal hyper parameters not only makes learning faster but also minimize to get stuck in local minima or overfitting problems (Abbaszadeh Shahri et al., 2020). To prevent the early convergence, a two-step termination criterion was considered in which if the root mean square error (RMSE) as the priority didn't achieve then the number of iterations (set to 1000) is replaced. Accordingly, numerous captured shallow and *DNLN* even with similar topology but different hyperparameters can assist to improve the generalization of selected model for new examples (Abbaszadeh Shahri et al. 2021b; Kriegeskorte and Golan, 2019). To control the stability of the achieved topology, the *UQ* was estimated using a proposed state-of-the-art ensemble-based automated randomly deactivated weight database (*ARDCW*) approach as presented in block B (Figure3) (Abbaszadeh Shahri et al., 2021c). Deactivating approach solely is applied on the weight database of the optimum topology. Subsequently, the model automatically is retrained using switched off weight components to monitor the variation of predicted output subjected to different scenarios. Therefore, the challenges of computational costs for multiple training of complex topologies through different optimizers are overcome (Abbaszadeh Shahri et al., 2021c).

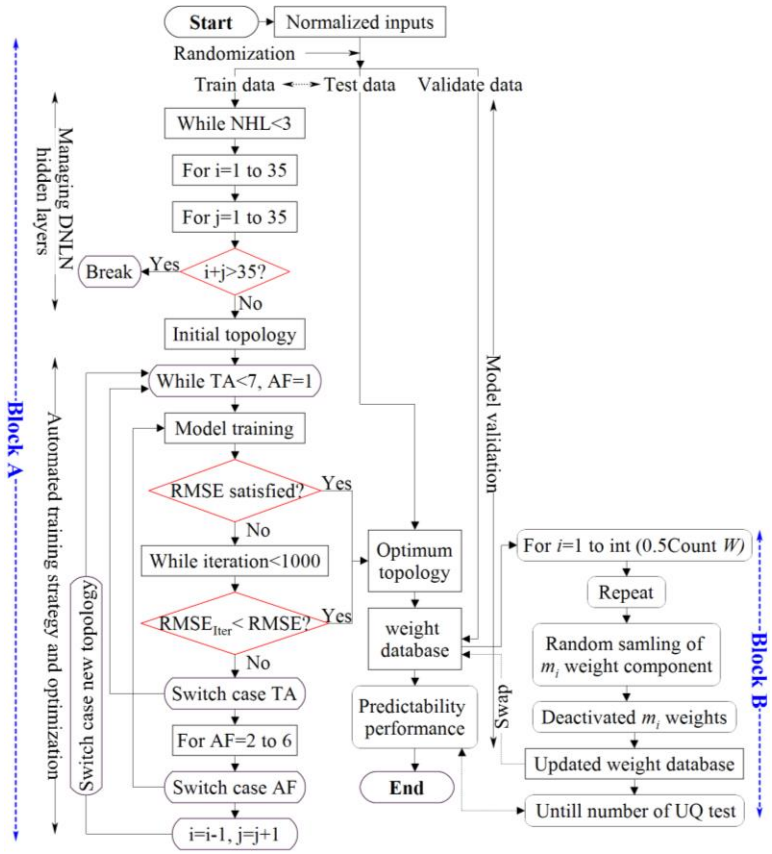


Figure 3. The layout of proposed hybrid automated approach to capture the optimum topology and *UQ* analysis

Table (1). Used hyperparameters to monitor the optimum model

Training algorithm (*TA*)	Activation function (*AF*)	Maximum number of hidden layers (user defined)	Maximum number of used neurons (user defined)
QP, *CGD*, *AM*, *QN*, *LM*, *SGD*	Sig, HyT, Lin, Relu, SS	3	35
η: 0.7 with a step size domain within [0.001,1.000] interval			
Note➔ *QP*: Quick propagation; ***CGD***: Conjugate gradient descent; ***AM***: Adaptive momentum; ***LM***: Levenberg-Marqurdt; ***QN***: Quasi Newton; ***SGD***: Stochastic gradient descent; ***Sig***: Sigmoid; ***Hyt***: Hyperbolic tangent; ***Lin***: Linear; ***Relu***: Rectified linear unit; ***SS***: Softsign			

3.3. Results and created 3D model

Considering the combination of employed hyperparameters, the optimum model was identified among more than 2000 monitored and ranked topologies with different internal characteristics. A series of the results subjected to implemented hyper parameters (Table 1) in terms of the variation of minimum *RMSE* for training stage are shown in Figure 4a. Using this process, the risk of overfitting, early convergence and get stuck in local minima

will be minimized because even similar topologies but with different hyperparameters are also examined. The overtraining problem also is detected when no improvements in the accuracy after a certain number of epochs can be observed (Abbaszadeh Shahri et al., 2021d). Furthermore, in automating procedure the best results among three runs of each model are saved. Referring to Figure 4a, the *DNLN* topology with structure of 3-20-15-1 subjected to *QN* and *Hyt* can be selected as optimal. Subsequently, the predictability of achieved topology using the randomized datasets is reflected in Figure 4(b-d).

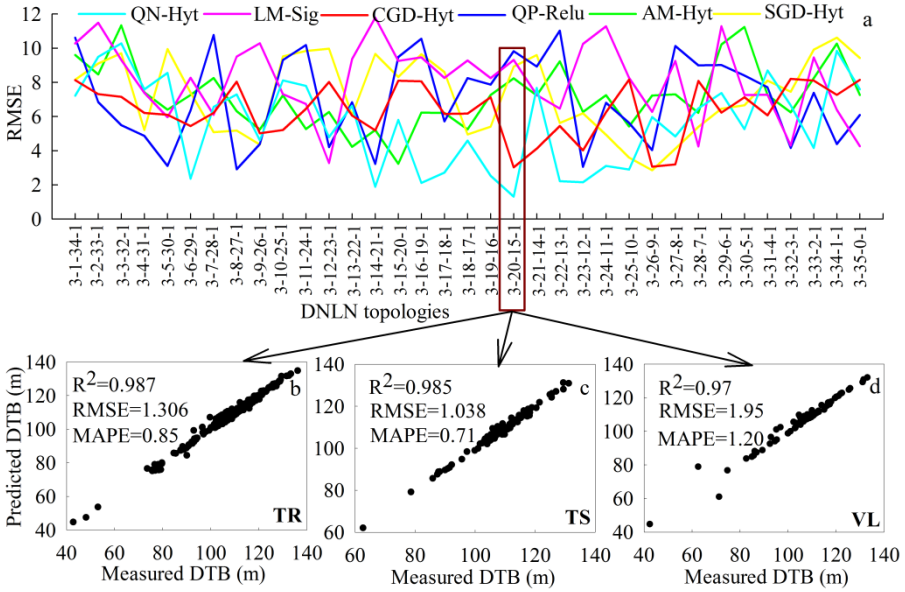


Figure 4. Identifying the optimum topology for prediction of *DTB* through variation of *RMSE* (a), and corresponding predictability using training (b), testing (c), and validation data (d) (*TR*: training data; *TE*: testing data; *VL*: validation data; *MAPE*: mean absolute percentage error)

Creating and simulation of the 3D model using *AIT*s is an important priority for subsurface geoengineering purposes (e.g., Zhou C et al., 2019; Zhang et al., 2020; Abbaszadeh Shahri et al., 2020, 2021b). This is because the traditional method relies on the knowledge and experiences of experts in the selection of assumptions, parameters, and data interpolation methods which are subjective and limited (Randle et al., 2019).
In Figure5, the step-by-step 3D modeling of the subsurface spatial *DTB* distribution for the study area in compare with the true scanned data is reflected. The rock outcrops then can be identified through the overlaid generated ground surface and spatial scanned *DTB* leading to a high resolution predictive 3D subsurface model with an adequate accuracy in geoengineering projects. In Figure5d, the result of analyzed *UQ* in terms of the predicted error using the *ARDCW* (Abbaszadeh Shahri et al., 2021c) for the achieved optimum *DNLN* topology is presented, where the higher errors in predicted *DTB* depict greater uncertainties at data points.

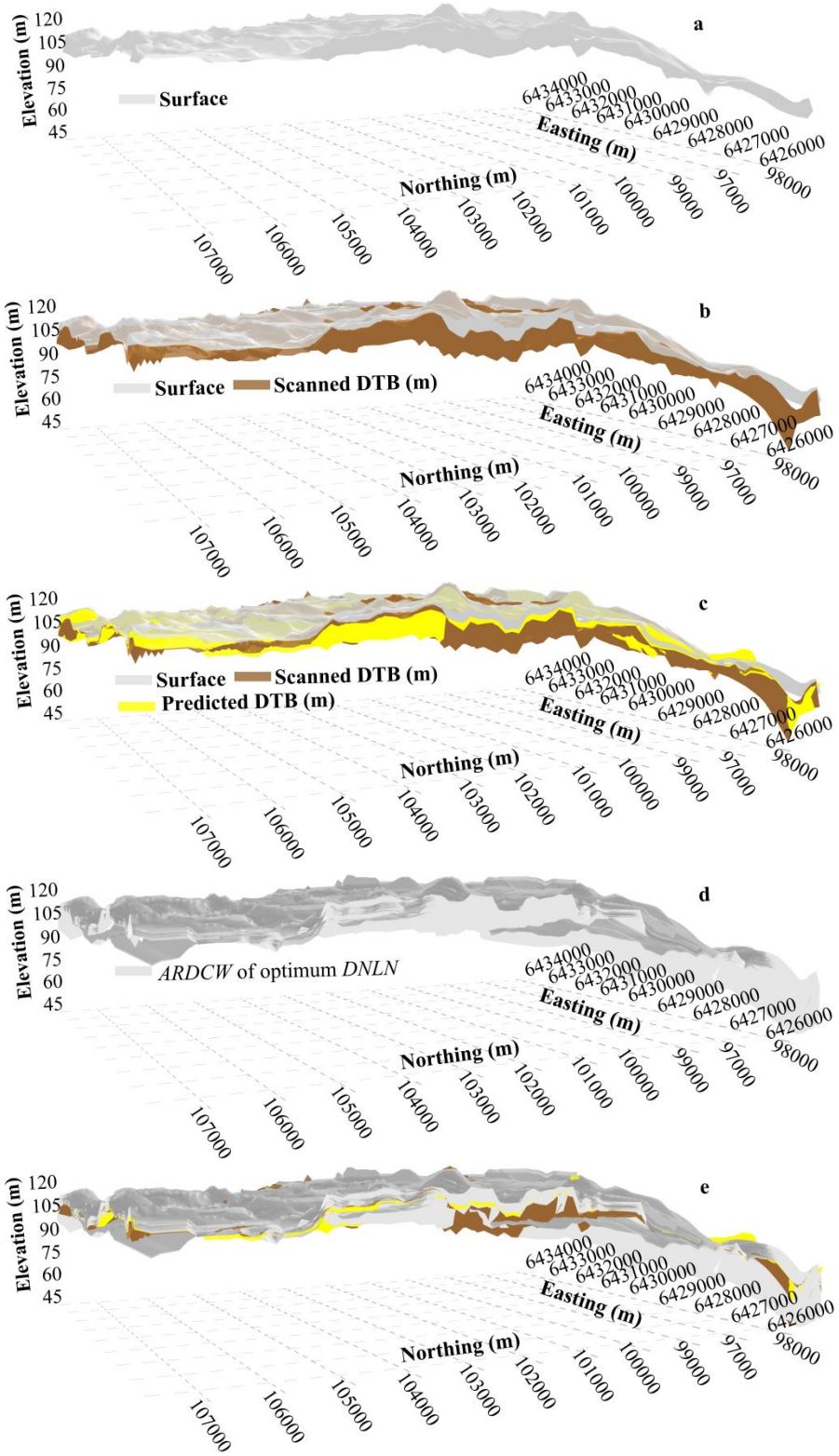


Figure5. The created 3D model of study area comprising the (a) ground surface, (b, c) incorporated the lateral distribution of scanned and predicted *DTB*, (d, e) estimated *UQ* using *ARDCW* accompanying with the scanned and predicted *DTB* using automating system

4. Discussion and validation

3D digital models provide more facilitates compared to traditional blueprints, as they can serve a way to improve communication between engineers, constructors, and their clients. Such models can play significant role when the project needs to undergo bidding and cost evaluation to list down all needed resources. This implies that for

the success of a geoengineering project 3D modeling due to ability in capturing the complexities of subsurface characteristics, improving the visualization, and thus assisting for better analysis of designs are important tools. Furthermore, 3D models can convey the use and the advantages of a designed project easier especially to those who have little knowledge of construction and engineering.
These descriptions show that the field practicability of each generated 3D model before use should carefully be evaluated in terms of different accuracy performance metrics. In this study the discussion on generated 3D model were presented through comparison of different analytical and evidential metrics with *QR* and traditional geostatistical ordinary kriging (*OK*).

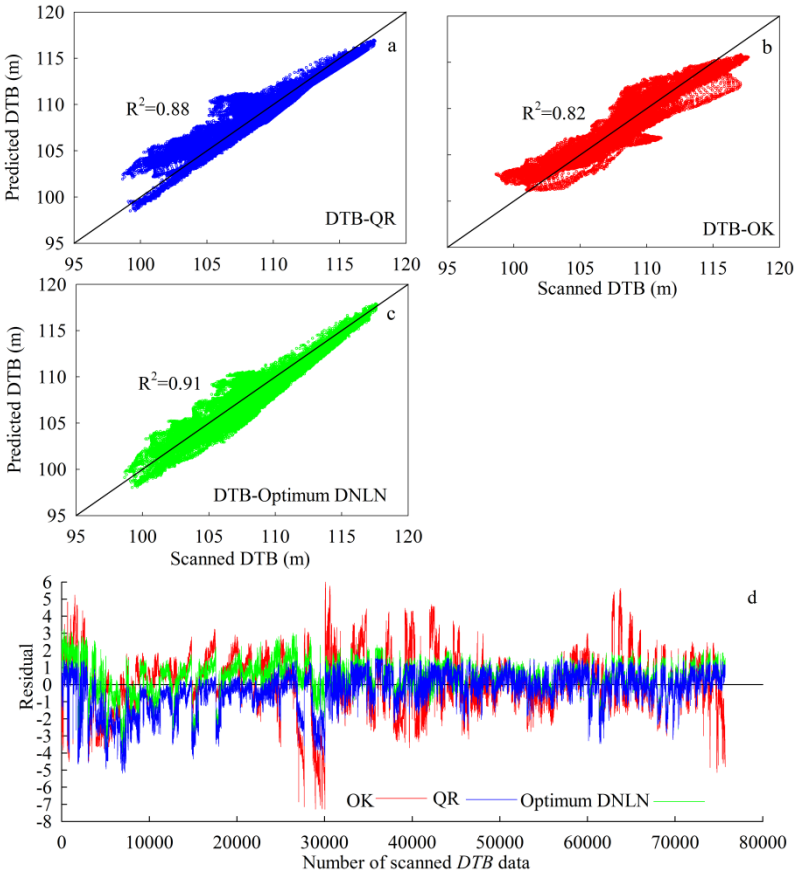
4.1. Precision performance

Among the traditional geostatistical interpolation algorithms, the *OK* can serve spatial estimation when a variogram is known (Wackernagel, 1995). However, *OK* due to using the interpolation can only predict the *DTB* values within the range of observations while the *QR* and *DNLN* because of regression and considering the minimal squared error are able to predict even in outside of the tabulated domain. In mathematical point of view, an n^{th} order polynomial for (n+1) *DTB* points can be obtained which passes exactly through all (n+1) data. This is the main reason why in training process the *OK* subjected to different variogram functions showed high R^2 (Table 2). Accordingly, during the training process the regression may not be as accurate as interpolated *DTB* within the observed domain but provide much better predictions in the range of below and beyond the observations. This reasonably describes why R^2 of predicted data in *OK* significantly is varied (Table 2). Moreover, since *OK* statistically contains several parameters that can be justified by the data, the achieved results can be referred to overfitting (Everitt and Skrondal, 2010; Christian and Griffiths, 2017) while variability of variagram parameters made different predictions on the same dataset. This implies that the *OK* subjected to big data does not generalize well, and the predicted new values provide closely or exactly response to a particular set of data.

Table (2). Summary of the *OK* results subjected to 8 to 14 lags and different variogram functions

Variogram function	R^2 validation	R^2 prediction
linear	[0.92, 0.97]	[0.73, 0.81]
power	[0.89, 0.96]	[0.23, 0.78]
Gaussian	[0.88, 0.90]	[0.08, 0.23]
spherical	[0.91, 0.96]	[0.76, 0.80]
exponential	[0.88, 0.99]	[0.79, 0.82]

Moreover, for proper evaluation the model never should be trained with the entire dataset because of too much learning and saturation (Burnham and Anderson, 2002; Tetko et al., 1995). The results for such training lead to capture the noise in the data in addition to the signal and thus can cause wild fluctuations in the model that does not represent the true trend (Abbaszadeh Shahri et al., 2020). The predictability of applied models in terms of scatter plots is given in Figure6.



Fiure6. Predictability of (a) *QR*, (b) *OK*, (c) optimum *DNLN*, and (d) calculated residuals subjected to scanned *DTB*

The predictive models analytically are dependent to the precision and accuracy of applied techniques. This implies that the acceptability of implemented method can therefore be judged on the sizes of the observed errors relative to a defined total allowable error (*TAE*) (Lakshmikantham and Sen, 2005). According to Figure7 and using the standard deviation (*SD*) of scanned *DTB*, a value of 6.811 for the *TAE* for the applied model is achieved. Considering the *SD* of each model in predicted *DTB*, values of 6.07, 5.40 and 6.07 corresponding to *QR*, *OK* and *DNLN* is calculated. Therefore, in compare with the calculated residuals (Figure 6d), the *OK* has exceeded from the *TAE* which can be interpret as instability of this model in handling the big data. The precision of *TAE* then was supported using *SD* of calculate residuals with values of 1.15, 1.45 and 0.98 for *QR*, *OK* and optimum *DNLN* respectively.

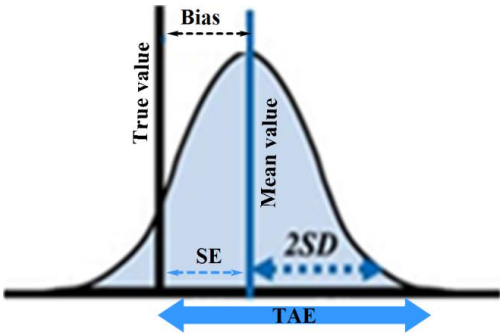


Figure7. Defining the *TAE* (*SD*: standard deviation, *SE*: standard error)

According to Akaike's information criteria (*AIC*) (Akaike, 1970) that penalizes appended terms to a model, additional variables that reduces *RMSE* and increases R^2 are not appropriate to help guide the model choice. The goal is to find the model that minimizes *AIC* and for the regression can be calculated as:

$$AIC = 2P + n\ log\left(\frac{RSS}{n}\right)$$

Where; *P* is the number of variables, *n* denotes the number of records and the models with *k* more extra variables are penalized by 2*k*. *RSS* as the residual sum of squares for *OK*, *QR* and optimum *DNLN* are 160000, 110899.5 and 81984.45 respectively. In this point of view the second term of the *AIC* for *OK*, *QR* and optimum *DNLN* models stand for 24602.68, 12549.58 and 2615.91 respectively. Therefore, the high achieved R^2 and thus low *RMSE* in training and validation of *OK* (Table 3) cannot lead for better performance than *QR* and optimum *DNLN*.

4.2. Statistical error analysis

In geoengineering purposes statistical analysis is an essential tool for the reliable interpretation of experimental results. Furthermore, the growing importance of decisions and opinions based on data can pivotally assess the quality of analyses and thus statistical error metrics can be used to decrease the decision risks. The results of the generated 3D models in terms of mean absolute percentage error (*MAPE*), variance account for (*VAF*), index of agreement (*IA*), general standard deviation (*GSD*), coefficient of determination (R^2), *RMSE* and difference ratio (*DR*) are reflected in Table (3). Accordingly, the model with lower *MAPE*, *DR*, *GSD*, and *RMSE* as well as higher values of *VAF*, *IA* and R^2 indicate for better performance.

Table (3). Applied statistical error criteria to evaluate and compare the employed models

Criteria	Equation	Optimum DNLN		QR		OK	
		VL	Pred	VL	Pred	VL	Pred
MAPE	$\frac{1}{n}\times\left[\sum_{1=1}^{n}\left\|\frac{p_i-o_i}{t_i}\right\|\times 100\right]$	1.20	0.8	1.7	0.8	3.4	1.2
RMSE	$\sqrt{\frac{1}{n}\times\left[\sum_{i=1}^{n}[(p_i-o_i)^2]\right]}\times\frac{100}{\bar{o}}$	1.95	1.033	3.206	1.19	4.47	1.69
VAF	$\left[1-\frac{var(p_i-o_i)}{var(p_i)}\right]\times 100$	0.98	0.90	0.93	0.85	0.85	0.46
GSD	$\frac{RMSE}{\bar{p}_i}$	0.0183	0.0096	0.0299	0.011	0.042	0.016
DR	$\frac{\bar{p}_i}{\bar{o}_i}$	1.001	0.996	1.005	1.003	1.001	1.000
IA	$1-\frac{\sum_{i=1}^{n}\|p_i-o_i\|}{\sum_{i=1}^{n}(\|p_i-\bar{o}_i\|+\|o_i-\bar{o}_i\|)}$	0.99	0.97	0.98	0.96	0.97	0.90
R^2	$1-\frac{Var\ (model\ error)}{Var\ (o_i)}$	0.97	0.91	0.95	0.88	0.90	0.82

4.3. Comparing 3D visualized models

In geoengineering applications, the interest of 3D models for emergency management responses is progressively increasing. Employing 3D models can help the geoengineers to determine early on the process all the necessary material involved for the construction of the project. Referring to high computational cost in available traditional modelling techniques, application of modern systems such as *DNLN* are motivated to represent and analyze 3D geospatial data for geoengineers, field workers and decision makers (Abbaszadeh Shahri et al., 2021b). Accordingly, in subsurface projects adding different level of stratigraphic or geological details are of great demand in application allowing visual representation of the interested geoengineering features (e.g. Mayoraz et al. 1992; Lemon and Jones 2003; Frank et al., 2007; Caumon et al., 2009). Such 3D predictive geospatial models can show the digitized geometric or topological objects based on the available featured factors leading to capture the complexities of buildings and infrastructures that closely resemble the view of the finished project. Accordingly, using 3D models the geoengineers can determine how to design the project based on the involved natural and man-made aspects which leads for more proper structural element adoption and thus making safer and stronger projects. An appropriate tuned *DNLN* can easily be adopted with new data to improve the diagnostic performance and computer vision modifications. It also can be supplied with future acquired data to reflect more details of the subsurface features. In this point of view, the predictability of applied models in

creating spatial 3D views supplemented by residual contours in the outlines of the area subjected to 75715 scanned *DTB* were compared and presented in Figure 8.

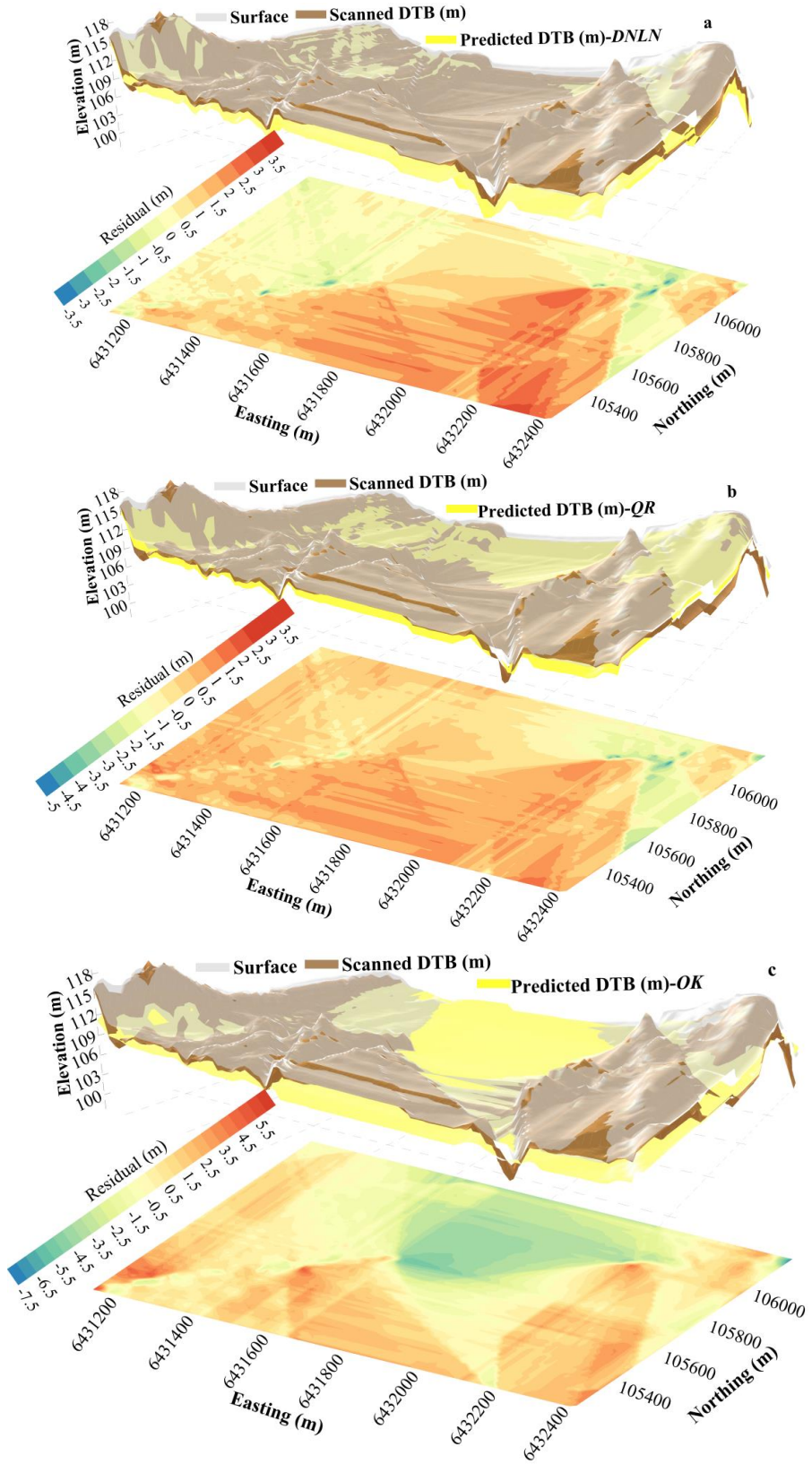


Figure 8. 3D visualizing of applied methods by incorporating the predictive models and scanned *DTB* data for (a) optimum *DNLN*, (b) *QR*, and (c) *OK*

4.4.2. Estimated *UQ* intervals in scanned *DTB*

Despite the various 3D suggested advanced analysis methods in geoengineering problems (e.g., Murdie et al., 2015; Zhang et al., 2021; Abbaszadeh Shahri et al., 2021c), the analyzed *UQ* due to used assumptions and propagated error in applied computational modelling technique can provide significant differences. Moreover, the problem associated with the generalization from a single study, difficulties in obtaining appropriate measures, and issues related to identifying the most accuracy for model representation should be considered (Abbaszadeh Shahri et al., 2021c). In this context, prediction intervals (*PI*) provide an indicator for the *UQ* of individual estimation. The PI has advantages over point estimates because of considering the variability in the data to provide a "reasonable" range of values for an observation (Chatfield, 1993; Meade and Islam, 1995). Therefore, *PI* must account for both the uncertainty in predicting the population mean, plus the random variation of the individual values (Lawless and Fredette, 2005), and thus it is always wider than a confidence interval. This implies that when the statistics cannot be employed to explain the future data, the *UQ* for the upcoming observations can be estimated using *PI* based on known populations of data. One of the main advantages of using *PI* is that it gives a range of likely weights and thus the modeler can have a sense of how accurate the predicted weight is likely to be (Hastie et al., 2001). In Figure 9, the estimated *UQ* intervals in terms of the level of 95% *PI* using *ARDCW-DNLN*, *QR* and *OK* for the scanned *DTB* area are presented. The distribution of classified *UQ* intervals lead to identify the sub-areas with higher uncertainties in which for reduction the need for more data and applicability of its corresponding in-situ test are highlighted. According to calculated *PI*, the *ARDCW-DNLN* with estimated *UQ* within [0.6, 5.2] showed narrower interval than *OK* and *QR* that depicts more concise results and thus higher accuracy. This implies that the bias and discrepancy of the *OK*, *QR*, and *ARDCW-DNLN* can be interpreted through the comparison of the estimated *UQ* in terms of *PI*.

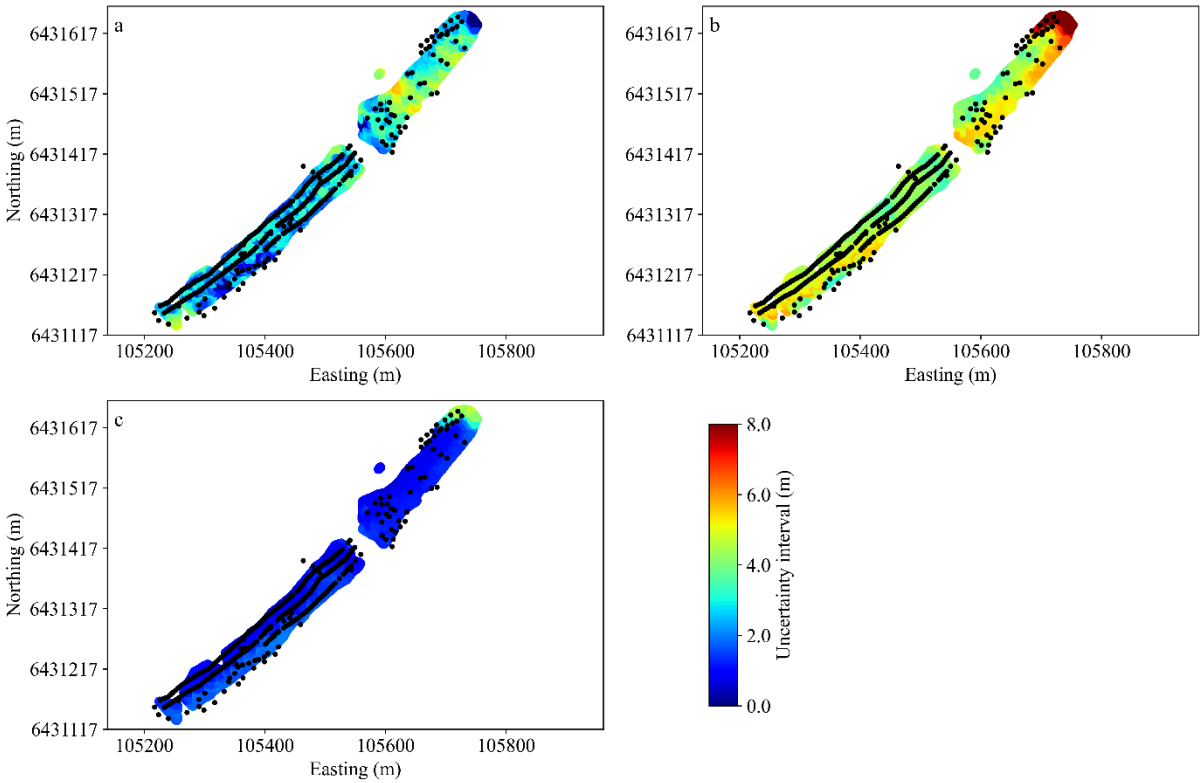


Figure 10. Comparative *UQ* analysis of the applied methods at a level of 95% *PI* for (a) *OK*, (b) *QR* and (c) optimum *DNLN*.

Concluding remarks

Referring to growing field of 3D modelling, an advanced predictive hybrid automated *DNLN* for visualization and *UQ* analysis using multi source geo-data in Sweden was proposed and developed. The presented automated *DNLN* framework can highly be benefited for subsurface geo-engineering applications through incorporating the accurate predictive spatial *DTB* with geological and geomechanical features. There different models using

proposed automated *DNLN* framework, *QR* and *OK* subjected to 65% of randomized 644 multisource datasets were trained. The predictability of captured optimum *DNLN* topology with structure of 3-20-15-1 then was compared with *QR* and *OK* using 75715 scanned *DTB* to provide better interpretation and easy convenient for post-modeling analysis. To evaluate the precision of applied models, the calculated *TAE* was supplemented by *SD* of calculate residuals. It was observed that the residual of *OK* in compare with the calculated *TAE* in predicted *DTB* is exceeded which can be referred to instability of this model in handling the big data. In a mathematical point view the reason for high R^2 in training the *OK* and then significant variation in prediction of big data were discussed and concluded that this traditional interpolation geostatistical method cannot generalize well in predicting of the big data. Referring to *AIC* in penalizing appended terms to a model, it was approved that the high achieved R^2 and thus low *RMSE* in training and validation of *OK* cannot lead for better performance than *QR* and optimum *DNLN*. Comparison of the applied models using different statistical metrics also showed *DNLN* can provide better results than *QR* and *OK*. Provided maps subjected to 95% *PI* for the scanned *DTB* revealed the narrower *UQ* estimation using *ARDCW-DNLN* within [0.6, 5.2] depicting more precise results and thus higher accuracy.

The presented 3D model can improve the visualization of geoengineers for better analyzing of crucial structural elements and thus better insights into the project based on the involved natural and man-made aspects leading to safer design stages. Using such digital but flexible model enable them to add the coordinates of the infrastructure to modify the inconsistencies of the plan without requiring to blueprints. This issue will give them the opportunity for dedicating more applicable solution to fix the difficulties in a right way. Using 3D models also enhance the bidding process and allow contractors to benefit from the automated intelligent modeling to yield higher quality and less expensive construction. The results approved those technological advancements using *DNLN* in geoengineering sectors dedicate easier task accomplishment and thus better outcomes. Although, *DNLN* still needs to overcome a few challenges before becoming a more versatile tool in real-world and thus geoengineering applications, but it is expected for large making investments to further technology in solving more complex problems in the future. The adopted 3D models not only speeds up the design process but also enable geo-engineers and decision makers play around with different ideas and identify potential design problems before they become actual issues. This implies that 3D modeling in construction can put all the phases together and dedicate a real view of the finished project. Swedish agencies also can provide 3D design data to potential bidders, or contractors to develop their own models for use during construction.

References

1. Abbaszadeh Shahri A, Larsson S, Renkel C (2020) Artificial intelligence models to generate visualized bedrock level: a case study in Sweden. Model. Earth Syst. Environ., 6, 1509–1528, doi: 10.1007/s40808-020-00767-0.
2. Abbaszadeh Shahri A, Shan C, Larsson S, Zäll E (2021a) Spatial distribution modelling of subsurface bedrock using a developed automated intelligence deep learning procedure: A case study in Sweden. J. Rock Mech Geotech Eng., doi: 10.1016/j.jrmge.2021.07.006.
3. Abbaszadeh Shahri A, Kheiri A, Hamzeh A (2021b) Subsurface topographic modeling using geospatial and data driven algorithm. ISPRS Int. J. Geo-Inf., 10(5), 341, doi: 10.3390/ijgi10050341.
4. Abbaszadeh Shahri A, Shan C, Larsson S (2021c) A new approach to uncertainty analysis using automated predictive deep learning in groundwater. Natural Research resources, In Process.
5. Abbaszadeh Shahri A, Asheghi R, Khorsand Zak M (2021d) A hybridized intelligence model to improve the predictability level of strength index parameters of rocks. Neural Comput & Applic., 33, 3841–3854, doi: 10.1007/s00521-020-05223-9.
6. Abdar M, Pourpanah F, Hussain S, Rezazadegan D, Liu L, Ghavamzadeh M, Fieguth P, Cao X, Khosravi A, Acharya UR, Makarenkov V, Nahavandi S (2021) A review of uncertainty quantification in deep learning: techniques, applications and challenges. Information Fusion, 76, 243-297, doi: 10.1016/j.inffus.2021.05.008.
7. Akaike H (1970) Statistical predictor identification. Annals of the Institute of Statistical Mathematics, 22, 203-277, doi:10.1007/bf02506337.

8. Anderson MP, Woessner WW, Hunt RJ (2015) Model dimensionality and setting boundaries. In Applied groundwater modeling (2nd Eds), Simulation of flow and advective transport, 117-180, Academic Press, Elsevier, doi: 10.1016/B978-0-08-091638-5.00004-3.
9. Apel M (2006) From 3d geomodelling systems towards 3d geoscience information systems: Data model, query functionality, and data management. Comput Geosci., 32(2):222–229, doi: 10.1016/j.cageo.2005.06.016.
10. Asheghi R, Hosseini SA, Saneie M, Abbaszadeh Shahri A (2020) Updating the neural network sediment load models using different sensitivity analysis methods: a regional application. Journal of Hydroinformatics, 22(3):562-577, doi: 10.2166/hydro.2020.098.
11. Athanasopoulou A, Bezuijen A, Bogusz W, Bournas D, Brandtner M, Breunese A, Burbaum U, Dimova S, Frank R, Ganz H, Grunicke U, Jung H, Lewandowska A, Nuijten G, Pecker A, Psomas S, Roessler K, Sciotti A, Sousa ML, Stille H, Subrin D (2019) Standardisation needs for the design of underground structures. EUR 29633 EN. Publications Office of the European Union, Luxembourg, JRC115352, doi: 10.2760/615209.
12. Bengio Y (2009) Learning deep architectures for AI. Foundations and Trends in Machine Learning, 2(1):1–127, doi: 10.1561/2200000006.
13. Bremnes JB (2004) Probabilistic forecasts of precipitation in terms of quantiles using NWP model output. Monthly Weather Review, 132(1):338–347, doi: 10.1175/1520-0493(2004)132<0338:PFOPIT>2.0.CO;2.
14. Bárdossy G, Fodor J (2001) Traditional and new ways to handle uncertainty in geology. Natural Resources Research, 10, 179–187, doi: 10.1023/A:1012513107364.
15. Burnham K P, Anderson DR (2002) Model selection and multimodel inference. 2nd Eds., Springer-Verlag, New York, doi:10.1007/b97636.
16. Caumon G, Collon-Drouaillet P, Le Carlier de Veslud C, Viseur S, Sausse J (2009) Surface-based 3D modeling of geological structures. Math Geosci., 41, 927–945, doi: 10.1007/s11004-009-9244-2.
17. Chen XW, Lin X (2014) Big data deep learning: challenges and perspectives. IEEE Access, 2, 514-525, doi: 10.1109/ACCESS.2014.2325029.
18. Chew LP (1989) Constrained delaunay triangulations. Algorithmica, 4(1–4):97–108, doi: 10.1007/BF01553881.
19. Chatfield C (1993) Calculating interval forecasts. Journal of Business & Economic Statistics, 11(2):121-135, doi:10.2307/1391361.
20. Christian B, Griffiths T (2017) Overfitting. In: Algorithms to live by the computer science of human decision. William Collins, Glasgow.
21. Cox DC, Baybutt P (1981) Methods for uncertainty analysis: a comparative survey. Risk Analysis, 1(4):251-258, doi: 10.1111/j.1539-6924.1981.tb01425.x.
22. Do Couto D, Gumiaux C, Jolivet L, Augier R, Lebret N, Folcher N, Jouannic G, Suc JP, Gorini C (2015) 3D modelling of the Sorbas Basin (Spain): New constraints on the Messinian Erosional Surface morphology. Marine and Petroleum Geology, 66, 101-116, doi: 10.1016/j.marpetgeo.2014.12.011.
23. Dong M, Neukum C, Hu H, Azzam R (2015) Real 3D geotechnical modeling in engineering geology: a case study from the inner city of Aachen, Germany. Bull Eng Geol Environ., 74, 281–300, doi: 10.1007/s10064-014-0640-6.
24. Elam KR, Rearden BT (2017) Use of sensitivity and uncertainty analysis to select benchmark experiments for the validation of computer codes and data. Nuclear Science and Engineering. 145(2):196-212, doi: 10.13182/NSE03-A2376.
25. Everitt BS, Skrondal A (2010) Cambridge dictionary of statistics. 4th Eds., Cambridge University Press, Cambridge, UK.
26. Frank T, Tertois AL, Mallet JL (2007) 3D-reconstruction of complex geological interfaces from irregularly distributed and noisy point data. Comput Geosci 33(7):932–943, doi: 10.1016/j.cageo.2006.11.014.
27. Ghaderi A, Abbaszadeh Shahri A, Larsson S (2019) An artificial neural network based model to predict spatial soil type distribution using piezocone penetration test data (CPTu). Bull Eng Geol Environ., 78, 4579–4588, doi: 10.1007/s10064-018-1400-9.

28. Gal Y, Ghahramani Z (2016) Dropout as a Bayesian approximation: representing model uncertainty in deep learning. In Balcan MF and Weinberger KQ (eds), Proc 33nd International Conference on Machine Learning, ICML, 48, 1050-1059, NY, USA.
29. Glimm J, Sharp DH (1999) Prediction and the quantification of uncertainty. Physica D, 133(1-4):152–170, doi: 10.1016/S0167-2789(99)00103-7.
30. Gomes GJC, Vrugt JA, Vargas EA (2016) Toward improved prediction of the bedrock depth underneath hillslopes: Bayesian inference of the bottom-up control hypothesis using high-resolution topographic data. Water Resour. Res., 52, 3085–3112, doi: 10.1002/2015WR018147.
31. Günther A (2003) SLOPEMAP: programs for automated mapping of geometrical and kinematical properties of hard rock hill slopes. Comput Geosci 29(7):865–875, doi: 10.1016/S0098-3004(03)00086-4.
32. Hack R, Orlic B, Ozmutlu S, Zhu S, Rengers N (2006) Three and more dimensional modelling in geo-engineering. Bull Eng Geol Environ., 65, 143–153, doi: 10.1007/s10064-005-0021-2.
33. Hastie T, Tibshirani R, Friedman J (2001) The elements of statistical learning, 2nd Eds., Springer-Verlag, New York, NY, doi: 10.1007/978-0-387-84858-7.
34. Hooimeijer FL, Maring L (2018) The significance of the subsurface in urban renewal. Journal of Urbanism: International Research on Placemaking and Urban Sustainability, 11(3): 303-328, doi: 10.1080/17549175.2017.1422532.
35. Houlding S (1994) 3D geoscince modeling, computer techniques for geological characterization. 1st Eds, Springer, Verlag, Berlin Heidenlberg, doi: 10.1007/978-3-642-79012-6.
36. Karpatne A, Ebert-Uphoff I, Ravela S, Babaie HA, Kumar V (2018). Machine learning for the geosciences: Challenges and opportunities. IEEE Transactions on Knowledge and Data Engineering, 31(8):1544–1554, doi: 10.1109/TKDE.2018.2861006.
37. Kriegeskorte N, Golan T (2019) Neural network models and deep learning. Current Biology, 29(7), R231–R236, doi: 10.1016/j.cub.2019.02.034.
38. Kumar P, Rani M, Pandey P, Sajjad H, Chaudhary BS (2019) Applications and challenges of geospatial technology, potential and future trends. Springer Nature, Switzerland AG, doi: 10.1007/978-3-319-99882-4.
39. Lakshmanan V (2012) Automating the analysis of spatial grids, a practical guide to data mining geospatial images for human & environmental applications. Springer Science, Dordrecht, Netherlands, doi: 10.1007/978-94-007-4075-4.
40. Lakshmikantham V, Sen SK (2005) Errors and approximations in digital computers. Mathematics in Science and Engineering, 201, 95-145, doi: 10.1016/S0076-5392(05)80055-4.
41. Lawless JF, Fredette M (2005) Frequentist prediction intervals and predictive distributions. Biometrika, 92(3):529-542, doi:10.1093/biomet/92.3.529.
42. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature, 521, 436-444, doi:10.1038/nature14539.
43. Lee J, Zlatanova S (2007) A 3D data model and topological analyses for emergency response in urban areas. In: Geospatial information technology for emergency response, 1st Eds, CRC Press, London, doi: 10.4324/9780203928813.
44. Mallet JL (1992) Discrete smooth interpolation in geometric modeling. Comp. Aided Design, 24(4):178–191, doi: 10.1016/0010-4485(92)90054-E.
45. Mallet JL (2008) Numerical Earth models. European Association of Geoscientists and Engineers, EAGE Publications bv.
46. Mayoraz R, Mann CE, Parriauz A (1992) Three-dimensional modelling of complex geological structures: new development tools for creating 3-D volumes. In: Hamilton DE and Jones TA (eds.), Computer modelling of geological surfaces and volumes, AAPG, Datapages Archives Inc., Tulsa, USA.
47. Meade N, Islam T (1995) Prediction intervals for growth curve forecasts. Journal of Forecasting, 14(5):413–430, doi:10.1002/for.3980140502.
48. Mielby S, Eriksson I, Campbell D, de Beer J, Bonsor H, Le Guern C, van der Krogt R, Lawrence D, Ryżyński G, Schokker J, Watson C (2016) Opening up the subsurface for the cities of tomorrow. Considering access to subsurface knowledge- Evaluation of practices and techniques, COST Sub-Urban WG2 Report, TU1206-WG2.0-001, Bruxelles, Belgium, www.cost.eu, www.sub-urban.eu.

49. Morgan MG, Henrion M (1990) Uncertainty: a guide to dealing with uncertainty in quantitative risk and policy analysis. Cambridge University Press, Cambridge.
50. Murdie R, Wellman F, Gessner K (2015) Automated estimation of uncertainties in a 3D geological model of the Sandstone Greenstone Belt, Yilgarn Craton, Western Australia. ASEG Extended Abstracts, 1, 1-4, doi: 10.1071/ASEG2015ab193.
51. Niu L, Li L, Zhu R, Huang M (2017) Research and implementation on 3D modeling of geological body. AIP Conference Proceedings 1890, 040088, doi: 10.1063/1.5005290.
52. Lemon AM, Jones NL (2003) Building solid models from boreholes and user-defined cross-sections. Computer & geoscience, 29(5):547-555, doi: 10.1016/S0098-3004(03)00051-7.
53. Perrin M, Zhu B, Rainaud JF, Schneider S (2005) Knowledge-driven applications for geological modeling. Journal of Petroleum Science and Engineering, 47(1-2):89-104, doi: 10.1016/j.petrol.2004.11.010.
54. Randle CH, Bond CE, Lark RM, Monaghan AA (2019) Uncertainty in geological interpretations: effectiveness of expert elicitations. Geosphere, 15(1):108–118, https://doi.org/10.1130/GES01586.1.
55. Schaaf A, De la Varga M, Wellmann F, Bond CE (2020) Constraining stochastic 3-D structural geological models with topology information using approximate bayesian computation using GemPy 2.1. Geoscientific Model Development, doi: 10.5194/gmd-2020-136.
56. Schmidhuber J (2015) Deep learning in neural networks: An overview. Neural Networks, 61, 85–117, doi: 10.1016/j.neunet.2014.09.003.
57. Spina R (2019) Big data and artificial intelligence analytics in geosciences: promises and potential. GSA Today Archive, The Geological Society of America, 29(1):42-43, doi: 10.1130/GSATG372GW.1.
58. Tetko IV, Livingstone DJ, Luik AI (1995) Neural network studies. 1. comparison of overfitting and overtraining. Journal of Chemical Information and Modeling, 35(5):826–833, doi:10.1021/ci00027a006.
59. Thornton J, Mariethoz G, Brunner P (2018) A 3D geological model of a structurally complex Alpine region as a basis for interdisciplinary research. Scientific Data, 5, 180238, doi: 10.1038/sdata.2018.238.
60. Toms BA, Barnes EA, Ebert-Uphoff I (2020) Physically interpretable neural networks for the geosciences: Application to earth system variability. Journal of Advances in Modeling Earth Systems, 12(9): e2019MS002002, doi: 10.1029/2019MS002002.
61. Turner AK (2006) Challenges and trends for geological modelling and visualisation. Bull Eng Geol Environ., 65, 109–127, doi: 10.1007/s10064-005-0015-0.
62. Uusitalo L, Lehikoinen A, Helle I, Myberg K (2015) An overview of methods to evaluate uncertainty of deterministic models in decision support. Environmental Modelling Software, 63, 24-31, doi: 10.1016/j.envsoft.2014.09.017.
63. Vangel MG (1996) Confidence intervals for a normal coefficient of variation. The American Statistician, 50(1):21–26, doi: 10.1080/00031305.1996.10473537.
64. Wackernagel H (1995) Ordinary Kriging. In: Multivariate geostatistics, 74-81, Springer, Berlin, Heidelberg, doi: 10.1007/978-3-662-03098-1_11.
65. Wei S, Hengl T, Mendes de Jesus J, Hua Y, Dai Y (2016) Mapping the global depth to bedrock for land surface modeling. Journal of Advances in Modeling Earth Systems, 9(1):65–88, doi: 10.1002/2016MS000686.
66. Wu Q, Xu H, Zou X, Lei H (2015) A 3D modeling approach to complex faults with multi-source data. Computers & Geosciences, 77, 126-137, doi: 10.1016/j.cageo.2014.10.008.
67. Xu N, Tian H (2009) Wire frame: a reliable approach to build sealed engineering geological models. Comput Geosci., 35(8):1582–1591, doi: 10.1016/j.cageo.2009.01.002.
68. Yan J, Liu Y, Han S, Wang Y, Feng S (2015) Reviews on uncertainty analysis of wind power forecasting. Renewable and Sustainable Energy Reviews, 52, 1322-1330, doi: 10.1016/j.rser.2015.07.197.
69. Yan F, Shangguan W, Zhang J, Hu B (2020) Depth-to-bedrock map of China at a spatial resolution of 100 meters. Sci. Data, 7, 2, doi: 10.1038/s41597-019-0345-6.
70. Zhan X, Lu C, Hu G (2020) Event sequence interpretation of structural geomodels: A knowledge-based approach for extracting tectonic sequences. In: Herzig A, Kontinen J (eds) Foundations of information

and knowledge systems. FoIKS 2020, Lecture notes in computer science, vol 12012, 318-333, Springer, Cham, doi: 10.1007/978-3-030-39951-1_19.
71. Zhang X, Zhang J, Tian Y, Li Z, Zhang Y, Xu L, Wang S (2020) Urban geological 3D modeling based on papery borehole log. ISPRS International Journal of Geo-Information, 9(6):389, doi: 10.3390/ijgi9060389.
72. Zhang TF, Tilke P, Dupont E, Zhu LC, Liang L, Bailey W (2019) Generating geologically realistic 3D reservoir facies models using deep learning of sedimentary architecture with generative adversarial networks. Petroleum Sciences, 16, 541–549, doi: 10.1007/s12182-019-0328-4.
73. Zhang W, Wang J, Lou J, Shan B, Liu J, Wang F, Song S (2021) Simplified 3D modelling method and its uncertainty analysis for adit fractures collected within short sampling windows. Quarterly Journal of Engineering Geology and Hydrogeology, 54 (4):qjegh2020-096, doi: 10.1144/qjegh2020-096.
74. Zhou C, Ouyang J, Ming W, Zhang G, Du Z, Liu Z (2019) A stratigraphic prediction method based on machine learning. Applied Sciences, 9(17), 3553; https://doi.org/10.3390/app9173553.
75. Åsander L (2015) Miljökonsekvensbeskrivning till vägplan Väg E20 delen Bälinge–Vårgårda. Sweden Administration Agency (Trafikverket), Projektnummer 139016, In Swedish (https://www.trafikverket.se/contentassets/60688a2dfff543ffae587e1fc19095f3/aktuella/01_mkb_kap1-5.pdf).

BIG – Branschsamverkan i grunden
Forskningsprogram för effektiv och säker grundläggning av vägar och järnvägar

BIG

BIG – Branschsamverkan i grunden - är ett forskningsprogram för effektiv och säker grundläggning av transportsystemets infrastruktur. Programmet etablerades under senhösten 2013, och påbörjade sin verksamhet den 1 januari, 2014.

Målsättningen är att sänka kostnader för byggande och underhåll av transportsystemets infrastruktur genom ett långsiktigt och systematiskt utvecklingsarbete inom geoteknikområdet.

I BIG samverkar Trafikverket, Chalmers tekniska högskola, Luleå tekniska universitet, Kungliga tekniska högskolan och Statens geotekniska institut.

Resultat från BIG projekt publiceras i vetenskapliga artiklar och parternas egna rapportserier. Alternativt så publiceras rapporterna i denna gemensamma BIG rapport serie.

Publicerade rapporter

A2014:03	Deformationer i undergrund – Litteratursammanställning och analys
A2014:07	Grundvattensänkning i morän
A2014:13	Höghastighetsspår i Sverige – på bank
A2014:14	On the fundamental relation between soil creep and cyclic pile-soil response
A2016:05	Modelling deformations below high-speed rail
A2017:07	Dynamic response of transition zone on soft clay
A2017:10	Effects of climate change on slopes in sensitive clay
A2017:14	Changes in undrained shear strength as a function of time under embankments
A2018:03	Three-dimensional effects in slope stability calculations
A2018:09	Permanent spont utmed järnväg i Uppsala - fallstudie av bygg- & driftskede
A2018:18	Osäkerhetsmodeller – för optimal geoteknisk resusanvändning.
A2018:22	Towards high quality deep soft soil samples
A2019:02	Degradation of a natural sensitive clay under cyclic loading
A2019:07	Ny provbank på lös lera
A2019:17	Rock and Groundwater Surface Modelling Using Artificial Intelligence
A2020:08	Towards contemporary erosion solutions