國立中央大學
National Central University

Modelling Subsurface Stratigraphy within the Taipei Basin using Markov Random Field Theory

Presenter: Stefan Nicholas

Advisor: Prof. Jia-Jyun Dong

Date: 2023/04/21

Introduction

Aims:

- Produce a random field model of the subsurface stratigraphy in the Taipei Basin.
- Assess the spatial correlation of this stratigraphy.
- Quantify the uncertainty within the geological model.

Key terminology:

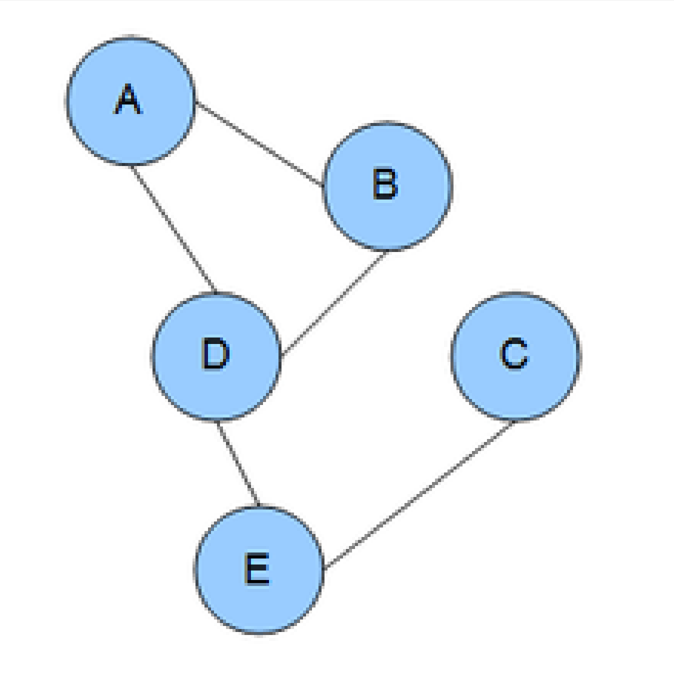
- **Spatial correlation** – refers to the tendency of neighbouring samples or locations in a spatial domain to be more similar to each other than those further away.
- **Uncertainty** – refers to the uncertainty of boundaries between different soil layers and lithological units.
- **Stochastic geological model** – a model that uses random or probabilistic methods to describe the uncertainty and variability of geological features and properties in the subsurface.

Why is it important?

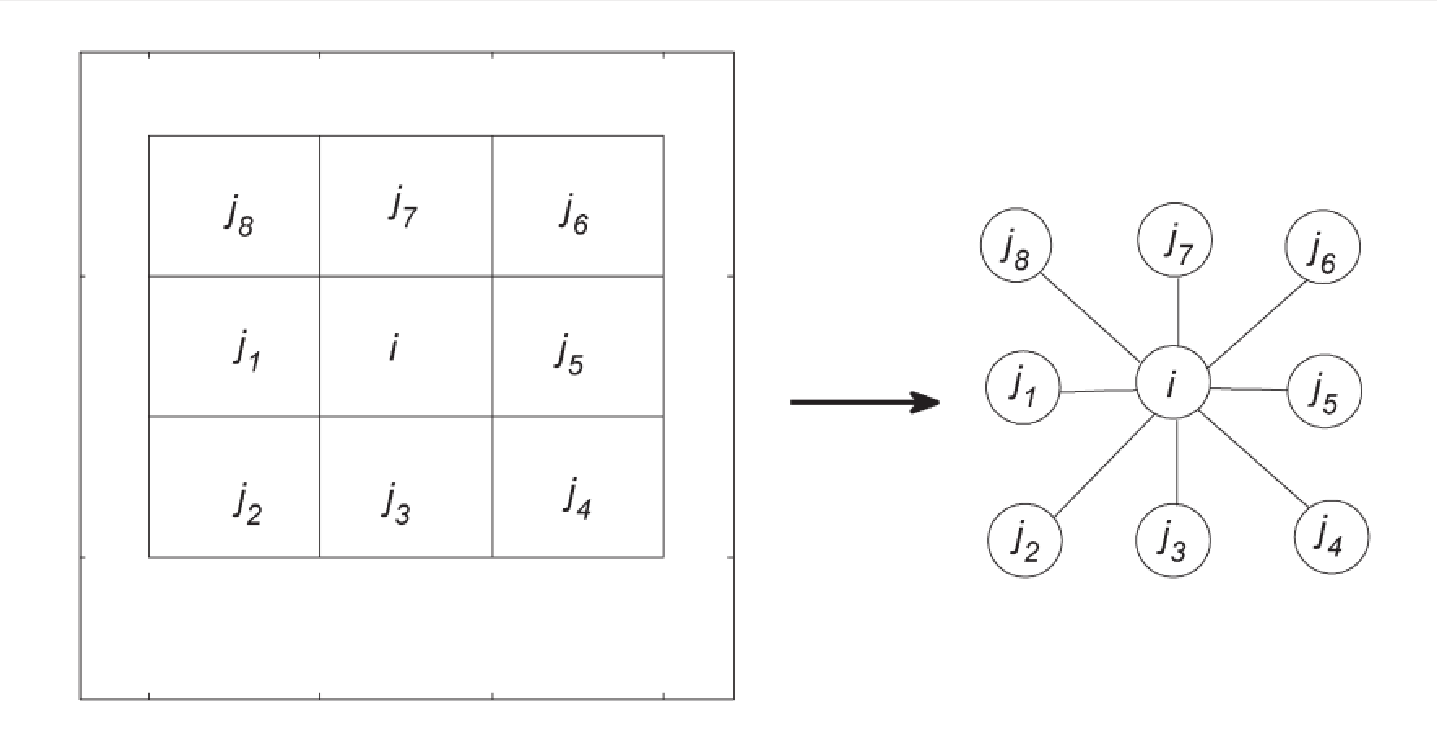
- Although there has been previous research about random field modelling, it hasn't been widely applied in real-case scenarios.
- Subsurface geological models are usually inaccurate due to highly variable strata and sparse site investigation data.
- Random field models could help reduce uncertainty in geological models, and therefore ground investigations as well.
- It is estimated that around 80% of problems discovered on construction projects are attributable to unexpected ground conditions.

Markov random fields

- Markov Random Fields (MRFs) are graphical models used to represent probability distributions over a set of random variables.
- Variables are represented by nodes in the graph, and the edges between nodes represent conditional dependencies between them.
- The main idea behind MRFs is that the probability of a particular configuration of variables depends only on the values of neighbouring variables, and not on the values of distant or unrelated variables. (Markov property).
- This can be used to model the spatial distribution of different types of soil/rocks in a given area.
- The different types of soil or rocks would be represented as the variables in the model, and their spatial relationships would be represented by the edges connecting the nodes.



Neighbourhood system



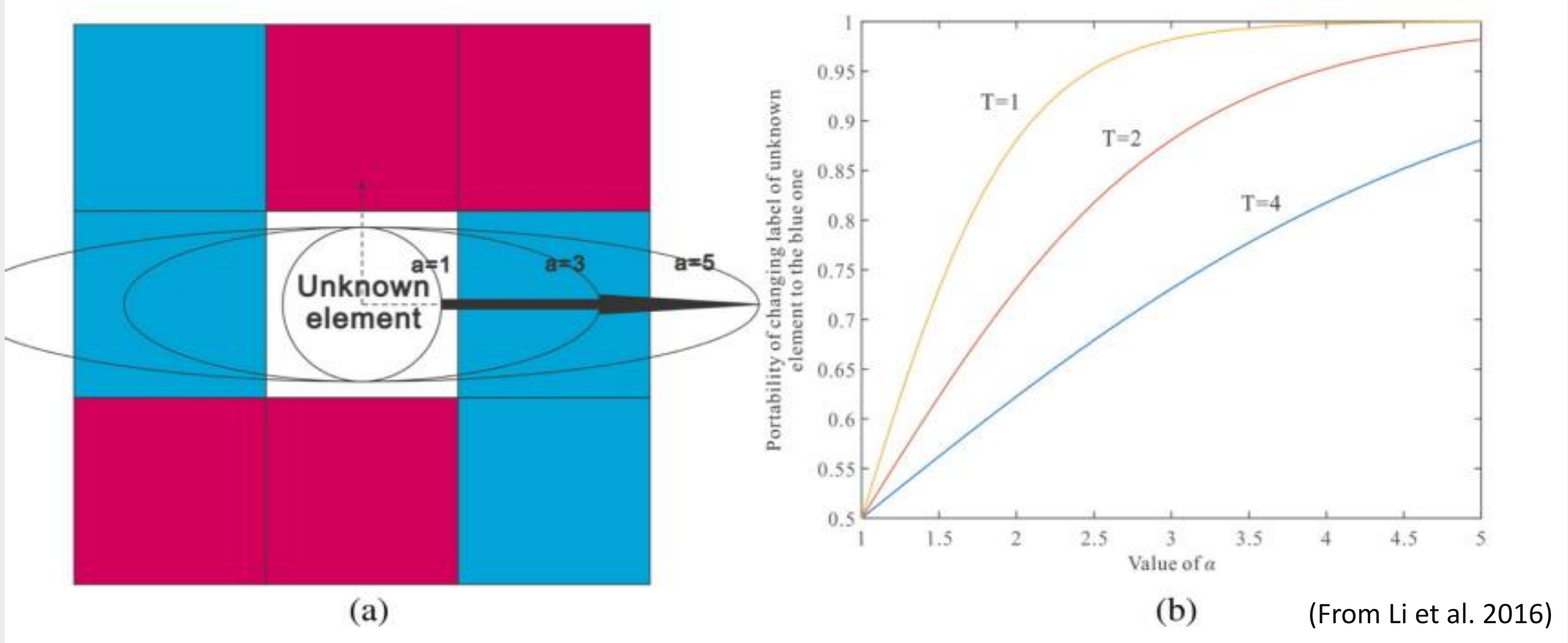
- Geological model is constructed by discretizing the geological body of interest into small square elements.
- The neighbourhood system defines the set of variables that are considered when computing the probability of a given variable, given the values of its neighbours.
- Neighbours are spatially related.
- Spatial correlation divided into 2 parameters

Spatial Correlation

The spatial correlation in the local neighbourhood system is divided into two components:

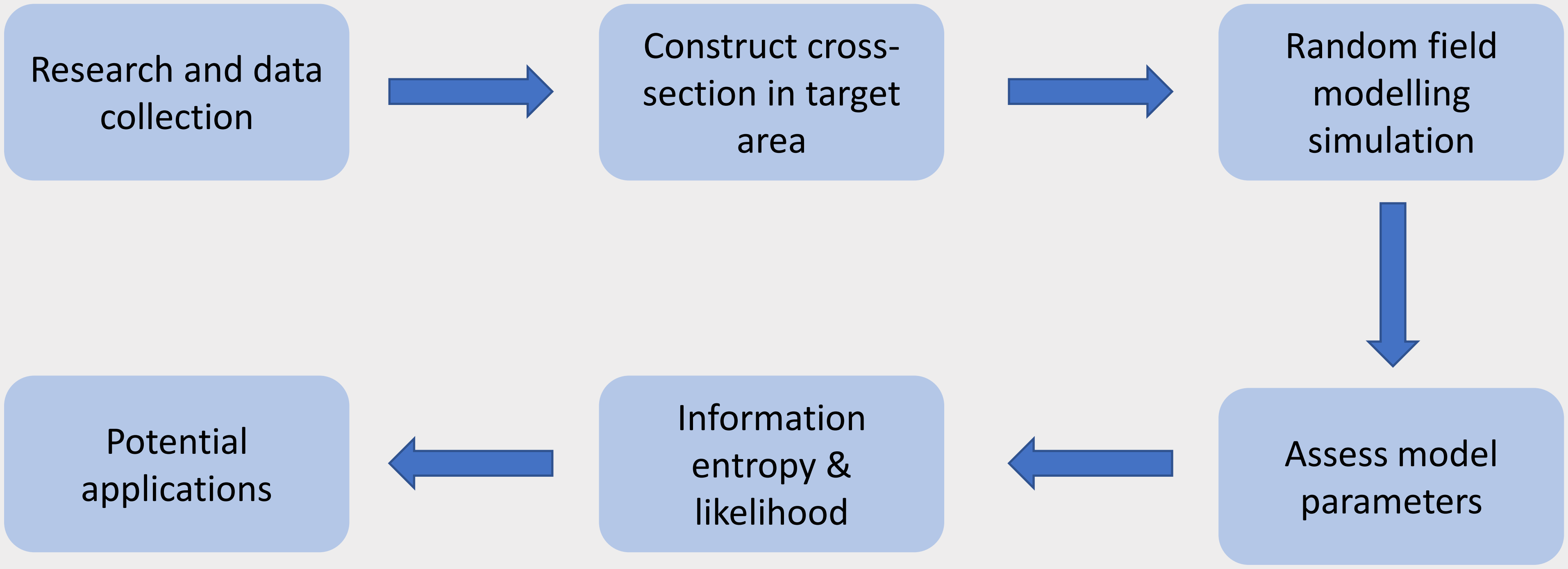
- ψ – orientation information of geological formations
- a – the ratio of strength of tangential correlation and normal correlation

a larger value of 'a' indicates a greater degree of correlation in the model, while a smaller value of 'a' allows for more spatial variability or heterogeneity.



(From Li et al. 2016)

Workflow


Research and data collection
Construct cross-section in target area
Random field modelling simulation
Assess model parameters
Information entropy & likelihood
Potential applications

Data Collection

Data used for cross section:

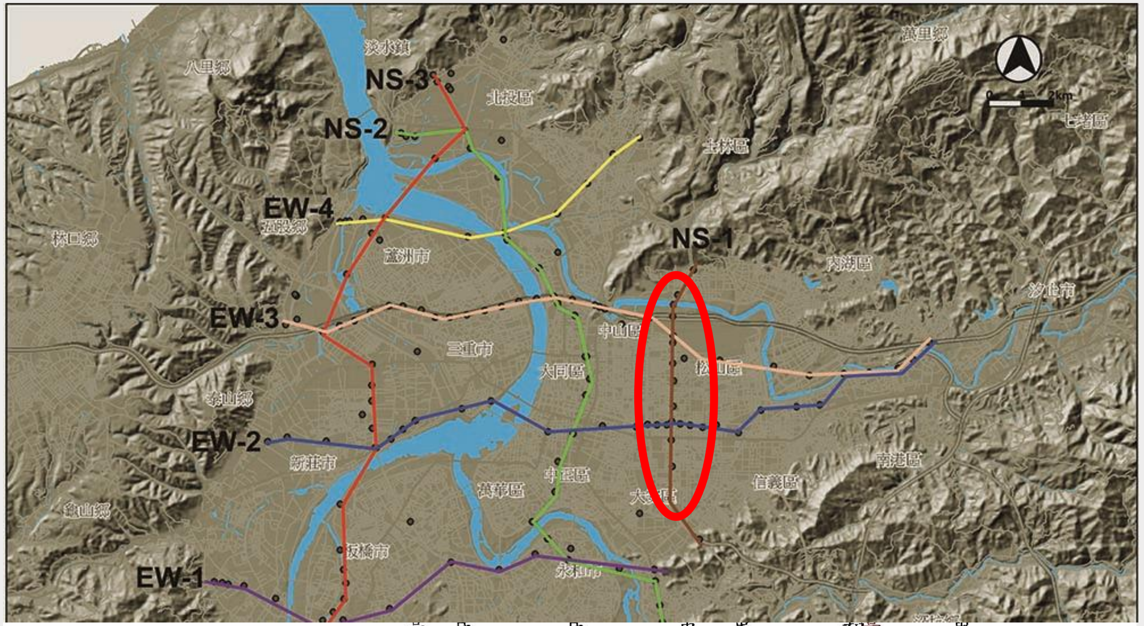
- Borehole data
- Lab test data

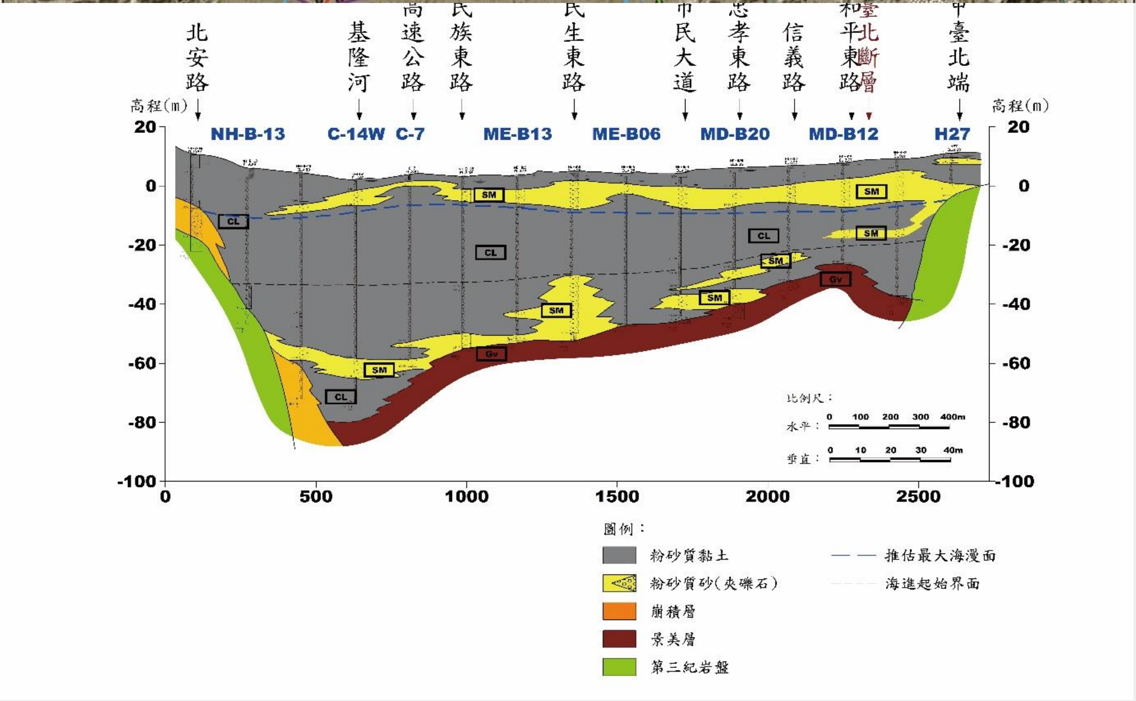
Borehole data and lab test results in the Taipei Basin from CGS.

A total of 22 boreholes were used.

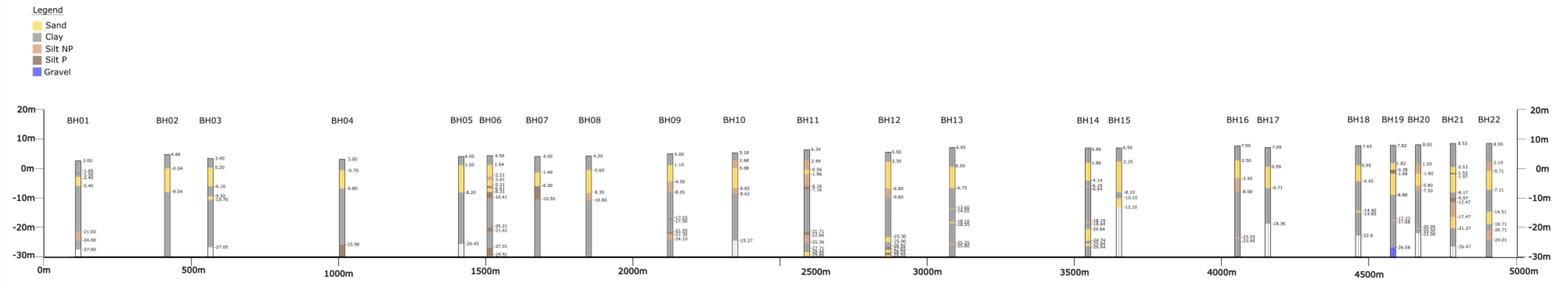
Data input for modelling:

1. Surface data – elevation and distance
2. Soil types
3. Layer boundaries
4. Location of boreholes



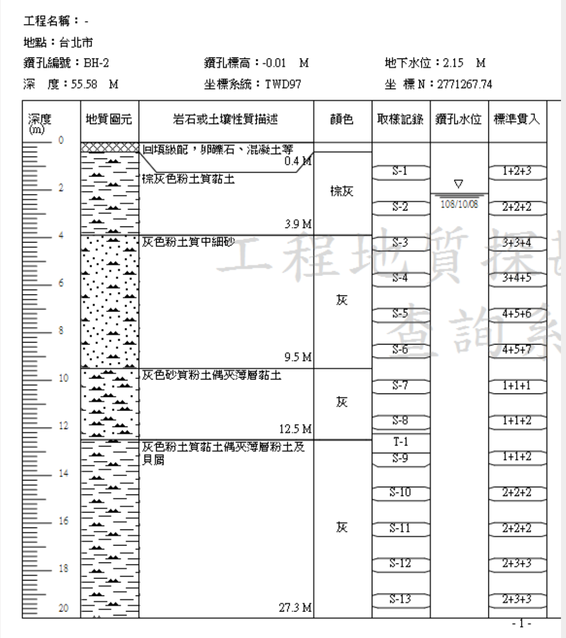


Cross-section borehole data

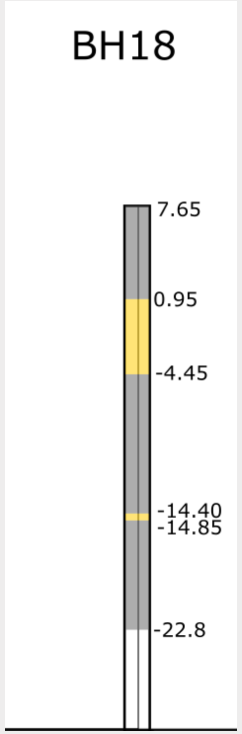


Difficulties:

- Not all borehole data matches lab test data
- Boreholes aren't uniformly distributed
- Made ground
- Lenses and non-continuous layers



Criteria for combining layers



Soil layer is the same above and below:

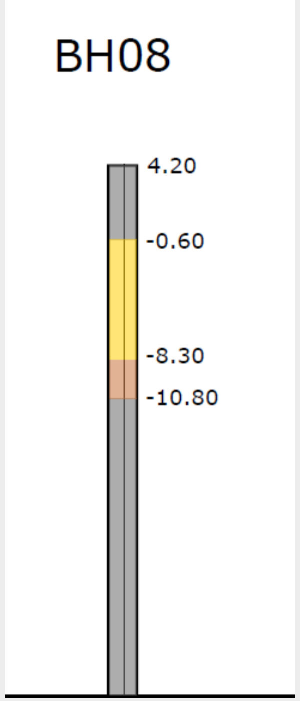
- <2m thick and the material is the same above and below, then combine with surrounding soil type.

Soil layer is different above and below:

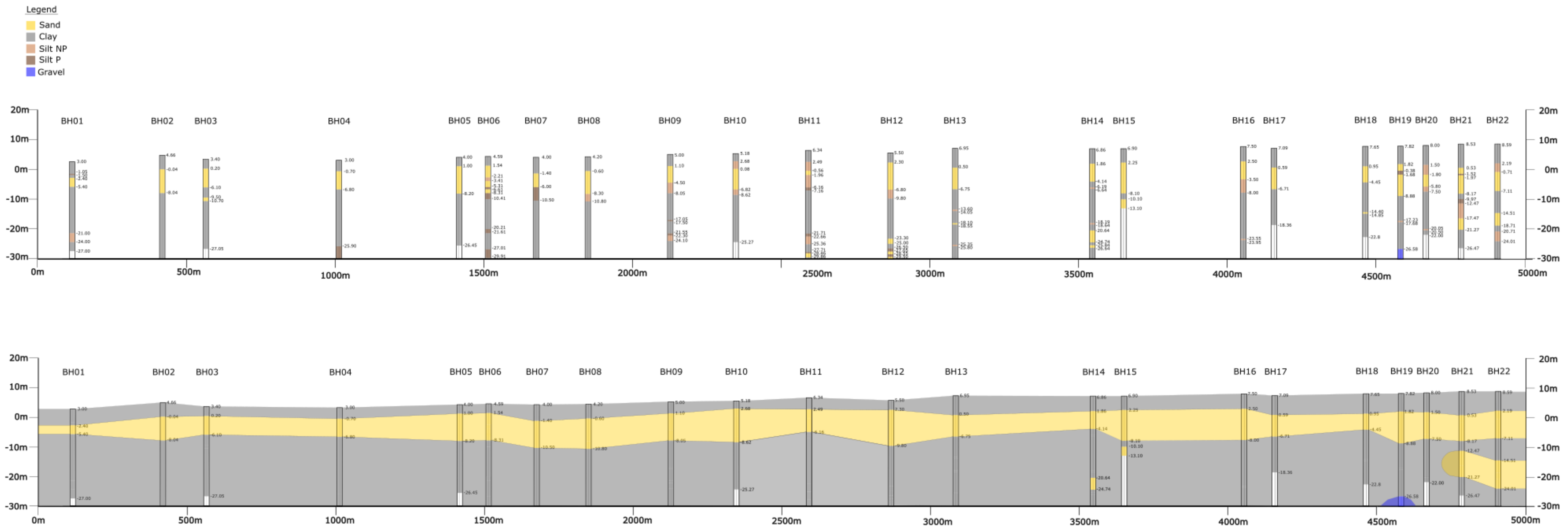
- Silt with plasticity – clay
- Silt without plasticity – sand

*If the silt without plasticity has a higher percentage of clay than sand then change to clay. (BH1)

*If the silt with plasticity has a very low plasticity index (<5%) change to sand. (BH7)



Cross-section for modelling



Assumptions for modelling:

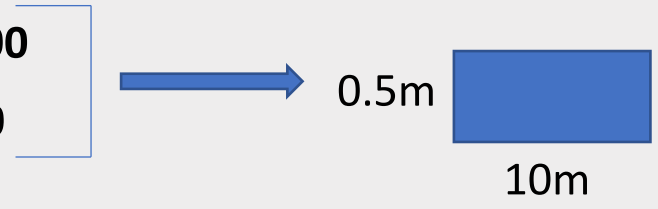
- Made ground area is clay
- Clay extends to 30m in shallower boreholes

Simulation setup

Realisations: **1000**

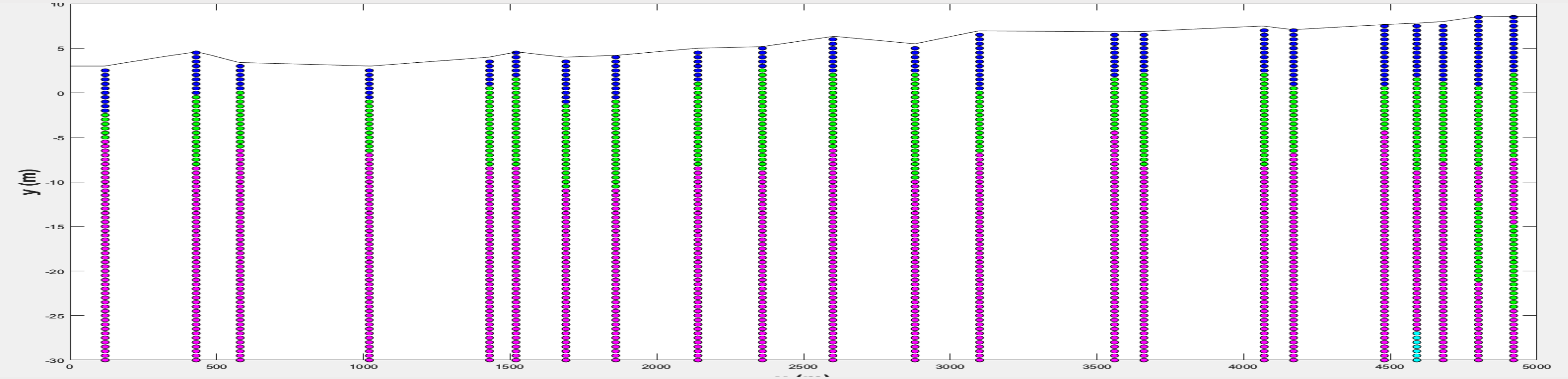
Number of elements in x direction: **500**

Number of elements in y direction: **80**



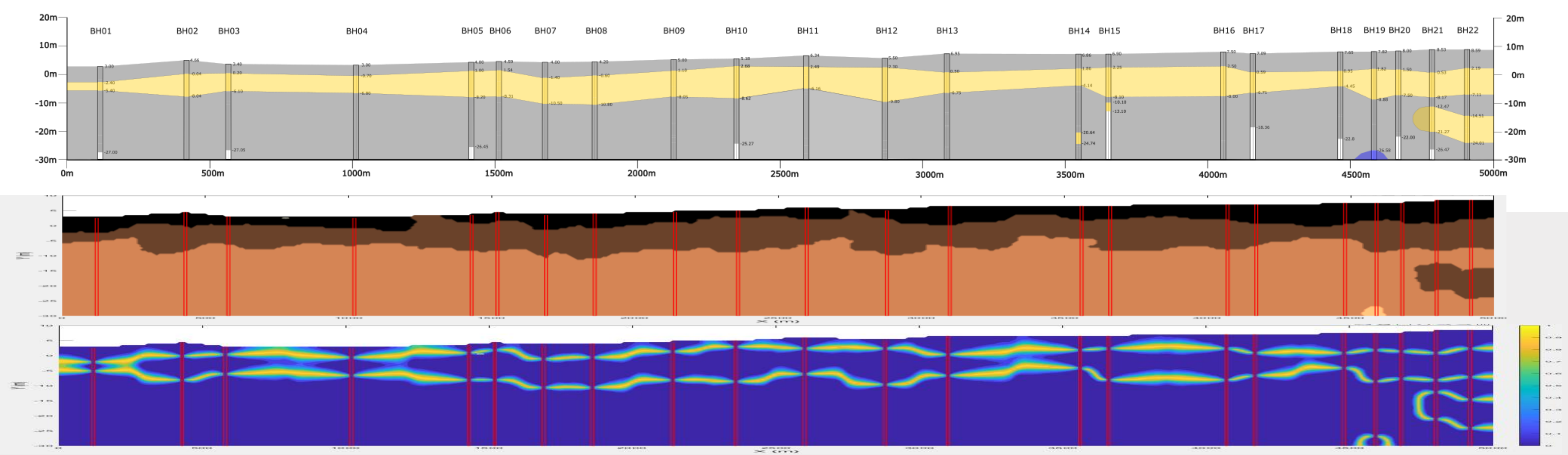
Model length in x direction: **5000**

Model length in y direction: **40**



Clay above and below sand layer is separated

Simulation results using all boreholes

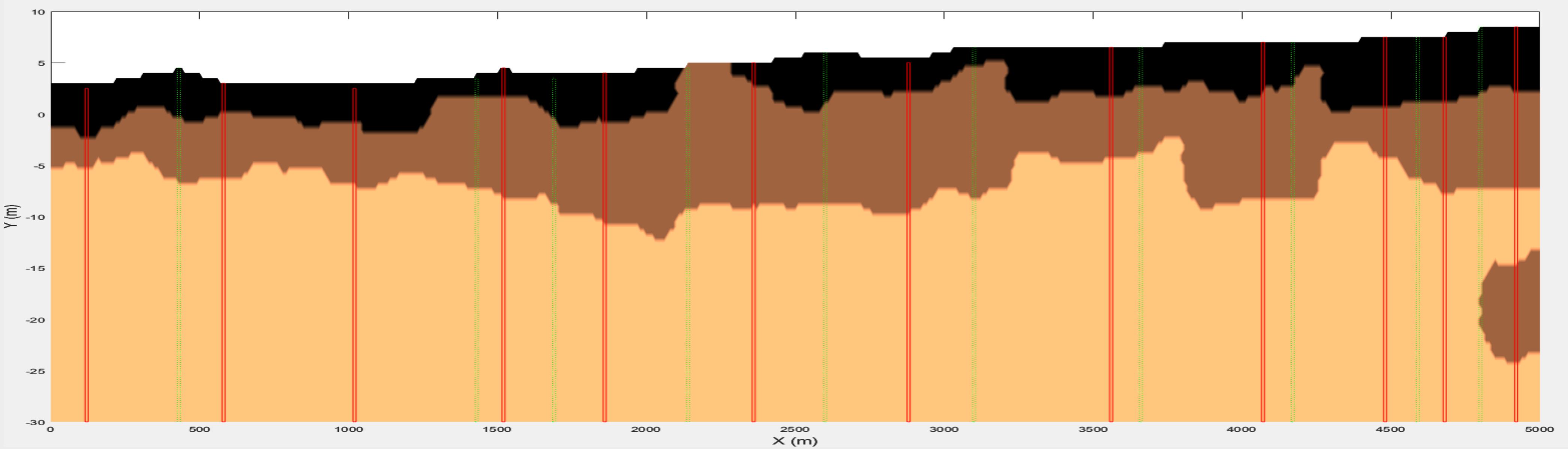


A=6 → A’=120

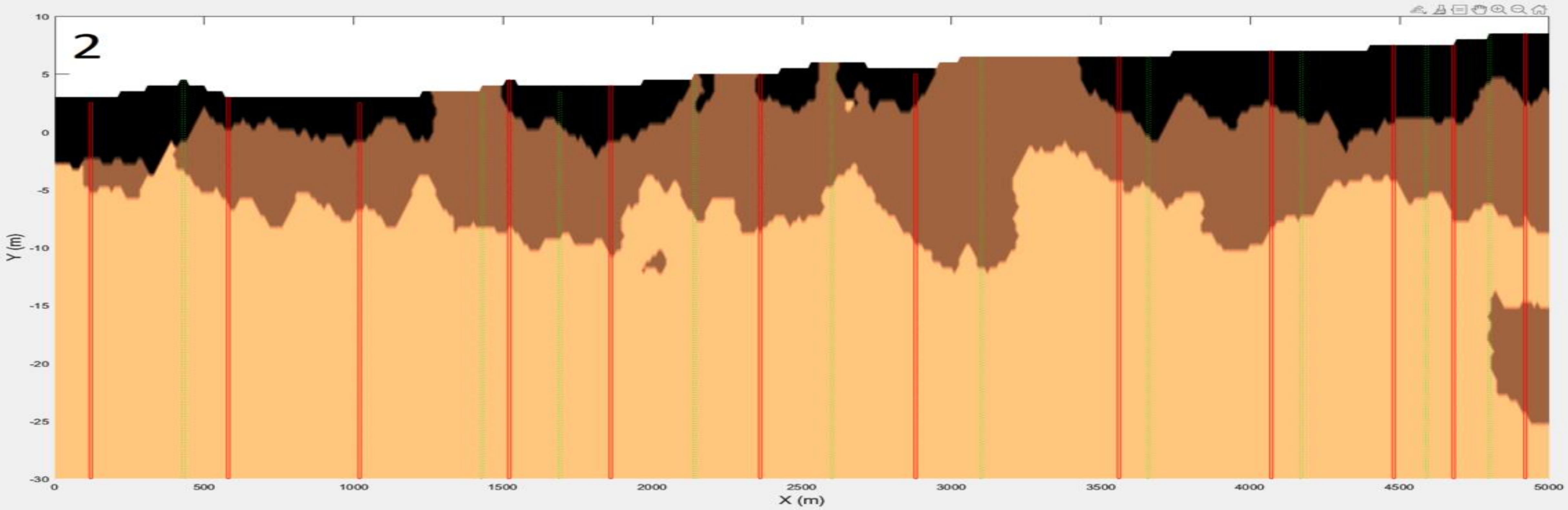
Observational and conditional boreholes

Red lines – conditional boreholes, used to make the model
Green lines – observational boreholes, used for comparison (not simulated)

12 conditional boreholes, 10 observational boreholes



Spatial parameter calibration

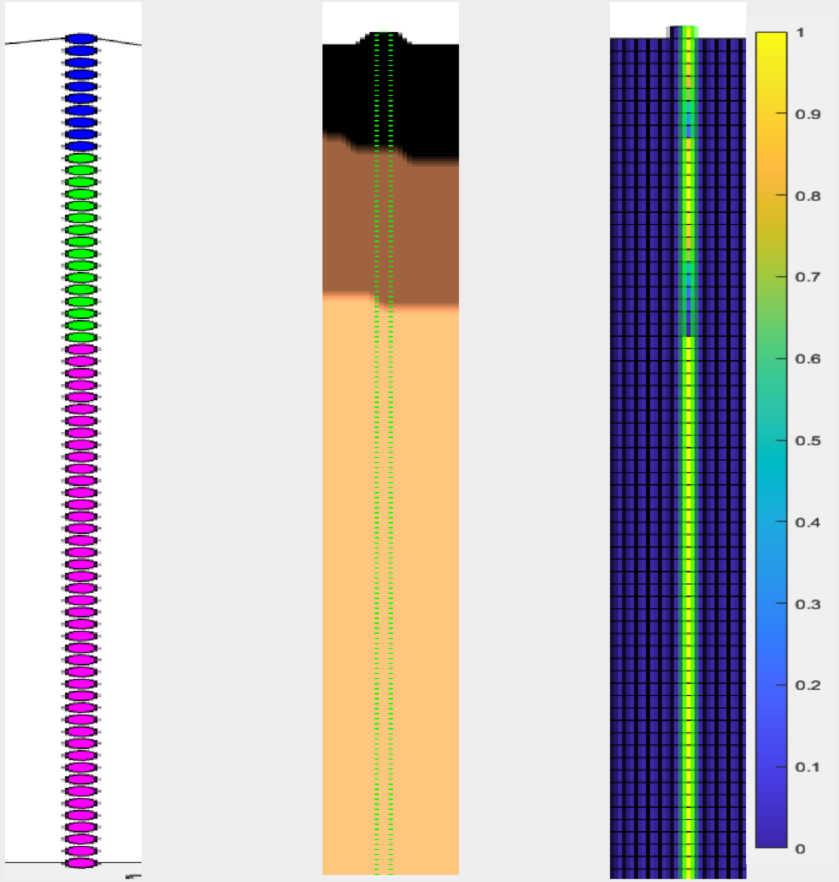


Maximum likelihood estimation

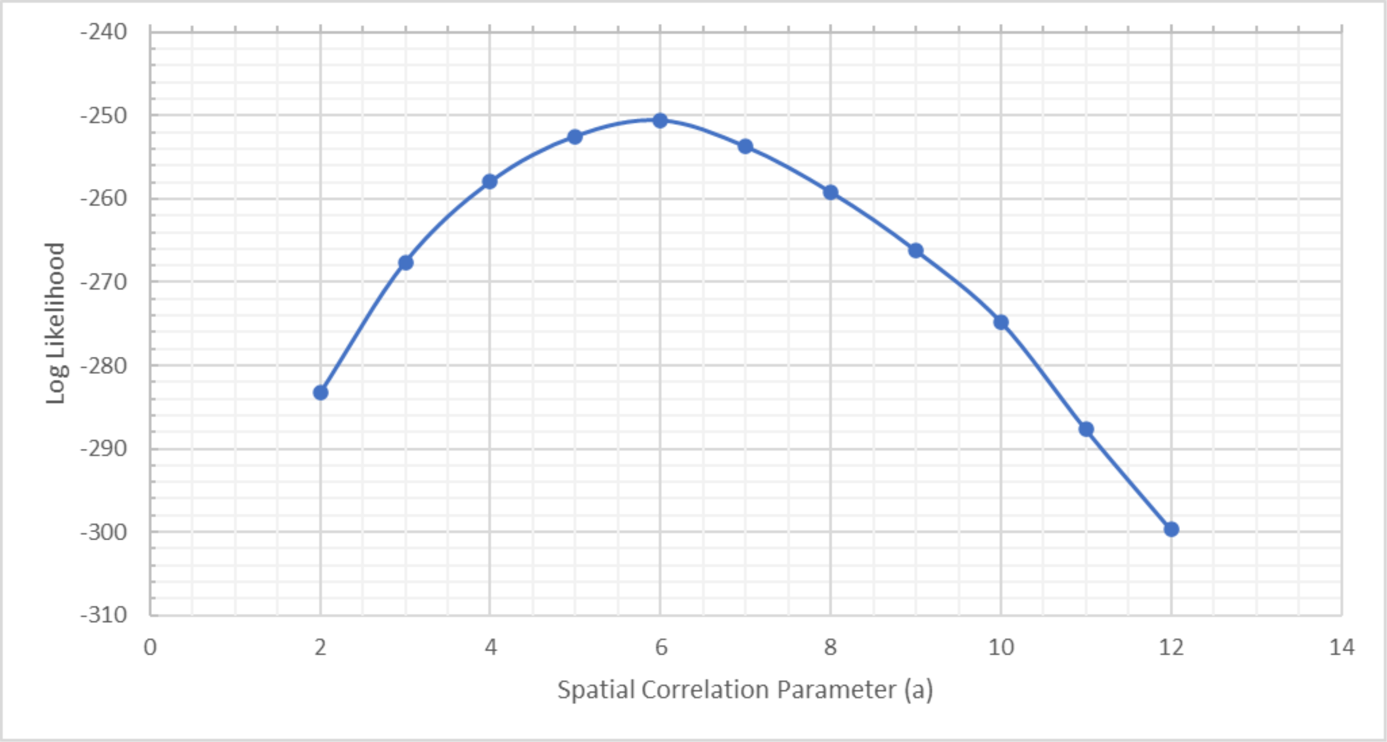
- Determines best model parameters that fit given data
- Lower likelihood = more uncertainty

Maximises log-likelihood function to estimate parameters

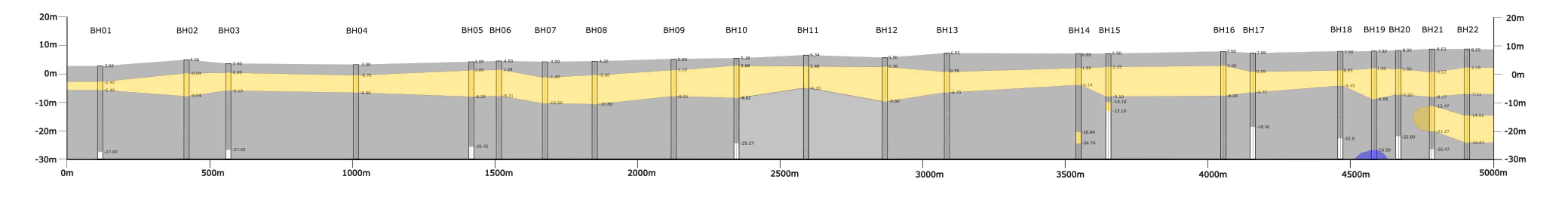
Borehole data – Simulated data – Likelihood

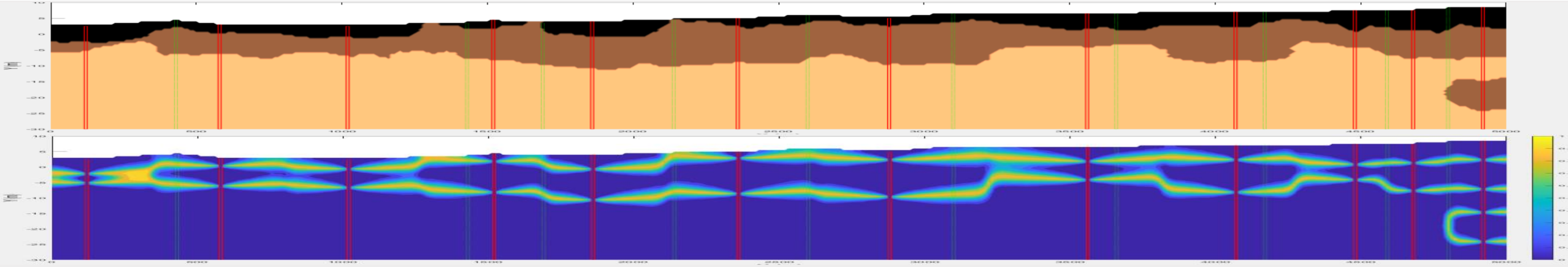


Graph of likelihood vs a value



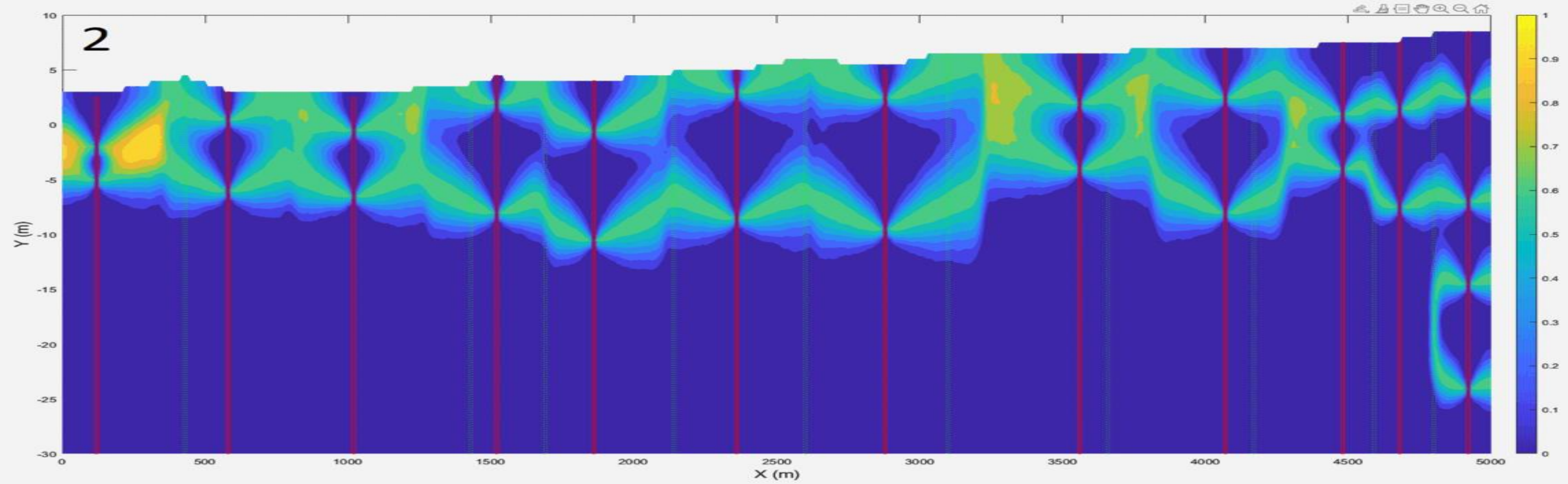
Comparison between cross section and modelling result





Information Entropy

- To quantify the uncertainty of this stochastic geological model, the concept of information entropy is adopted.
- The entropy is 0 when no uncertainty exists (E.g. only 1 possible lithological unit).
- Entropy is highest when all lithological units are equally probable.



More things to consider?

- Number of realisations
- Scale of model
- Does the random field represent possible realisations
- Grid/element size
- Neighbourhood size
- Randomness of model
- Number and spacing of boreholes
- Different spatial correlation for different lithologies
- Computational cost

Conclusion

- Preliminary results show that the best spatial correlation parameter (a) is about 6 (a'=120) for the north-south section chosen within the Taipei basin.
- Maximum Likelihood can be useful for finding the most optimal a value in stochastic geological models.
- However, there are many things that need to be considered for this type of modelling.
- Overall Markov random field modelling can be a useful tool to quantify uncertainty in geological models.

Future Work

- Difference in spatial parameters for different lithologies
- Anisotropy of spatial correlation
- Correlating spatial parameter with deposition environment or different zones with the Taipei Basin
- 3D Modelling

Thanks for listening!

Simulation Steps

Pre-process stage:

- First the geological body is first discretized by a suitable mesh scheme.
- Followed by processing the geometric information of the meshed plot and constructing neighbourhood systems.
- Then, all known data are assigned to the corresponding elements.
- Finally, values for parameters ψ and a are selected.

Stochastic simulation stage:

- The first step is aimed at creating an initial configuration by filling all the blank elements with lithological units.
- A practical way is sampling the geo-material type of elements which are the neighbours of those elements with known geo-material type, and then spread to the whole domain.
- Following this stochastic simulation process, a series of realisations are generated, which can be considered as samples of simulated subsurface lithological profile.

Post-process stage:

- In the post-process stage, based on the simulated realisations, the uncertainty of the predicted lithological profile is assessed by the information entropy.