Report
R-20-19
December 2022

Methodology for modelling of thermal properties of the Forsmark site

Part 2 – Background and methodology development

Pär-Erik Back
Jan Sundberg
John Wrafter
Lars Rosén

SVENSK KÄRNBRÄNSLEHANTERING AB

SWEDISH NUCLEAR FUEL
AND WASTE MANAGEMENT CO

Box 3091, SE-169 03 Solna
Phone +46 8 459 84 00
skb.se

SVENSK KÄRNBRÄNSLEHANTERING

ISSN 1402-3091
SKB R-20-19
ID 1984113

December 2022

Methodology for modelling of thermal properties of the Forsmark site

Part 2 – Background and methodology development

Pär-Erik Back, Statens geotekniska institut

Jan Sundberg, JK Innova AB

John Wrafter, Lars Rosén
SWECO

Preface

A series of methodology reports support the programmes for investigation and modelling during the execution of planned underground constructions at Forsmark. The series includes the following disciplines: geometric modelling of ground elevation and regolith, deterministically modelled geological structures, discrete fracture network (DFN) modelling (stochastic, semi-stochastic and deterministic modelling of structural-hydraulic fracture data), rock mechanics modelling, thermal properties modelling, integrated hydrological and hydrogeological modelling, hydrogeochemical modelling, and transport modelling. Report numbers (ID), acronyms, and titles are shown below. The acronyms are recommended for internal referencing.

ID	Acronym	Title
R-20-10	DGMM	Methodology for deterministic geologic modelling of the Forsmark site
R-20-11	DFNMM1	Methodology for discrete fracture network modelling of the Forsmark site. Part 1 – Concepts, Data and Interpretation Methods.
R-20-12	DFNMM2	Methodology for discrete fracture network modelling of the Forsmark site. Part 2 – Application examples.
R-20-13	RMMM	Methodology for rock mechanics modelling of the Forsmark site.
R-20-14	HGMM	Methodology for hydrological and hydrogeological modelling of the Forsmark site.
R-20-15	HCMM	Methodology for hydrochemical modelling of the Forsmark site.
R-20-16	ERMM	Methodology for elevation and regolith modelling of the Forsmark site.
R-20-17	TRPMM	Methodology for site descriptive and safety assessment transport modelling of the Forsmark site.
R-20-18	THPMM1	Methodology for modelling of thermal properties of the Forsmark site. Part 1 – Recommended data and interpretation methods.
R-20-19	THPMM2	Methodology for modelling of thermal properties of the Forsmark site. Part 2 – Background and methodology development.

The methodology for thermal modelling at the Forsmark site is presented in two reports, Part 1 and Part 2. The main methodology report, Part 1 (SKB R-20-18; Back and Sundberg 2022) is an updated version of previous thermal strategy reports (Sundberg 2003, Back and Sundberg 2007). The previous versions were focused on site descriptive modelling at rock domain level. This third version has an updated objective: To describe a methodology to be used in conjunction to the development of deposition areas of the spent fuel repository at the Forsmark site (the construction and operational phases).

This report, Part 2, presents background information on the methodology for modelling of thermal properties at the Forsmark site. It is a compilation of the developments that have been made over time regarding the most important steps in the methodology. Important background information and definitions are documented, as well as evaluations and discussions leading to choices made in the development of the methodology. The report should preferably be read in conjunction with Part 1.

A reference group has been connected to the project, with the following participants: Diego Mas Ivars, Eva Hakami, Lillemor Claesson Liljedahl, Peter Hultgren, and Anders Winberg. In addition, valuable information has been supplied by Jesper Petersson and Thomas Andolfsson.

Significant contributions to the development of the previous version of the methodology have been made by Johan Andersson, Rolf Christiansson, Lars O Ericsson, Harald Hökmark, Michael Stephens, Assen Simeonov, Raymond Munier, and Peter Dowd (University of Adelaide).

Summary

The methodology for thermal modelling at the Forsmark site is presented in two reports, Part 1 and Part 2. The methodology report (Part 1, Back and Sundberg 2022) is an updated version of previous thermal strategy reports. The present report, Part 2, presents background information on the methodology for modelling of thermal properties at the Forsmark site. It is a compilation of the developments that have been made compared to the previous version (Back and Sundberg 2007) regarding the most important steps in the methodology.

The choice of simulation scale is one such important step. Different aspects on scale is discussed, such as the scale that data represents, the cell size in computer simulations, and the simulated rock volume, both regarding lithological and thermal simulations. Taking all aspects into account, the conclusion is that the simulation scale for the modelling should be 1 m or larger.

Thermal rock classes (TRCs) are defined in order to facilitate the modelling and simulation of lithology. A TRC is a categorical entity without defined spatial boundaries between each class. The reason that TRCs need to be defined instead of using the rock types directly is that there is a large number of rock types but only a limited number of classes can be handled in the simulation of TRC lithology due to software limitations. Two alternatives of Thermal Rock Classes have been developed. The first alternative, or definition, was used during the site investigations and site descriptive modelling (Back et al. 2007). A new TRC definition is proposed for the construction and operation of the repository. The new definition assumes that it is possible to use *rock type* in Boremap as a basis for defining TRCs, and at the same time include *rock occurrences* that exist within the simulation scale in each borehole. Thus, the TRCs are coupled to a defined scale, preferably the simulation scale.

It is concluded that the new TRC definition facilitates the lithological modelling. Regarding the thermal models, there are several alternative modelling approaches. Continuous measurement profiles in boreholes is the preferable method. However, the laboratory data approach was tested in order to study the effect of the new TRC definition on the thermal conductivity distributions. It was concluded that the approach works and that the statistical distributions can be determined for each TRC. It is believed that the distribution models can be used in the stochastic simulations without problems.

The statistical modelling of TRC lithology and the stochastic simulation of TRC lithology are performed with computer software. Different options for geostatistical modelling are evaluated, such as object-based versus pixel-based approaches, and two-point versus multipoint geostatistics. It was concluded that the software T-PROGS (two-point approach) has proven to represent the spatial distribution of lithological units reasonably well and is therefore recommended. A test was performed with T-PROGS to study its performance. A conclusion was that running times for the simulations are reasonable and that sufficiently large 3D models can be handled.

Similarly, software is required for statistical modelling of thermal properties and stochastic simulation of thermal properties. A number of alternative software packages were evaluated regarding the ability handle 3D data, performing 3D simulation, variogram modelling capabilities etc. Software packages that do meet the requirements are S-GeMS and ISATIS.

Anisotropic thermal properties at Forsmark are discussed, and the possibility to introduce anisotropy in the thermal modelling. However, it was concluded that there is no reason to introduce anisotropy in the thermal conductivity modelling since available data is very sparse. It is therefore recommended that the anisotropy is handled by the recipients of the thermal model. However, also for this subject more data should be obtained.

The report ends with a discussion of application scenarios for the thermal modelling methodology. These include pilot test of developed modelling strategy, modelling of a single main tunnel, and modelling of a batch/set of deposition tunnels.

Sammanfattning

Metodiken för termisk modellering i Forsmark presenteras i två rapporter, del 1 och del 2. Metodrapporten (del 1, Back and Sundberg 2022) är en uppdaterad version av tidigare termiska strategirapporter. Föreliggande rapport, del 2, presenterar bakgrundsinformation om metodiken för modellering av termiska egenskaper i Forsmark. Det är en sammanställning av den utveckling som har gjorts jämfört med tidigare version (Back and Sundberg 2007) när det gäller de viktigaste stegen i metodiken.

Valet av simuleringsskala är ett sådant viktigt steg. Olika skalfrågor diskuteras, såsom dataskalan, cellstorleken i datasimuleringar och den simulerade bergvolymen, både vad gäller litologiska och termiska simuleringar. Med hänsyn till alla aspekter är slutsatsen att simuleringsskalan för modelleringen ska vara 1 m eller större.

Termiska bergartsklasser (TRC) definieras för att underlätta modellering och simulering av litologi. En TRC är en kategorisk enhet utan definierade rumsliga gränser mellan varje klass. Anledningen till att TRCer måste definieras, istället för att använda bergarterna direkt, är att det finns ett stort antal bergarter men endast ett begränsat antal klasser kan hanteras i simuleringen av TRC-litologi på grund av programvarubegränsningar. Två alternativ för TRC har utvecklats. Det första alternativet, eller definitionen, användes under platsundersökningar och platsbeskrivande modellering (Back et al. 2007). En ny TRC-definition föreslås för byggande och drift av förvaret. Den nya definitionen förutsätter att det är möjligt att använda *rock type* i Boremap som grund för att definiera TRC, och samtidigt inkludera *rock occurrences* som finns inom simuleringsskalan i varje borrhål. Därigenom kopplas TRCerna till en definierad skala, företrädesvis simuleringsskala.

En slutsats är att den nya TRC-definitionen underlättar den litologiska modelleringen. När det gäller de termiska modellerna finns det flera alternativa angreppssätt. Kontinuerliga mätprofiler i borrhål är den metod som rekommenderas. Emellertid testades ett angreppssätt baserat på laboratoriedata för att studera effekten av den nya TRC-definitionen på värmeledningsfördelningarna. Slutsatsen är att tillvägagångssättet fungerar och att statistiska fördelningar av värmeledning kan tas fram för respektive TRC. Troligen kan fördelningsmodellerna användas i de stokastiska simuleringarna utan problem.

Den statistiska modelleringen av TRC-litologi och den stokastiska simuleringen av TRC-litologi utförs med programvara. Olika alternativ för geostatistisk modellering har utvärderats, såsom objektbaserade kontra pixelbaserade tillvägagångssätt samt tvåpunkts- vs multipunkts-geostatistik. Slutsatsen är att programvaran T-PROGS (tvåpunktsmetod) har visat sig kunna beskriva den rumsliga fördelningen av litologiska enheter relativt väl och därför rekommenderas denna programvara. Ett test utfördes med T-PROGS för att studera dess prestanda. En slutsats var att simuleringstiderna är rimliga och att tillräckligt stora 3D-modeller kan hanteras.

På liknande sätt krävs mjukvara för statistisk modellering av termiska egenskaper samt stokastiska simuleringen av termiska egenskaper. Ett antal alternativa mjukvaror utvärderades avseende förmågan att hantera 3D-data, utföra 3D-simulering, variogrammodellering etc. Programvarupaket som uppfyller kraven är S-GeMS och ISATIS.

Anisotropa termiska egenskaper vid Forsmark diskuteras, samt möjligheten att införa anisotropi i den termiska modelleringen. Slutsatsen är dock att det inte finns någon anledning att införa anisotropi i modelleringen på grund av knapphändiga data. Rekommendationen är därför att anisotropin hanteras av mottagarna av den termiska modellen. Även för detta rekommenderas mer mätdata.

Rapporten avslutas med en diskussion om tillämpningsscenarier för den termiska modelleringsmetodiken. Dessa inkluderar pilottest av modelleringsstrategin, modellering av en enskild huvudtunnel samt modellering av en uppsättning deponeringstunnlar.

Glossary

The glossary is common to both Part 1 and Part 2.

Boremap is a core and borehole digital mapping system used by SKB during site investigation. All mapped Boremap parameters for rock and fractures are stored in SKB database SICADA.

Categorical variable is a discrete variable that has a limited set of classes, e.g. Thermal Rock Classes (TRCs).

Change of support, see upscaling.

Conditional simulation is a type of simulation where actual observations or measurements are honoured, i.e. the simulated value in a cell will be equal to the measured value at that specific location.

Cumulative Distribution Function (CDF) is a function describing the statistical distribution of a population. The value of the y-axis is the proportion of values that is lower than the x-value, i.e. the scale of the y-axis is from 0 to 1 (0 % to 100 %).

Declustering is a geostatistical technique to handle data that occur in spatial groups, so called clusters. Each data value is given a weight and clustered samples are given less weight than others. The weights are considered when the statistics are calculated, resulting in mean and variance that are more representative.

Gaussian simulation is a type of stochastic simulation where simulated values follow a Gaussian (normal) distribution. If measurements are not Gaussian, a Gaussian transformation must first be applied.

Hard data are data acquired by physical measurement, such as primary data; see soft data for comparison.

Heat capacity is the capacity for a material to store thermal energy. In the report, heat capacity is used in a general manner, including both specific and volumetric heat capacity.

Histogram is a graph that shows the distribution of the occurrence of different values, separated into a finite set of classes.

Indicator simulation is a stochastic simulation technique for simulation of different classes of a categorical variable. The classes are defined by indicators and cut-offs between the classes. Indicator variogram models are used to simulate the spatial occurrence of the different classes.

Kriging (linear) is an interpolation method, resulting in the best linear unbiased estimator. Under correct assumptions, the method gives a mean error equal to 0 and minimises the variance of the errors. Kriging is often the best option for making prognosis of mean properties (estimation) but it is not a good option for characterising uncertainty because of its smoothing effect; compare stochastic simulation.

Lag is the separation distance between classes of spatial data. In the variogram, the lag is plotted on the x-axis. The spatial correlation will usually decrease when the lag increases.

Markov chains describe the change of state in a system over space (or time). The changes of state are called transitions. The Markov property means that the conditional probability distribution of the state depends only on the state of the neighbouring cell. Markov chains can be used for calculating transition probabilities of categorical variables.

Nugget is the (apparent) intersection of the variogram with the y-axis, i.e. the variance at separation distance (lag) zero. It results from measurement errors and/or micro-variability at scales smaller than the sample support, appearing in the form of white noise (Journel and Huijbregts 1978).

Probability Density Function (PDF) is a function describing the statistical distribution of a population. The value of the y-axis is the probability density, which is the derivative of the Cumulative Distribution Function (CDF), i.e. the maxima and minima of a PDF correspond to the inflection points in a CDF.

Range is a distance representing the zone of influence of a sample. It is the distance on the x-axis where the variogram reaches a more or less pronounced plateau.

Rock domain is a region of the rock mass for which the properties can be considered essentially the same in a statistical sense. Several rock units are assembled into a rock domain.

Rock occurrence is a part of a drill core, with the same rock, having a length less than 1 m. It is used in Boremap for fine-scale classification of lithology (cf. Rock type).

Rock type is defined in a section of drill core with a length larger than 1 m. It is the rock type that has the largest proportion within the interval: Each rock used in Boremap has a SKB Rock type ID number, e.g. 101057 for granite to granodiorite (SKB MD 132.005).

Rock unit is term used for a volume of rock judged to have a reasonably statistically homogeneous distribution of lithology (rock types) and fracturing statistics. It may contain several different rock types judged to be similar. A rock unit may also contain small-scale inclusions of very different rock types. Each rock unit is defined by its location and is described in terms of rock type distribution and fracture statistics.

Sill represents the variance beyond the zone of influence, i.e. the value on the y-axis of the variogram at the plateau.

Simulation scale is the size of a grid cell in the simulation, i.e. the resolution of the model.

Simulation volume is the rock volume that is modelled, i.e. all the grid cells in a realisation from a stochastic simulation.

Soft data are data that are acquired in the form of expert knowledge, without physical measurements; see hard data for comparison.

Spatial correlation indicates the strength and direction of a linear relationship between two spatially separated data values. Two values are considered positively correlated if an increase in one value results in an increase of the other value. Spatial thermal conductivity data are positively correlated, up to the range.

Stochastic simulation is the general term for techniques for assigning random values to stochastic variables, according to a model describing the random properties. Spatial stochastic simulation is used when the values should be distributed in space, which requires a spatial model, e.g. a variogram model. It can be performed on continuous variables (such as thermal conductivity) or categorical variables (such as Thermal Rock Classes). Stochastic simulation is used when the uncertainty of a parameter must be quantified (uncertainty analysis or risk analysis); compare Kriging.

Subordinate rock is a general term and used in different ways depending on scale. Normally the term is used for rock types that are subordinate to the dominating rock type in a rock domain. The term rock occurrence is used for specific reference to sections of rock with borehole lengths less than 1 m.

Support is the term used for measurement scale in geostatistical nomenclature. It is the volume, shape, and orientation that a measurement represents (Starks 1986).

Thermal lithological modelling includes the statistical modelling of TRC lithology and the stochastic simulation of TRC lithology.

Thermal properties encompass thermal conductivity, thermal diffusivity, heat capacity, temperature, and the coefficient of thermal expansion. The stochastic modelling (Chapter 5 in Part 1) is restricted to the parameter thermal conductivity but there is a relationship between thermal conductivity and heat capacity, which implies that heat capacity also can be characterised by this approach. The other thermal properties are only described by data.

Thermal Rock Class (TRC) is a concept defined in this report. There are many rock types, but simplifications are required in the modelling and therefore thermal rock classes are defined. Each TRC will have a defined distribution and correlation structure of thermal conductivity values.

Transition probability is the probability of a change, a transition, between two states of a categorical variable. Example: The probability of transition from Ävrö granite to Quartz monzodiorote.

TRC lithology is the result of classifying lithological data into Thermal Rock Classes (TRCs). Thus, TRC lithology is equivalent to the lithology expressed as TRCs for the purpose of thermal modelling.

Unconditional simulation is a method that distributes simulated data spatially without honouring measurements at specific locations.

Upscaling, or change of support, refers to the change of the scale of data or simulated values. The upscaling results in a distribution (change in variance and sometimes even in the mean). The variance is reduced when the scale increases.

Variogram is a graph that describes how the variance changes as a function of separation distance (lag). A variogram illustrates the spatial correlation. A variogram model can be fitted to the experimental variogram to model the spatial correlation.

Contents

1 Introduction

1.1 Background

The Swedish Nuclear Fuel and Waste Management Co (SKB) is responsible for the handling and final disposal of the nuclear waste produced in Sweden. Forsmark has been selected as the site for the final repository for spent nuclear fuel.

The deposited nuclear waste canisters will emit heat due to radioactive decay and the thermal conductivity distribution in the rock will influence the temperature of the buffer surrounding the canister. A low thermal conductivity leads to a larger required distance between the canisters and tunnels than in the case of a high thermal conductivity. It is important to describe the thermal conductivity distribution of the rock in order to design the repository for a suitable buffer temperature. Attribution of thermal properties to the bedrock is referred to as thermal modelling. In this work, the combined use of measurements and simulations are employed.

This report presents a further developed thermal modelling methodology for description, prediction and visualisation. The new methodology, version 3.0 (Back and Sundberg 2022), is an updated[1] version of an earlier strategy applied in the final SDM-Site model versions resulting from the surface-based investigations at the Forsmark and Oskarshamn areas (see Back et al. 2007, Sundberg et al. 2008a, b). The previous methodology for thermal modelling is described in detail in Back and Sundberg (2007). The strategy was based on unconditional stochastic modelling of both lithology and thermal properties. One important aim of the current development was to improve how the spatial variability in lithology and thermal transport properties[2] is handled. In addition to the thermal conductivity, the heat capacity is also important but is less significant for the maximum temperature development.

The objective for the previous strategy was mainly to describe the thermal property distribution and other thermal aspects at rock domain level. Focus was on the lower tail of the thermal conductivity distribution. The new developed methodology is adapted so that it can also be applied to limited rock volumes during the repository construction and operational phases. More focus is on prediction of the thermal property distribution to enable future optimization of the repository's design. This requires conditioned stochastic simulation and local data.

1.2 Objective of the report

The objective of this report is to present background material, as a documentation of the development of the thermal modelling methodology of thermal properties in conjunction to the choice and preparation of deposition areas of the spent fuel repository at the Forsmark site. The thermal methodology is described in Part 1 (Back and Sundberg 2022) with supporting information (background material and development) in Part 2 (this report). The supporting information in this report refers to some of the most important steps of the thermal modelling methodology which have been developed gradually, from the early stages of the site descriptive modelling to the phase for development of deposition areas.

[1] Such updating is part of SKB's overall modelling strategy.

[2] The thermal property distribution must be characterised in a proper way without the assumption of a normal distribution. The reason for the deviation from a normal distribution is that different rock types have widely different thermal conductivities.

1.3 Outline of the methodology

The methodology for thermal modelling is presented in Figure 1-1, as well as in Part 1 (Back and Sundberg 2022). It consists of a procedure in ten steps:

1. Data collection.
2. Data interpretation.
3. Expert knowledge.
4. Statistical modelling of TRC lithology.
5. Statistical modelling of thermal properties.
6. Spatial thermal boundaries.
7. Stochastic simulation of TRC lithology.
8. Stochastic simulation of thermal properties.
9. Realisations of thermal properties.
10. Output to thermal dimensioning and evaluation of results.

The methodology, as outlined in Figure 1-1, is applied for each rock volume of interest, i.e. on sub-volumes of the repository. The approach was originally developed for modelling of a whole rock domain (Back et al. 2007) but it has now been adapted to include conditioned modelling of sub-volumes.

Starting at the upper part of Figure 1-1, the thermal modelling work starts with data collection (1). The captured data is prepared and interpreted in the next step (2). This data interpretation step includes the choice of simulation scale, defining Thermal Rock Classes (TRCs), preparing lithological data, classifying lithological data as TRCs, preparing thermal data, and upscaling (change of support) of thermal data (if required). The result of these sub-steps is a defined set of TRCs and corresponding thermal data for each TRC, representing the simulation scale, i.e. the resolution of the thermal model.

Expert knowledge and input from existing models (3), mainly geology, will influence the data interpretation step (2) as well as the modelling steps (4 to 10). For example, thermal models from the SDM can be used as a starting point, with new data being used to update and refine the models.

Expert knowledge (3) is imperative, both during the data interpretation and the modelling steps. Such expert knowledge is utilized to bring the general knowledge down to the model level and to compensate for bias in the data. More explicitly, expert knowledge is applied for defining TRCs, for the statistical modelling of TRC lithology (4) and thermal properties (5). The statistical modelling of TRC lithology utilizes lithological data to construct models of the transition between different TRCs, thus describing the spatial statistical structure of each TRC. The result is a set of transition probability models that are used in the lithological simulation of TRCs (7). This simulation can be either conditional or unconditional. The intermediate result of this stochastic simulation is a number of realisations of TRC lithology, each one equally probable.

A prerequisite for the simulations is that the spatial boundaries have been defined (6). There are computational limitations on how large volume of rock can be simulated, and consequently, the modelling will have to be performed on sub-volumes. This is also in accordance with application of the methodology on a successively evolving repository during the construction and operational phases.

Based on the thermal data, statistical thermal models are constructed for each TRC (5). A statistical thermal model consists of both a statistical distribution model and a variogram model. These are used in the stochastic simulation of thermal conductivity (8), either conditional or unconditional, and the result is a number of equally probable realisations of thermal conductivity for each TRC. Steps 4 and 7 can be carried out in parallel with steps 5 and 8 because they are independent.

In the next step (9), the realisations of TRCs (geology) and thermal conductivity are merged, i.e. each realisation of geology is filled with simulated thermal conductivity values. The result is a set of realisations of thermal conductivity that considers both the difference in thermal properties between different TRCs and the variability within each TRC. In addition, realisations of other thermal properties, primarily heat capacity (specific or volumetric), can be created in step 9 by applying a statistical relationship between thermal conductivity and heat capacity.

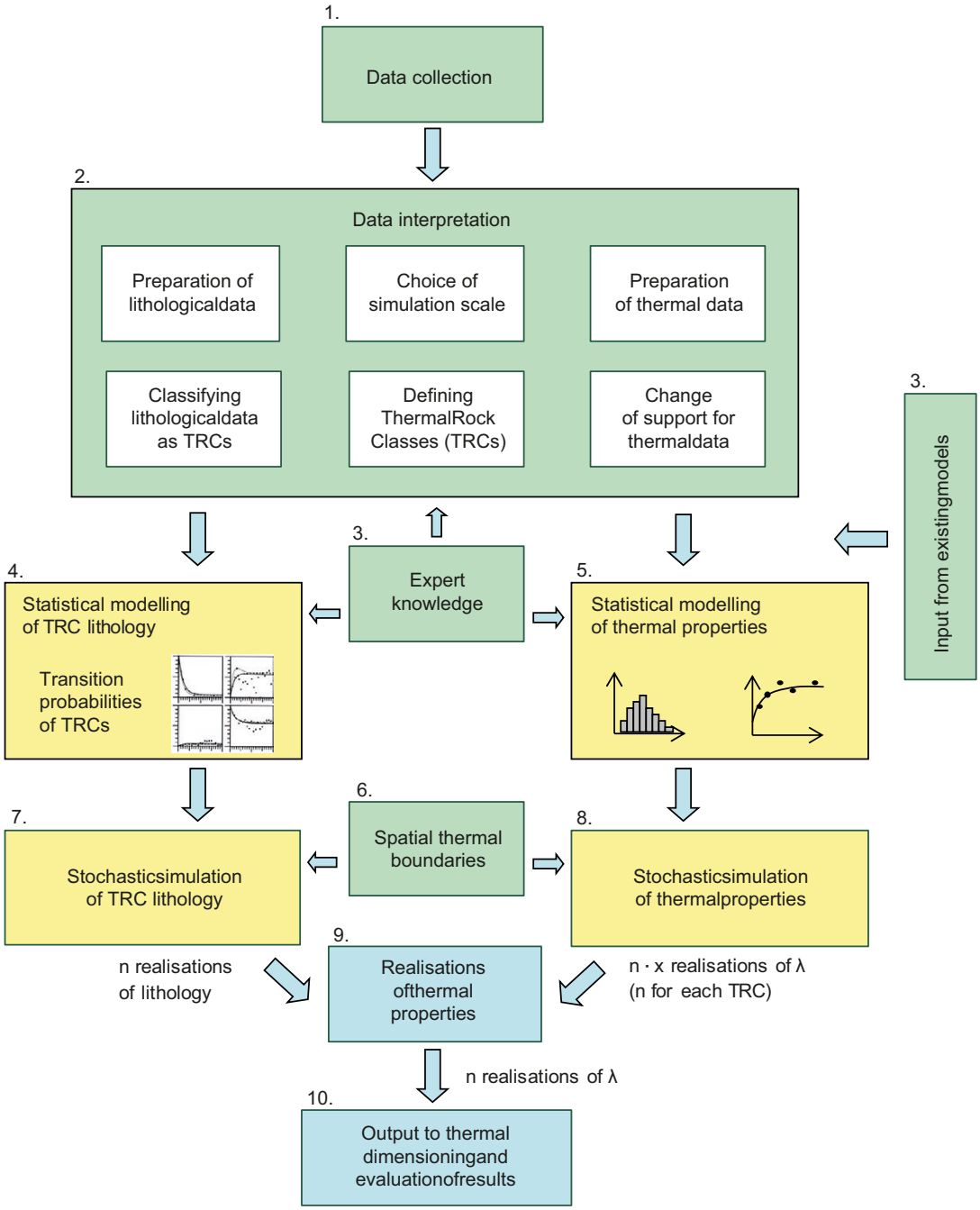


Figure 1-1. *Schematic description of the procedure for thermal modelling. The figure illustrates the workflow when applying the methodology, i.e. each box represents a certain work package.*

As the last step in the methodology (10), the realisations can be used directly in the thermal dimensioning. In addition, the modelling results can be analysed and presented in several ways, both regarding thermal and lithological properties in the modelled rock volume, the most obvious being predicted thermal conductivity values in sub-volumes of rock. Other examples include 3D illustrations for visualization, histograms and statistical parameters (descriptive statistics), proportions of different TRCs, typical lengths and volumes of rock bodies, confidence intervals for specific properties, etc. If the result is desired at a scale different from the simulation scale, upscaling of the realisations can be performed to a desired scale appropriate for the thermal design and dimensioning of the repository.

1.4 Outline of the report – Part 2

The report provides background information about the development of the thermal modelling methodology, primarily the most important steps of the methodology and the ones that have changed the most over time. Those steps are described in separate chapters, following the outline of the methodology (Section 1.3) as implemented in Part 1. The report should preferably be read in conjunction with Part 1, as a complement.

In Chapter 2, background information is presented on the choice of simulation scale (step 2 in Figure 1-1). The scale issue is imperative and is discussed from various perspectives. Chapter 3 describes the development of the concept Thermal Rock Classes (TRCs), and how the definition has changed. Background information on how these TRCs are modelled and simulated (step 4 and step 7) is presented in Chapter 4. This chapter includes software-specific information and results of simulation tests. In Chapter 5, the various types of anisotropy in thermal properties in Forsmark are discussed, with evaluation and conclusions. Chapter 6 is a compilation of information on software for thermal statistical modelling (step 5) and simulation (step 8), including qualitative comparisons and conclusions. The report ends with a chapter discussing possible applications of the thermal modelling methodology, i.e. application scenarios.

2 Choice of simulation scale

2.1 Objective

The objective of this chapter is to suggest a desired simulation scale for the final stage in the thermal stochastic modelling. The choice of simulation scale is an important step in the thermal modelling; see Part 1 (Chapter 5 in Back and Sundberg 2022).

2.2 Scales in thermal modelling

2.2.1 Different scales

There are several different scales involved in the thermal simulations and the modelling. The most important ones are:

1. The scale that data represents:
   a. Scale (or scales) for lithological data.
   b. Scale (or scales) for thermal data.
2. Cell size in simulations (simulation scale):
   a. Cell size at lithology simulations (TRC).
   b. Cell size at thermal simulations.
   c. Common cell size (merged).
3. Size of the simulation volume:
   a. Simulation volume for lithology simulations (TRC).
   b. Simulation volume for thermal simulations.
   c. Common simulation volume (merged).
4. Desired scale for output from the simulations.

These different scales are discussed below.

2.2.2 The scale that data represents

Data can represent different scales. Here, data refers to both lithological data and thermal measurements. Upscaling may be required for several reasons:

- To reach the same scale for lithological data and thermal data.
- To achieve a scale that is reasonable from a simulation point of view (this was done during the earlier site descriptive modelling).

Upscaling implies that uncertainty is introduced and should therefore be avoided if possible. In other words, it is preferable that the data represent the same scale as the cell size at the simulations (simulation scale).

Thermal data in the range of 0.01–0.2 m will almost certainly require upscaling (existing TPS, Transient Plane Source, data are in cm scale). However, upscaling can probably be avoided if data represents a larger scale. Thermal measurement data at a scale of about 0–1 m (currently the range is from 0.12 m and larger) can possibly be obtained from boreholes through so-called DTS-technique (Distributed Temperature Sensing) combined with heating; the infinite line source theory (Carslaw and Jaeger 1959).

TRC data at the same scale as above requires an algorithm (decision rule) for the assignment of TRCs to particular simulation cells. In the site-descriptive modelling for Forsmark, this was done by a random procedure in the 1 m simulation scale, along a limited borehole length. This approach involves uncertainties due to random effects. The approach to assigning TRCs is improved in the updated strategy (Back and Sundberg 2022).

2.2.3 Cell size in simulations

The cell size in the simulations is identical to the resolution of the simulations. It is also referred to as the simulation scale. In principle, the simulations of TRC lithology and thermal properties could be made on different scales with subsequent upscaling to a common scale. In the strategy however, it is assumed that the same cell size is used in both types of simulation.

The appropriate cell size depends on:

- The scale the data represents.
- The size of the simulation volume.
- Computing capacity.

There is a mutual dependence between data scale and cell size. If data is available at a given scale, the cell size may need to be adapted to it. As an example, if the measurement data represents the scale 1 m (assumption) then it is advisable to also use the cell size 1 m in the simulations. On the other hand, if the cell size has been determined, then data need to be adopted to that scale.

The cell size has a minimum limit, as a result of limited computing capacity. In the site descriptive modelling for Forsmark, the cell size was 1 m^3 and the simulation volume 50 × 50 × 50 m^3, equivalent to a model size of 125 000 cells.

2.2.4 Size of the simulation volume

Important factors that may limit the simulation volume are cell size, number of TRCs, complexity of the lithology and computing capacity. There is a mutual dependence between cell size and simulation volume: An increased simulation volume may require an increased cell size in order to keep the number of cells at a manageable level. However, this may influence the definition of the different TRCs.

In the case of conditioned simulation, the simulation volume should primarily be decided to match the physical rock volume of interest. This means that the simulation volume may deviate from the volume where the properties of the rock domain are reasonably constant. The consequence is that it is not necessarily the lithological properties of the rock domain that should be recreated during simulations, rather the rock distribution in the physical volume of interest.

2.2.5 Desired scale for output from the simulations

The simulations generate realisations with a resolution of the cell scale (simulation scale). These realisations can be used directly as input to the desired application. Additional simulation volume could be necessary, e.g. to meet boundary conditions in an application for temperature and heat transport calculations.

2.3 Framing of appropriate scales

2.3.1 Method

As discussed above, scale issues encompass (1) the scale that data represents, (2) the cell size in simulations, and (3) the size of the simulation volume. All these issues are interdependent and important in the strategy. In addition, there is also the desired scale for application of the simulation result; see Section 2.2.5.

One way to frame the appropriate scales is to perform a type of backcasting. The approach is as follows:

A. Begin the framing at a specific starting point; see below.

B. Determine the consequences regarding scale based on this starting point.

C. Determine the possible interval of the scale.

The consequences determined in step B is general consequences of various types, as discussed in the following sections. These steps are then repeated for different starting points, each resulting in a suggested interval for the different scales. As a last step, all scale intervals are combined, and an assessment is made regarding the most appropriate scales to use in the strategy.

Five different starting points can be used:

1. Starting point; Thermal data.
2. Starting point; Lithology and subordinate rock types.
3. Starting point; Required simulation volume in the repository.
4. Starting point; Simulation time (computing limitations).
5. Starting point; Desired scale for output.

These starting points are evaluated below.

2.3.2 Thermal data

Existing small-scale data

In the site descriptive modelling in Forsmark, the thermal main data set was based on TPS-data in cm scale. Both thermal data and parts of the lithological data were present in a small scale, which gave a cell size of 0.1 m as a reasonable compromise. This gives a simulation volume much too small to be able to reproduce the domain properties. This problem was solved by upscaling, using stochastic simulation and subsequent averaging of the thermal properties by the SCA method; see Appendix A and Sundberg (1988). For practical reasons, 1 m^3 was chosen as the resulting cell size, i.e. upscaling was performed from the data scale to 1 m. This was performed for both lithology and thermal properties, although with different upscaling algorithms. The final simulations were made as a second step, using a cell size of 1 m and simulation volume of 125 000 m^3.

The number of lithological classes was four in the earlier Forsmark site descriptive modelling (limited to five in the software used). The number of rock types was approximately 3–4 times as many. This was solved by creating thermal rock classes of sets of rock types with similar thermal properties in each class. Rock types included in one TRC may not occur close to each other in reality. The thermal conductivity distribution in each TRC were based on the domain proportions of each rock type. This means that the thermal conductivity distribution may be different at a location (e.g. a single deposition tunnel) with slightly anomalous composition compared to the rock domain as a whole.

The thermal data gave thermal conductivity distributions for TRC. For the stochastic modelling, the spatial correlation was also required. Variogram models for the different thermal rock classes were created from density loggings, and for the main rock type also from TPS data. The use of density loggings was possible, since a relationship between thermal conductivity and density had been established earlier (see Back et al. 2007, Section 3.4). However, the relationships have larger uncertainties compared to the Laxemar site and especially for some TRCs the relationship was weak.

The primary purpose of calculating variograms based on density loggings is to estimate the range, i.e. the separation distance over which spatial dependence is apparent. In other words, samples that are separated by a distance less than the range are considered having spatially correlated thermal conductivity values.

The thermal conductivity distributions and variogram models may be possible to reuse in the initial steps in future modelling. The data on thermal conductive does not put limitations on the upscaling.

Possible large-scale data

Thermal data at a larger scale may possible be acquired for boreholes in the repository in Forsmark by the so-called DTS method with integrated heating at a scale of about 0.1–1.0 m. Data will be both continuous and represent a larger scale. The former enables much better quality regarding the spatial distribution of thermal conductivity, which significantly reduces uncertainty in the variogram modelling. In addition, the larger scale data may enable a one-step modelling without the need of upscaling.

However, larger scale data may imply a redefinition of the thermal rock classes (TRCs). Thermal data over e.g. 1 m will involve different rock types compared to the original approach in the site descriptive modelling. New TRCs including subordinate rock types may need to be developed. Thus, the TRCs in the developed strategy may deviate from the TRCs used in the original approach.

2.3.3 Lithology and subordinate rock types

The scale of lithological data is basically identical to the resolution of the borehole mapping (previously 0.01 m). The relevant TRC scale depends on the size of dykes of subordinate rock types. This would require a cell size of approximately 0.05 – 1.0 m for good representation of subordinate rock types in the simulations.

Irrespective of size, mafic rock (mainly amphibolite veins and dykes) make up approximately 5 % of the rock mass at Forsmark. Approximately 2 % of the rock mass (all rock domains considered together) consist of these rock types having a true thickness of less than 0.5 m (Stephens et al. 2007). Approximately 0.2 % occur in the size range of 0.01 to 0.05 m (Back et al. 2007, Figure 5-2).

For practical reasons, the selected TRC scale should match the scale of the thermal data. Consequently, the scale of thermal data and the scale of subordinate rock types are interdependent. A conclusion is that a common scale should be around 0.5 – 1 m. A scale much smaller might lead to an unrealistic large number of cells and a scale much larger implies insufficient representation of subordinate rock types.

However, if thermal data at larger scale are used (up to 1 m), the scale for the TRCs may need to be changed. Table 2-1 shows TRCs used in the earlier site descriptive modelling.

At a larger scale, e.g. the amphibolites may need to be represented in a different way compared to the original approach. This may be performed by defining one TRC based on the main rock type, including small dykes of amphibolite, and one TRC with larger dykes at a size relevant at the simulation scale.

Table 2-1. TRCs for rock domain 29 used in Back et al. (2007).

TRC	Rock name/code	Proportion in domain 29 (%)	Mean thermal conductivity (TPS)	Composition, mode of occurrence, etc.
57	Granite to granodiorite, 101057	74	3.68	Both felsic group B rocks. Dominating granites.
	Granite, aplitic, 101058	1	3.85	
51	Granodiorite to tonalite, 101051	5	2.85	Felsic to more intermediate group A and C rocks.
	Felsic to intermediate volcanic rock, 103076	0.4	2.54	
61	Pegmatite, pegmatitic granite, 101061	13	3.33	Both felsic group D. rocks. Late tectonic dykes, segregations, veins.
	Granite, 111058	2	3.47	
17	Amphibolite, 102017	4	2.33	Both mafic group B rocks.
	Diorite, quartz diorite and gabbro, 101033	0.2	2.28	Dykes and small irregular bodies.

2.3.4 Required simulation volume

The required simulation volume is dependent on how long time period the forecast covers and the number of pilot holes drilled per year (or tunnels excavated per year during operation), Table 2-2.

Table 2-2. Example of assumptions needed to determine the relevant simulated volume.

Forecast	1–3 years
Pilot holes for deposition tunnels	5–10 per year
Deposition tunnels	5 per year
Operation	200–250 canisters/year
Deposition tunnel, length	300 m
Distance between tunnels	40 m
Resulting forecast volume	1 year: 5–10 tunnels or borehole: assume 7 3 years: 15–20 tunnels or borehole: assume 20

The assumptions regarding forecast, investigations, and operation results in a required simulation volume that covers approximately 7–20 tunnels. The simulation area is 12 000 m^2/tunnel (300 × 40 m) and at a simulation height of 80 m this means approximately 1 × 106 m^3/tunnel. With 7–20 tunnels, the corresponding prognosis volume is 7–20 × 106 m^3. This is 50–170 times larger than the simulation volume during the site descriptive modelling.

However, during the initial stages of the construction work much smaller volumes are likely to be modelled and verified, maybe a volume corresponding to one tunnel (prognosis volume: 1 × 106 m^3). Such a volume is only 8 times larger than in the earlier simulations.

2.3.5 Simulation time (computing limitations)

Computing limitations have the potential to affect simulation and computing times in several ways:

- Simulation time in thermal simulations.
- Simulation time in lithological simulations.
- Calculation time in post-processing, i.e. problem-specific algorithms developed specifically for the thermal modelling.

Of these three, it is believed that the lithological simulations are most computer intensive and thus most sensitive to limitations in computational capacity. Thus, the simulation time for the lithological simulations is probably critical for deciding the number of cells in the simulation volume.

To investigate calculation times for stochastic geological simulation of rock volumes with software T-PROGS, simulations of different model sizes and with and without conditioning have been performed; see Chapter 4. The simulations have been carried out as follows: Data for rock domain 45 in Forsmark has been used. Four (4) thermal rock classes have been used (the maximum number in T-PROGS is five material classes). The same correlation properties were used as in the simulations conducted in 2007.

The results can be summarized as follows:

- Simulation times increase linearly with the number of cells for cubic unconditional models.
- For a cuboid unconditional model, the simulation time is shorter than what would have been predicted for a similarly large cubic model.
- Simulation times for conditional simulations were identical to the simulation times for the unconditional simulations with the same model.
- The simulation time for a cubic model with 10^6 cells was 40 minutes and the simulation for the large cuboid model ($720 \times 160 \times 40 = 4.6 \times 10^6$) took 157 minutes.

2.4 Conclusions regarding scale

2.4.1 Scale problem

In conditioned simulation, the scale problem is accentuated compared to unconditioned simulation. The scale inflicts the strategy in a complicated manner.

Simulation scale: On the one hand, we must have a sufficiently large simulation scale to be able to simulate a sufficiently large volume with reasonable simulation times. On the other hand, e.g. the assignment of subordinate rocks to simulation scale 1 m, may create shortcomings in the representation of e.g. amphibolites.

Data scale: Avoiding upscaling of thermal data has several advantages but also limitations. The main advantages are:

- Less complex modelling.
- Uncertainties resulting from the upscaling procedure are eliminated.

The limitations are related to data; larger scale thermal data is required in order to avoid upscaling.

I both cases above, new TRCs will probably need to be defined. However, the number of TRCs are limited to five in the software used.

2.4.2 Thermal data

Existing thermal data are in the cm scale and would need some sort of upscaling in order to be used properly in simulations at a larger scale. The upscaling procedure involve upscaling of both the thermal conductivity distribution and the spatial correlation structure. In the earlier Forsmark site descriptive modelling, upscaling of the thermal distribution was performed using stochastic simulation at the dm scale.

Main conclusions are:

1. The thermal conductivity distributions and the variogram models are possible to reuse, especially in the initial steps in future modelling. However, uncertainties exist, especially in the variogram models and these should preferable be improved.
2. Continuous profiles of new data enable much better modelling of the spatial correlation structure, i.e. the variogram models of thermal conductivity can be significantly improved. This would significantly reduce the uncertainty in the results compared to a reuse of the existing models.
3. New and larger scale data may enable one-step modelling, without the need of upscaling, and improve existing thermal conductivity distribution models with site specific data.
4. Larger scale data may influence how the different TRCs should be defined. From a thermal point of view there are no limitations on the scale. Data can always be scaled up but with the weaknesses mentioned above.

2.4.3 Lithological data

The lithological data are in all scales, from cm scale and upwards. The main conclusions are:

- Upscaling of small-scale lithological data is not problematic provided that the target scale is limited; see below.
- The target scale may be limited with respect to how the subordinate rock types, e.g. amphibolites and pegmatites, are represented.
- Larger scale of data may affect how the TRCs are defined (see below), i.e. which rock types will belong to a specific TRC. For example, there will be more focus on the location of subordinate rock types relative to the main rock types.

2.4.4 TRCs

The simulation scale and larger scale data may affect the classification of TRCs, especially if the scale reaches 1 m or more. For example, one TRC may include both the main rock type and small dykes of amphibolites, and another TRC will solely include larger amphibolites at a size relevant at the simulation scale. However, there is a limitation to 5 TRCs in the software used and new TRCs have consequences for the distribution of thermal data for each TRC.

2.4.5 Required simulation volume and computer capacity

- With 7–20 tunnels in a simulation, the resulting prognosis volume is $7–20 \times 10^6$ m^3 (based on a simulation volume per tunnel in the order of 1×10^6 m^3 (300 × 40 × 80). This is 50–170 times larger than the simulation volume during the site descriptive modelling and may be challenging for software and computer capacity.
- During the initial stages of the construction work, much smaller volumes are likely to be predicted and verified, maybe a volume corresponding to one tunnel. Such a volume is only 8 times larger than in the earlier simulations.
- Tests of computer simulation time for previous Forsmark data indicate that the calculation time increases linearly with the number of cells for cubic models (less for cuboid models). There are possibilities to further reduce processing time with computers designated for processing purposes.
- Simulation scale in the simulations should preferably be 1 m or larger, given the required simulation volume (0.5 m may be sufficient for smaller volumes).

2.4.6 Simulation volume for output to end users

The size of the simulation volume should consider the end user's requirements, for example regarding boundary conditions.

2.4.7 Desired simulation scale

Taking all aspects above into account, the preliminary conclusion is that the simulation scale for the modelling should be 1 m or larger. However, this scale is larger than the occurrence of important subordinate rock types, which may influence the description of TRCs and the modelling results. Test modelling in various scale are therefore recommended before a final decision is made regarding simulation scale and the division of the lithology into TRCs.

3 Definition of thermal rock classes

3.1 The TRC concept

Thermal rock classes (TRCs) are defined in order to facilitate the modelling and simulation of lithology. A TRC is a categorical entity without defined spatial boundaries between each class. The reason that TRCs need to be defined instead of using the rock types directly is that there is a large number of rock types but only a limited number of classes can be handled in the simulation of TRC lithology (Chapter 4) due to software limitations. By defining TRCs the complexity of the simulations can be kept at a reasonable level. The TRCs implies a rougher classification than rock types but the TRCs are sufficiently detailed to handle thermal conductivity. The TRCs are not necessarily generic, they may differ between different rock volumes, depending on the occurrence of rock types.

Two alternatives of Thermal Rock Classes have been developed. The first alternative, or definition, was used during the site investigations and site descriptive modelling (Section 3.2). A new TRC definition is proposed to be used during construction and operation of the repository (Section 3.3).

The following three sections of Chapter 3 elaborate various implications of the new TRC definition; scale issues (3.4), lithological implications (3.5), and thermal implications (3.6). The new TRC definition is tested in Section 3.7 and in Section 3.8 a concluding summary is given regarding the new TRC definition.

3.2 TRC definition for SDM

In the site descriptive modelling, the following was taken into consideration when defining TRCs (Back and Sundberg 2007):

- The most important rock types from a thermal point of view were defined as separate TRCs. The importance of a rock type is determined based on (1) how common its occurrence is in the rock domain, and (2) the shape and absolute values of the thermal conductivity distribution. Rock types with low thermal conductivity values are most important.
- Rock types with similar thermal conductivity are grouped into one single TRC, if required. The reasons for combining them are documented. The similarity is assessed based on plotted histograms and calculated statistical parameters.
- Rock types with very different spatial variability and correlation should not be grouped in the same TRC. These aspects can be assessed by comparing variograms of the rock types.
- Geological aspects, such as composition, age, genesis, and mode of occurrence are considered when a TRC is defined. If these aspects differ a lot between two rock types, they should only be combined if the difference in thermal conductivity is insignificant.

A check of the physical properties of the different rock types of a TRC can be performed, e.g. by comparing density values, plotted Streckheisen diagrams etc for tentative compositions of the rock types. Such tests can increase the confidence in the expected thermal conductivity of a TRC.

The definition of TRCs for rock in domain 29 (RFM029) is given in Table 3-1, for the site descriptive modelling. The code for a TRC is defined by using the two last digits of the rock code for the dominating rock type in that class.

Table 3-1. Thermal Rock Classes (TRCs) for rock domain 29 at Forsmark during the site descriptive modelling (Back et al. 2007).

Thermal Rock Class	Rock types	Rock codes
TRC 57	Granite to granodiorite	101057
	Granite, aplitic	101058
TRC 61	Pegmatite	101061
	Granite, fine to medium grained	111058
TRC 51	Granite, granodiorite and tonalite	101051
	Felsic to intermediate volcanic	103076
TRC 17	Amphibolite	102017
	Diorite	101033

3.3 New TRC definition for construction and operational phases

The new definition of TRCs, for application during the construction and operational phases of the repository, assumes that it is possible to use *rock type* in Boremap as a basis for defining TRCs, and at the same time include *rock occurrences* that exist within the simulation scale in each borehole. Thus, the TRCs are coupled to a defined scale, preferably the simulation scale. This is a major difference compared to the SDM definition in Section 3.2.

The following definition is suggested for new TRCs for rock in domain RFM029, for the construction and operational phases:

- TRC-57, comprising rock types 101057 and 101058, and all other rock occurrences.
- TRC-51, comprising 101051 (rock code 103076 excluded; see Section 3.5.1) and all rock occurrences.
- TRC-61, comprising 101061 and 111058, and all rock occurrences.
- TRC-17, comprising 102017 and 101033, and all rock occurrences.

This definition offers a number of advantages for the thermal modelling compared to the SDM definition in Section 3.2. One important advantage is that Boremap data of rock type can be used directly to assign TRCs to data at the simulation scale (suggested scale is 1 m). In addition, rock occurrence in Boremap can also be used directly. This eliminates many of the uncertainties and difficult assessments in the SDM definition and simplifies the process of defining TRCs. It is also more in accordance with the geologists' classification of rock.

The new TRC definition is further discussed in the following sections. It was tested in Section 3.7 and considered successful. This TRC definition is therefore recommended for the construction phase of the repository and it is the approach suggested in Part 1 (Back and Sundberg 2022).

3.4 Scale issues for new TRCs

In conditioned simulation, the scale problem is accentuated compared to unconditioned simulation. The scale impacts the strategy in a complicated manner. These issues are discussed in Chapter 2 (Choice of simulation scale) and the main conclusions are summarized below.

Simulation scale: On the one hand, a sufficiently large simulation scale is required to be able to simulate a sufficiently large volume with reasonable simulation times. On the other hand, the assignment of subordinate rocks to a simulation scale of for example 1 m, may create shortcomings in the representation of some rock types, such as amphibolites and pegmatites, which generally occur with a borehole thickness of less than 1 m.

Data scale: Avoiding upscaling of thermal data has several advantages but also limitations. The main advantages are:

- Less modelling steps.
- Uncertainties in the simulation output resulting from the upscaling procedure are eliminated.

The limitations are related to required data; avoiding upscaling requires thermal data of a larger scale, and a sufficient amount of such data.

The main conclusions from a thermal point of view are:

1. The thermal conductivity distributions and the variogram models are possible to reuse. However, uncertainties exist, especially in the variogram models.
2. Continuous profiles of new field data enable much better modelling of the spatial correlation structure. This would significantly reduce the uncertainty in the results compared to a reuse of the existing models.
3. New and larger scale data may enable one-step modelling, without the need of upscaling, and improve existing thermal conductivity distribution models with site specific data.

The main conclusions from a lithological point of view are:

- Upscaling of small-scale lithological data is not problematic provided that the target scale is limited; see below.
- The target scale may be limited with respect to how the subordinate rock types, e.g. amphibolites and pegmatites, are represented.
- Larger scale of data may affect how the TRCs are defined, i.e. which rock types will belong to a specific TRC. For example, there will be more focus on the location of subordinate rock types relative to the main rock types.

Taking all aspects above into account, the conclusion in Chapter 2 is that the desired simulation scale for the modelling should be 1 m or larger. However, this scale is larger than the occurrence of important subordinate rock types, which may influence the description of TRCs and the modelling results.

3.5 Lithological implications of new TRCs

3.5.1 Subordinate rocks in Forsmark

The dominant rock types within the planned repository area in Forsmark are different in the different rock domains. Granite to granodiorite, metamorphic medium-grained (101057) is dominating in domain 29 (RFM029) and granite commonly affected by albitization, fine-grained and metamorphic, dominate in domain 45 (RFM045) (101057_104 or 101058_104). Subordinate rocks of importance are amphibolite (102017), granodiorite-tonalite (101051) and pegmatite (101061). These occur both as rock type and as rock occurrence. Lithological borehole data in Boremap are divided into two data sets: rock type (> 1 m borehole length) and rock occurrences (< 1 m borehole length).

To explore the possibility of integrating rock occurrences into new TRCs, discussions were held with site geologists. The following conclusions could be drawn (Jesper Petersson, personal communication) and are in general supported by Back et al. (2007) and Stephens et al. (2007):

- Pegmatite (101061) occurs as concordant veins, mainly in 101057, and dykes which runs in all rocks, which are often discordant. Rock type 111058 occurs in a similar manner to pegmatites for discordant dykes. Pegmatite dykes definitely appear as swarms and to some extent also as veins, although the latter are relatively evenly distributed.
- Amphibolites occur mainly in 101057 and occur in varying ways, as intruded bodies of varying thickness and as bulbous accumulations.
- Rock type 101051 often appear as larger bodies and extends from a few meters up to tens of meters. These may be relatively homogeneous, i.e. tonalitic, but this should be checked. Rock type 103076 (part of TRC 51) occurs only in the outer margins of the Forsmark tectonic lens and should be excluded.

The conclusion from a lithological point of view is that it seems reasonable to integrate *rock occurrences* into TRCs. There is an already established criterion for separating rock occurrence and rock type: the 1 m length criterion. If this could be used also in the TRC definitions, it would further simplify the work.

3.5.2 Analysis of subordinate rocks

There is a need to investigate e.g. the size distribution of pegmatite and amphibolite to find out potential obstacles and difficulties. Also, the verification of the assumption that 101051 is homogeneous within each body is an issue.

Size distribution of pegmatite

The borehole length distribution of pegmatites is based on data from 12 boreholes: KFM01A, 01C, 01D, 04A, 05A, 06A, 07A, 07B, 08A, 08C, 09B, and 10A (almost all used in Forsmark Thermal site descriptive modelling 2.2, Back et al. 2007). Data comes from Boremap and encompass both rock occurrence and rock type:

- Total borehole length is 7 895 m.
- Rock domains 29 and 45 only. No distinction made between these two domains.

The Boremap data is divided into even lengths with even increments of 0.5 m; see Figure 3-1 and Table 3-2 where the data are in proportions to borehole length. In Figure 3-2, the same data are illustrated but with higher resolution and non-even increments. The data from Figure 3-1 is also illustrated in proportions to total borehole length; see Figure 3-3.

The following conclusions can be drawn:

- Pegmatite makes up 13.9 % of the total analysed borehole length.
- Most pegmatites are small, 45 % less than 0.5 m.
- Relatively small proportion around 1 m, which reduces the risk of discretization errors.
- The results are not in conflict with modelling using a simulation scale of 1 m and the integration of pegmatite rock occurrence in the main rock type.

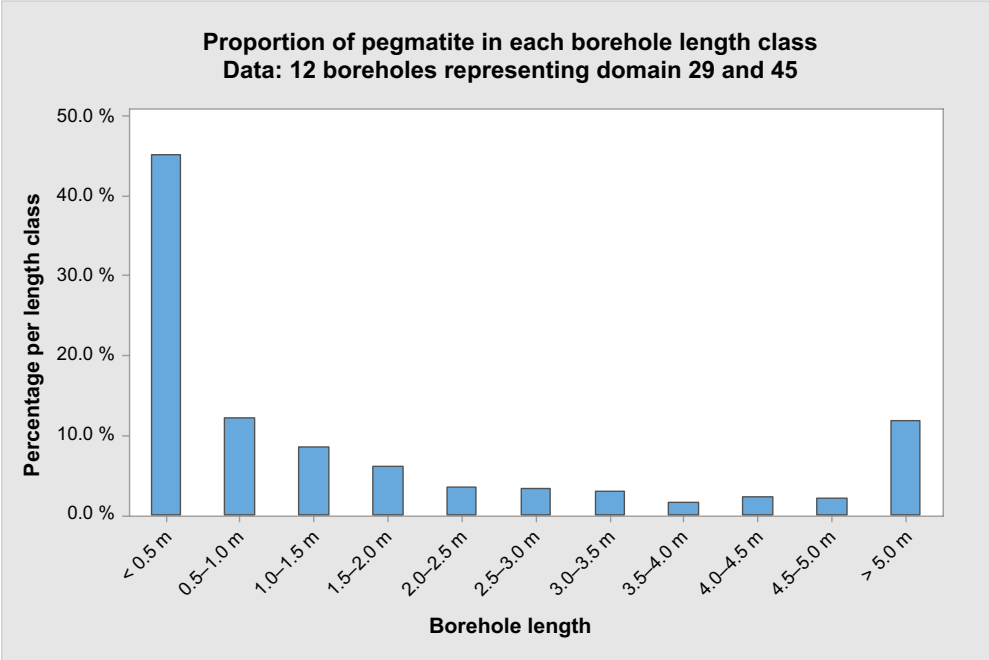


Figure 3-1. *Proportion of pegmatite in each borehole length class with even increments.*

Table 3-2. Proportion of pegmatite in each borehole length class with even increments.

Borehole length interval	Percentage pegmatite (101061) in different length classes relative to total borehole length in 12 boreholes	Proportion of pegmatite in each length class
< 0.5 m	6.24 %	45.1 %
0.5–1.0 m	1.69 %	12.2 %
1.0–1.5 m	1.18 %	8.49 %
1.5–2.0 m	0.86 %	6.21 %
2.0–2.5 m	0.49 %	3.54 %
2.5–3.0 m	0.48 %	3.44 %
3.0–3.5 m	0.42 %	3.04 %
3.5–4.0 m	0.24 %	1.72 %
4.0–4.5 m	0.32 %	2.34 %
4.5–5.0 m	0.30 %	2.15 %
> 5 m	1.63 %	11.8 %
All lengths	13.9 %	100 %

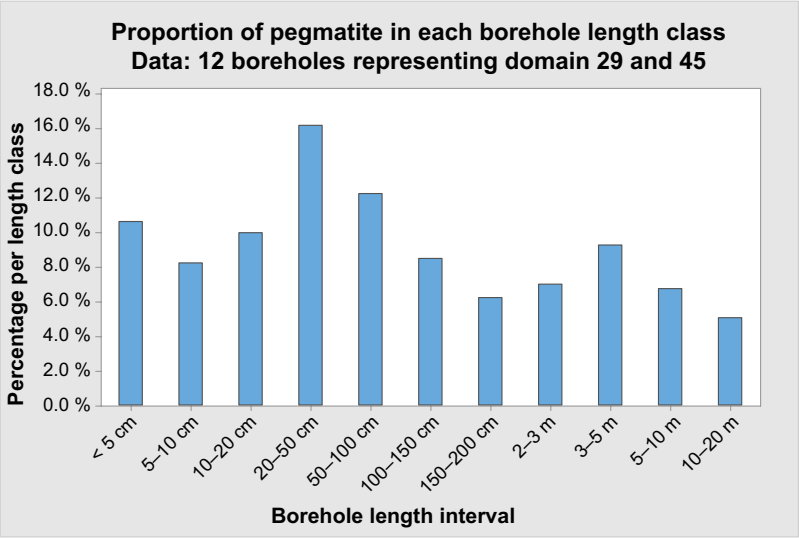


Figure 3-2. *Proportion of pegmatite in each borehole length class with higher resolution for lengths 0–100 cm.*

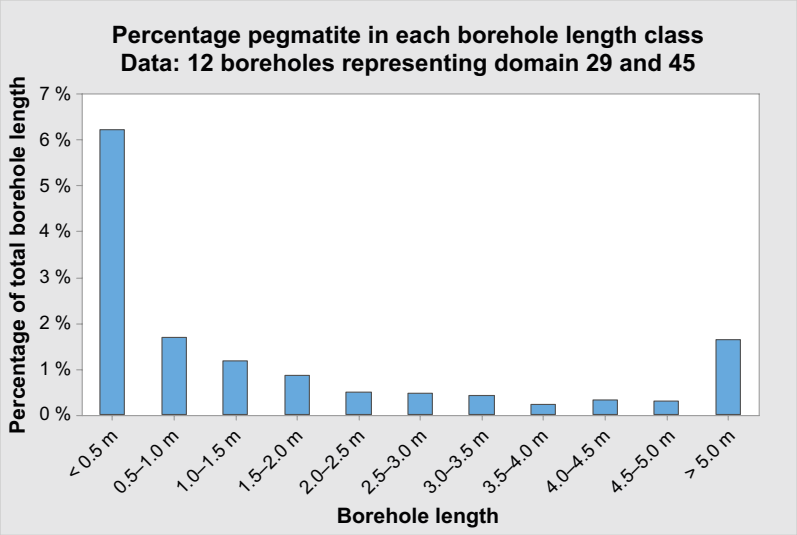


Figure 3-3. *Proportion of pegmatite in each borehole length class (in proportions to total borehole length).*

Pegmatite – Dyke or vein

Frequency of different types of pegmatite are shown in Figure 3-4 to Figure 3-6, based on data from 12 boreholes: KFM01A, 01C, 01D, 04A, 05A, 06A, 07A, 07B, 08A, 08C, 09B, and 10A (almost all used in Forsmark Thermal site descriptive modelling 2.2). Data comprises only rock occurrences.

A dyke forms when a magma intrudes into a fracture and then crystallizes. A vein is formed by crystals growing from the fracture wall toward the middle, successively filling the void. Veins are normally smaller, and dykes are larger. When mapping pegmatites these two types of pegmatite were identified. They are largely (but not exclusively) differentiated by their lengths in the drill core; veins were set to 0–20 cm and dykes were set to 20–100 cm if evidence for intrusion is visible in the drill core. Rock occurrences that could not be classified as a vein or a dyke were mapped as "*unspecified*".

The following conclusion can be drawn from the rock occurrence data:

- The number of pegmatite veins is much larger than the number of dykes.
- They are largely (but not exclusively) differentiated by their lengths in the drill core (veins are smaller).
- A substantially large proportion of the pegmatites is not categorized as vein or dyke. These unspecified pegmatites are mostly below 0.1 m in borehole length, indicating that "unspecified" may have originated from veins or small pegmatites.

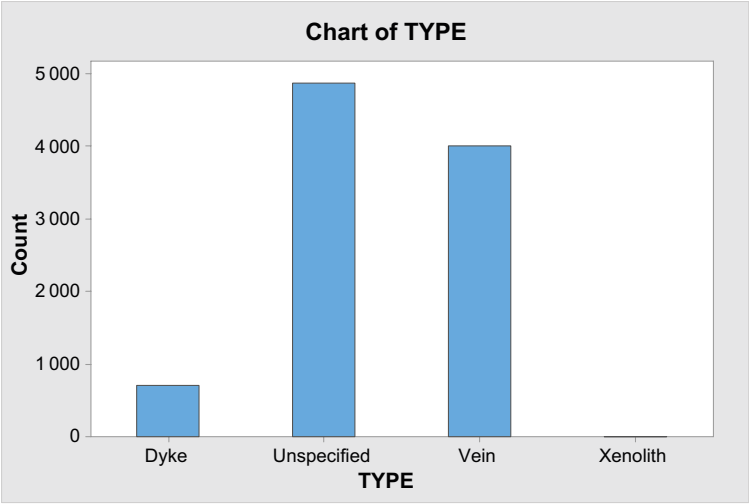


Figure 3-4. *Subdivision of number of pegmatites into different types (only rock occurrences).*

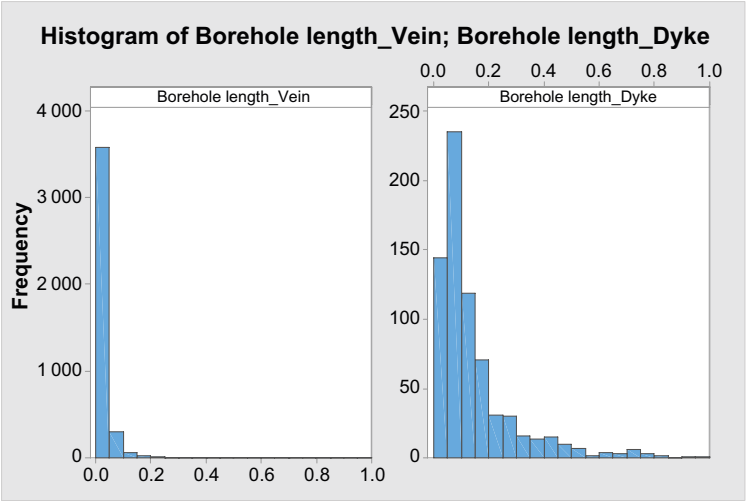


Figure 3-5. *Frequency plot on borehole length of pegmatite veins and dykes.*

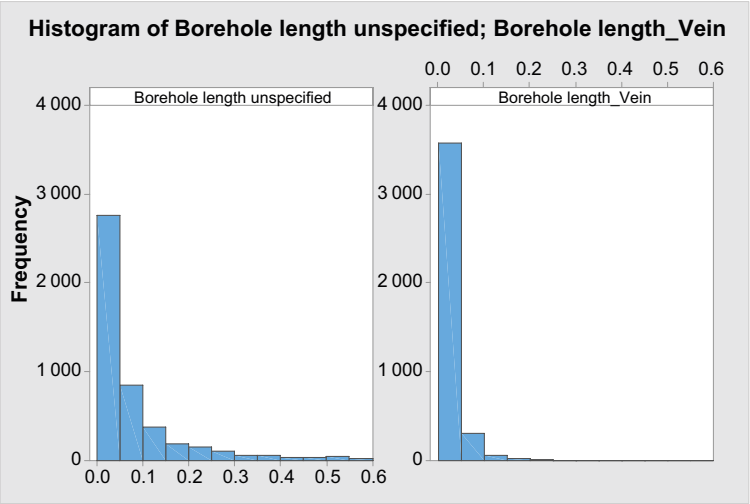


Figure 3-6. *Frequency plot with smaller increments on borehole length of pegmatite veins.*

Orientation of pegmatite dykes

Orientation of pegmatites are based on data from *Rock type* (p-rock) and *Rock occurrence* from 12 boreholes: KFM01A, 01C, 01D, 04A, 05A, 06A, 07A, 07B, 08A, 08C, 09B, and 10A (almost all used in Forsmark Thermal site descriptive modelling 2.2). For strikes > 180º, measured strike minus 180º has been used.

Table 3-3 shows a summary of the directions of the tectonic foliation in 15 boreholes based on stereographic projections in Stephens et al. (2007). The foliation in the margins is more pronounced but is also influenced by the direction of the margin boundary. The general direction of foliation is NNW–SSE (about 160°) and is steeply dipping. Boreholes with anomalous foliation orientations occur near the fold axis of the major synform. Figure 3-7 shows a rose diagram of the directions of pegmatite from *rock type*. Rose diagrams of the directions of pegmatite from *rock occurrence* are presented for a selection of boreholes in Figure 3-8. Where different orientations for veins and dykes are observed, these types are plotted separately.

Table 3-3. Summary of foliation directions per borehole. Interpretations of stereographic projections in Geology Forsmark 2.2 (Stephens et al. 2007, Appendix 6).

Borehole ID	Foliation – strike	Foliation – Dip
1A	170	subvertical
1C	170	subvertical
1D	160	subvertical
4A	140	subvertical
5A	150	subvertical
6A	100	Ca 60 degrees
6C	100	Ca 60 degrees
7A	165	Vertical
7B	155	subvertical
7C	150	Ca 60 degrees
8A	Variable	
8B	165	subvertical
8C	Bimodal: 60 and 160	subvertical
9B	165	Vertical
10A	135	subvertical

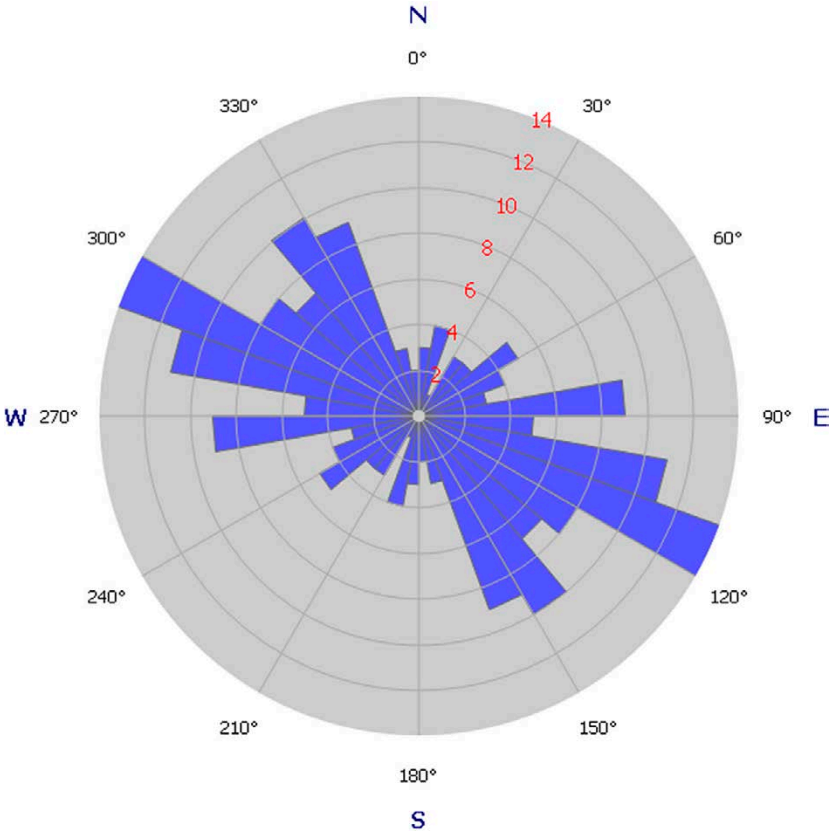


Figure 3-7. *Pegmatites in rock type: p-rock 1A, 1C, 1D, 5A, 7A, 7B, 9B (boreholes in which foliation generally strikes NNW–SSE).*

When comparing Figure 3-8 and Table 3-3 it can be seen that the foliation direction is often concordant to the high frequency of pegmatites from *rock occurrence* (mainly veins). However, there is also a large number of discordant pegmatite *rock occurrences*, particularly those mapped as dykes. The latter is much more accentuated in Figure 3-5 that shows histograms from *rock type*.

The following conclusion can be drawn for pegmatite:

- The veins are to a large extent concordant to the foliation (however not in KFM01A, 1C, and 6A).
- The dykes can also be concordant to the foliation but are mostly discordant.
- That veins are to a generally concordant to the foliation strengthens the approach of creating new thermal rock classes that includes the rock occurrences of pegmatite in the main rock type.
- This also enables separate orientations for larger dykes and smaller veins in the simulations, since the latter is included in the dominant TRC.

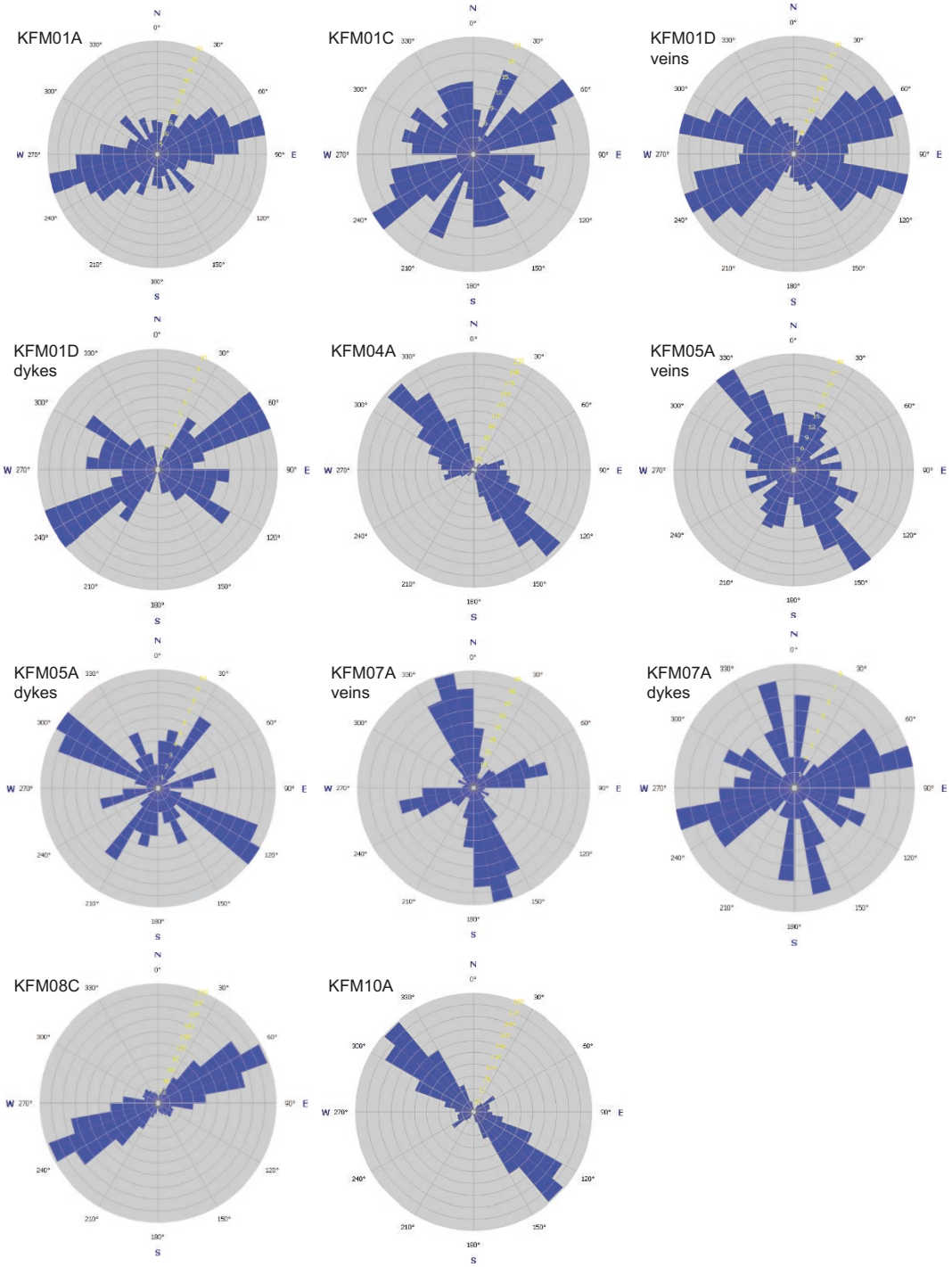


Figure 3-8. *Rose diagram for strike orientations of pegmatites in a representative selection of boreholes. Data from rock occurrence (< 1 m).*

Amphibolite

At Forsmark, amphibolite, which occurs as small, irregular and dyke-like bodies, comprise approximately 5 % of the rock mass. Having little or no quartz, the amphibolites have thermal conductivities considerably lower than the dominant granitoid host rocks. An assessment of the true thickness of amphibolites is reported by Stephens et al. (2007). The amphibolites are predominantly thin geological entities with numerically relatively few occurrences with a true thickness greater than 0.9 m; see Figure 3-9.

A somewhat different picture emerges when the volume of each true thickness class is plotted; Figure 3-10. This plot shows that mafic rocks with a true thickness of 1 m or more make up approximately 1 % of the total rock volume based on the borehole data. Analysis of individual rock domains has not been performed. However, an anomalously large amphibolite body, present along a borehole length of about 40 m in borehole KFM06C (745–785 m), occurs in domain 45. This occurrence is equivalent to approximately 15 % of all amphibolite with borehole lengths longer than 1 m in 16 boreholes (citation from Back et al. 2007).

Table 3-4 shows percentage amphibolite in each length class in relation to (a) total borehole length and (b) amphibolite occurrences only.

Table 3-4. Borehole length distribution of amphibolite (102017) based on data from 12 boreholes (almost all used in Forsmark SDM 2.2). No distinction made between rock domains.

Borehole length interval	(a) Proportion of total borehole length comprising amphibolite	(b) Proportion of amphibolite in different length classes
< 5 cm	0.18 %	3.6 %
5–10 cm	0.24 %	4.9 %
10–20 cm	0.44 %	8.7 %
20–50 cm	0.87 %	17.4 %
50–100 cm	0.96 %	19.1 %
Sum < 1 m	2.69 %	53.7 %
> 1 m (based on assumption of 5 % total; Stephens et al. 2007)	2.3 %	46 %

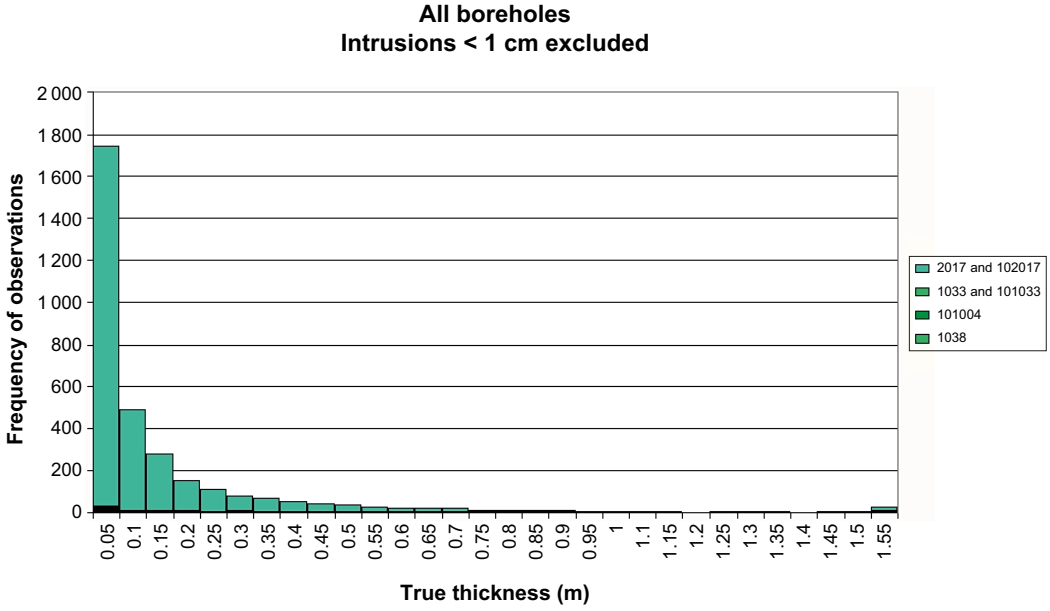


Figure 3-9. *Thickness distribution of mafic rocks (mainly amphibolite – code 102017) based on data from all cored boreholes excluding KFM07C, KFM08C, KFM10A (Stephens et al. 2007).*

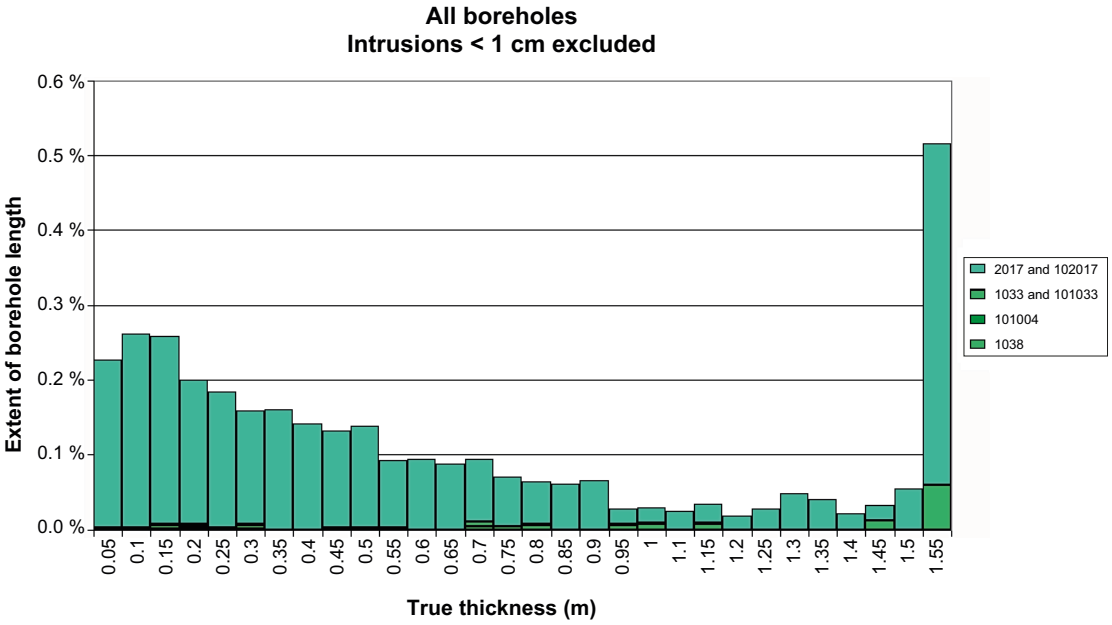


Figure 3-10. *Percentage of borehole length in each thickness class for mafic rocks (mainly amphibolite – code 102017) based on data from all cored boreholes excluding KFM07C, KFM08C, KFM10A (Stephens et al. 2007). The y axis can also be considered as a measure of volume percent.*

The following conclusions can be drawn for amphibolite:

- A substantially large proportion of the amphibolites are thin with true thickness less than 0.2 m.
- The results support the approach of including rock occurrence of amphibolites in the actual rock type and forming new TRCs.

Granodiorite – tonalite

Granodiorite to tonalite (101051), is not a single rock type. Compositionally, rock type with code 101051 varies from tonalite and granodiorite to more subordinate granite (Stephens et al. 2007). Calculations of thermal conductivity from density loggings have shown that individual bodies of this rock type are rather homogenous with respect to thermal conductivity, but that there are quite large differences between these bodies. During the site descriptive modelling it was therefore decided to subdivide granodiorite to tonalite (101051). The thermal conductivity values were found to fall into three distinct groups: high (> 3.17 W/(m·K)), medium (2.8–3.18 W/(m·K)), and low (< 2.8 W/(m·K)) (Back et al. 2007).

The following preliminary conclusions can be drawn for granite to tonalite (101051):

- Density loggings can be used to distinguish between the different subtypes of 101051
- Each body of 101051 seems to be rather homogeneous from density interpretations and supported by information from site geologist (Petersson, 2020, personal communication).

3.6 Thermal implications of new TRCs

3.6.1 Modelling approaches

A new definition of TRCs has implication for the spatial thermal models that are required for performing stochastic simulations of thermal conductivity. These models are of two types:

1. Thermal distribution models.
2. Thermal variogram models.

Both types of models correspond to the 1 m scale, identical to the scale suggested for the new TRC definition.

Three different approaches can be used to determine distribution models and variogram models for TRCs, as described below. These approaches are summarized in Table 3-5.

Table 3-5. Approaches for developing spatial statistical thermal models, based on available data.

	Laboratory data	Field measurements	Density log data
Thermal distribution model of TRC	Laboratory data approach (RBA)	Field measurement approach	Density logging approach[1]
Thermal variogram model of TRC	Laboratory data approach[1]	Field measurement approach	Density logging approach

[1] Approach associated with substantial uncertainty.

3.6.2 The laboratory data approach

Distribution modelling

This approach can be used to determine distribution models for TRCs in situations when only small-scale laboratory data of thermal conductivity are available. The basis for the approach is the Boremap lithological data, both rock type and rock occurrence (rock type is a section of borehole core with a length exceeding 1 m and rock occurrence is a part of a borehole core with a length less than 1 m). This information is then combined with laboratory data of thermal conductivity, randomly assigned to boreholes. Therefore, the approach could also be referred to as the Randomized Boremap Approach (RBA), to indicate that it utilizes randomly drawn data from Boremap. The approach is described below for a situation when the simulation scale is 1 meter, but slightly smaller or larger scales can also be used if necessary. The basis for the approach is Sections 5.4.3 and 5.4.4 in Part 1 (Back and Sundberg 2022), i.e. *preparing lithological data* and *classifying lithological data as TRCs.* The procedure is as follows:

1. A set of borehole cores is selected. These cores should be representative of the rock volume that is modelled.
2. Each core is divided into parts of length 0.1-meter. Each part is assigned a rock code from Boremap. Both *rock type* and *rock occurrence* are considered.
3. Each core is divided into 1 meter long sections, i.e. 10 parts of 0.1 meter. Each meter section is assigned a TRC. The TRC is identical with the rock code from Boremap of the dominating rock type. If there are more than one dominating rock type within a section, the one with the largest share in the meter section is used for the classification.
4. A set of thermal laboratory data from rock samples is compiled for each rock code. Descriptive statistics is calculated for each data set and normal distributions determined. It is assumed that thermal conductivity for each rock code is normally distributed, which is a reasonable assumption.
5. Each 0.1-meter part in a core is assigned a randomly selected thermal conductivity value from the normal distribution corresponding to the particular rock code. Thus, almost[3] the whole core is assigned attributed thermal conductivity values conditioned on the rock code.

[3] Except for a few rock codes of minor importance where laboratory data is lacking.

6. The thermal conductivity for each 1 m-section is calculated from the thermal conductivity values in the corresponding ten 0.1 m-parts. The self-consistent approximation (SCA) for an n-phase material (Sundberg 1988) can be used for the calculation, or the geometric mean as an approximation.
7. All thermal conductivity values at the 1 m scale are sorted by TRC. Then, a distribution of thermal conductivity can be presented for each TRC. A model of this distribution can be used in the subsequent stochastic simulation.

A disadvantage of this procedure is that no spatial autocorrelation is accounted for within each 1 meter section. This results in a too large variance reduction at the 1 meter scale and an underestimation of the variance at that scale. One way to combat this is to consider correlation in step 5 above. The approach is relatively simple: First, a random value of thermal conductivity is selected for the top 0.1 m-part in a meter-section. Secondly, the value for the next 0.1 m-part is randomly selected but with consideration of the autocorrelation over a distance of 0.1 meter.

The value λ_1 for the top 0.1 m-part is calculated as:

$$\lambda_1 = \mu_i + \sigma_i \cdot Z_1 \tag{3-1}$$

where μ_i is the mean thermal conductivity for rock code i, σ_i is the corresponding standard deviation and Z_1 is the standard normal random variable for the top part (randomly selected value from the distribution derived in step 4). The value λ_2 for the second 0.1 m-part is calculated as:

$$\lambda_2 = \mu_i + \sigma_i \cdot [Z_1 \cdot \rho + Z_2 \cdot (1 - \rho^2)^{0.5}] \tag{3-2}$$

where Z_2 is the standard normal random variable for the second 0.1 m-part and ρ is the correlation coefficient for autocorrelation over distance 0.1 m. This calculation is repeated for all 0.1 meter parts within a meter section (Z_1 for the next part will be equal to the expression within brackets with Z_2 having a random value from the distribution).

If a standardized variogram is available, the correlation coefficient can be estimated as:

$$\rho = 1 - nugget \tag{3-3}$$

Correlation is ignored by setting ρ = 0. On the contrary, setting ρ =1 results in all 0.1 m-parts having the same value within a meter-section, i.e. maximum correlation.

Major advantages of the laboratory data approach are that:

- Real lithological data is used.
- The new TRC definitions correspond to the dominating rock type in Boremap, which makes the approach easy to apply.
- The approach is in fact a type of upscaling to the simulation scale, but no stochastic simulation is required (except the simple random selection of thermal conductivity for each 0.1 m-part).

Some disadvantages are:

- Large laboratory data sets are required in order to derive reliable distributions of thermal conductivity.
- It is assumed that laboratory data represents the 0.1 meter scale. This assumption results in an overestimation of the variance.
- Correlation within meter-sections can be accounted for but the correlation coefficient is uncertain.

A test of the approach is presented in Section 3.7.

Variogram modelling

In the laboratory data approach, the variogram is calculated from data acquired from laboratory tests. Prerequisites for this approach is:

- Data must have 3D coordinates.
- The data set must be sufficiently large.
- Enough data must come from locations within the correlation distance (the range).

These criteria must be met for all dominating rock types corresponding to each TRC, only in this case can variograms be calculated. However, there are two additional problems that need to be overcome:

- The variograms will represent dominant rock types, not necessarily TRCs that include rock occurrences.
- The variogram will represent the scale of the laboratory data, not the simulation scale. Hence, upscaling of the variogram is required.

The first problem is probably small but will introduce some uncertainty. The second problem can be solved by upscaling techniques that were applied in the site investigations. However, due to all uncertainties, the laboratory data approach is not recommended for developing variogram models.

Conclusions

The laboratory data approach can be used for deriving distribution models of thermal conductivity for TRCs. There are some uncertainties, such as the effect of autocorrelation over short distances, but they could be overcome. However, the approach is not the first-hand choice, mainly because several assumptions that introduce uncertainty must be made.

For variogram modelling, the laboratory data approach is also uncertain. Large data sets are required and even with such the scale of data deviate from the desired scale. The conclusion is that the laboratory data approach is not the first-hand choice for variogram modelling either.

3.6.3 The field measurement approach

Distribution modelling

The field measurement approach can be applied when continuous data from large-scale field measurements of thermal conductivity are available at a flexible scale or a scale equivalent to the simulation scale, i.e. approximately 1 meter. The use of the DTS-H method has been suggested (Sundberg 2019). The procedure to determine the thermal distribution for a TRC is straight-forward:

1. A set of boreholes, or borehole sections, is selected. These boreholes should be representative of the rock volume that is modelled.
2. The rock code (with corresponding TRC) for each measurement of thermal conductivity is determined. This is made from information in Boremap.
3. All data are sorted by TRC. Then, a distribution of thermal conductivity can be presented for each TRC. A model of this distribution can be used in the subsequent stochastic simulation.

In the described approach, it is assumed that all data represent the same scale, i.e. have the same support. If this is not the case, data should be grouped according to scale and each scale be handled as a separated population.

The advantages of this approach are that the principle is simple and that data will be available at the simulation scale, so no upscaling is needed. In addition, the measurements will represent the corresponding TRC and no assumptions are required.

Potential disadvantages of the field measurement approach are all related to the data collection. Generally, there are no disadvantages, if data of sufficient quality can be acquired at reasonable cost.

Variogram modelling

The field measurement approach for developing variogram models is straight-forward. The data prerequisites are the same as for the laboratory data approach. In addition, the measurement scale should be equal to the simulation scale, to eliminate the need for upscaling.

The procedure is as follows: First, field measurements of thermal conductivity are sorted by TRC, i.e. based on dominating rock type. Then, variograms are calculated for each TRC.

There are several advantages of this approach. One important advantage is that the variogram will be based on true measurements of thermal conductivity for each TRC, including rock occurrences. No additional assumptions are needed. In addition. the variogram will represent the simulation scale, i.e. no upscaling is required. Therefore, uncertainty is expected to be significantly lower for this approach compared to other approaches.

Conclusions

The field measurement approach has significant advantages for both distribution modelling and variogram modelling:

- Data can be provided for the appropriate scale, without upscaling.
- Data will automatically consider rock occurrences in a TRC.
- Data may be continuous in boreholes.
- No major assumptions are required.

These important advantages make the field measurement approach the preferred option, both for developing thermal distribution models and variogram models. A prerequisite is of course is that data of sufficient quality and quantity will be available.

3.6.4 The density log approach

Distribution modelling

This approach can be used when density logging data is available and a reliable relationship between thermal conductivity and density logging data can be determined. A theoretical relationship between thermal conductivity and density can be expected from a mineralogical point of view; see Sundberg et al. (2009). If this is confirmed for specific rock types, thermal conductivity can be estimated from density log data, as in Back et al. (2007). The approach is described below for a case where the simulation scale is 1 meter, with a resolution of 0.1 m. A different scale and a different resolution could of course be used if necessary. The procedure is as follows:

1. A set of boreholes, or borehole sections, with density log data is selected. These boreholes should be representative of the rock volume that is modelled.
2. Each borehole is divided into parts of length 0.1-meter (other resolution is possible). Each part is assigned a rock code from Boremap. Both *rock type* and *rock occurrence* are considered.
3. The thermal conductivity is calculated for each 0.1 m-part based on established relationships between thermal conductivity and density for the actual rock code.
4. Each borehole is divided into sections of 1 meter. Each section is assigned a TRC, corresponding to the rock type from Boremap.
5. The thermal conductivity for each 1 m-section is calculated from the thermal conductivity values in the corresponding ten 0.1 m-parts. The self-consistent approximation (SCA) (Sundberg 1988) can be used for the calculation, or the geometric mean as an approximation.
6. All thermal conductivity values at the 1 m scale are sorted on TRC. Then, a distribution of thermal conductivity can be presented for each TRC. A model of this distribution can be used in the subsequent stochastic simulation.

The advantage of this approach is that it results in a large data set of thermal conductivity for each TRC. However, there are several obstacles in the procedure. Most importantly, it is not known whether or not reliable relationships between thermal conductivity and density can be established for all important rock types. If this is possible to achieve will depend on the mineralogical composition of each rock type but also on the quality of the density logging data. In addition, filtering of density logging data is probably required. Furthermore, it is expected that there will be substantial uncertainty in the calculated thermal conductivity values. The uncertainty will be present in the thermal distribution models and consequently propagated into the thermal simulation where the models are used. Due to the uncertainties, this approach is not recommended.

Variogram modelling

The density log approach for developing variogram models depends on density log data. It does not necessarily rely on a determined relationship between thermal conductivity. However, a basic assumption for the approach is that autocorrelation in density is similar to autocorrelation in thermal conductivity. Only if this assumption is true will calculated variograms be reliable.

The procedure is the same as in Section 3.6.3 when thermal distributions are developed, except for step 6 where a variogram is calculated instead of a distribution.

The main advantage of this approach is that large amounts of spatial data can be used in the calculations, which make variograms reliable. However, it must be stressed that this assumes that the spatial statistical structure of thermal conductivity is the same as for density.

The drawback of the approach is of course that the basic assumption about similar spatial structure may not be very good for all TRCs. This will introduce uncertainty. In addition, the success of the approach will depend on the quality of density logging data, and if suitable filtering of data can be made.

Conclusions

The density log approach is not recommended for modelling of thermal TRC-distributions due to all uncertainties. However, the density log approach may be more useful for variogram modelling. The main advantage of the approach is that large, continuous data sets may be compiled, which is ideal for variogram modelling. The success of the approach will depend on the quality of density logging data and how well the spatial structure of thermal conductivity is reflected by density. All in all, it is expected that the approach of developing variogram models from density logs could be useful, but some uncertainty is inevitable.

One possibility is the use density log data in combination with some other approach for variogram modelling. One method may not give reliable variograms but by combining two approaches more reliable variogram models may be derived.

3.7 Test of new TRCs

3.7.1 Basis for the test

Previous sections have illustrated the implications of the new TRC definitions for both the lithologic models and the thermal models used in the simulations. It was concluded that the new TRC definition facilitates the lithological modelling in that Boremap data can be used straight forward. Regarding the thermal models, there are several alternative approaches that can be used depending on available data; see Table 3-5. It was considered necessary to test one of these approaches in order to study the effect of the new TRC definition on the thermal conductivity distributions. Because of the availability of data, the laboratory data approach (Randomized Boremap Approach) was selected for the test. This approach is based on laboratory measurements of thermal conductivity.

Moreover, including rock occurrence in the new TRCs may cause less lithologic anisotropy in the model (and thus less thermal anisotropy). The need for correction of thermal properties in specific directions needs to be investigated. A simple way to do this is to investigate possible anisotropy in the test results. This has been done in Section 3.7.4.

3.7.2 Implementation of the test

The laboratory data approach was tested on Forsmark data. The test was performed according to the procedure of seven steps described in Section 3.6.2. In addition, a test of considering autocorrelation within each meter-section was also performed. Furthermore, the lengths of rock occurrences in TRC-57 was investigated. The goal was to get an idea of the effect on the thermal distribution from such occurrences.

Step 1

Four different boreholes were used in the test: KFM01A, KFM01B, KMF02A, and KFM03A, all representing rock domain 29. The reason for the choice of boreholes was practical; data preparation had previously been performed for these boreholes in the site investigation phase, which facilitated the test.

The Thermal Rock Classes in the test are presented in Table 3-6.

Table 3-6. The different Thermal Rock Classes in the test.

Thermal Rock Class	Dominating rock type	Rock occurrences
TRC-57	101057	All
TRC-51C	101051[1]	All, excluding 103076[2]
TRC-61	101061	All
TRC-17	102017	All

[1] The rock was assumed to be tonalitic, i.e. of type C according to Back et al. (2007).

[2] No such rock (103076) was present in the test.

Step 2–3

All borehole cores were divided into 0.1 m-parts and 1 m-sections. Each 0.1 m-part was assigned a rock code based on information stored SICADA (from core mapping using Boremap system). In addition, each 1 m-section was assigned a TRC based on the dominating rock type for that section.

Step 4

Thermal conductivity at the 0.1 m scale was assumed to follow a distribution derived from laboratory data. All distributions were assumed to be normal. They were all based on the distributions used by Back et al. (2007).

Step 5–6

Each 0.1 m-part was assigned a random value of thermal conductivity from the distribution corresponding to the rock at the specific section. Then, thermal conductivity was calculated for each 1 m-section, as the geometric mean of ten 0.1 m-parts. In addition, a second test was made where the value for the upper 0.1 m-part was chosen to represent the whole meter-section. This gives an upper bound of the variability at the 1 m scale.

Step 7

Distributions of thermal conductivity values and descriptive statistics were compiled for each TRC, both for the geometric mean (denoted “geomean”) and for the top 0.1 m values (denoted “scale 0.1 m”).

Autocorrelation

Because the approach above does not consider autocorrelation, an attempt was made to illustrate how this may influence the results. This test was limited to borehole KFM01A. The procedure followed the description in Section 3.6.2. Because the correlation coefficient at the 0.1 m scale is not fully known, a range of different coefficients were used, from 0 to 1.

Rock occurrences

In order to study the effect of rock occurrences, all such in TRC-57 were identified for the four boreholes. Lengths were calculated and statistics compiled.

3.7.3 Results

The derived thermal conductivity distributions are presented as histograms in Figure 3-11 to Figure 3-14 (TRC2 in the figures refers to the new TRC definition). Descriptive statistics are presented in Table 3-7.

The geometric mean calculations result in distributions with lower variance compared to the 0.1 meter scale. This is expected because the variance is reduced when the scale is increased. The distributions based on the 0.1 meter scale have unrealistic large variance, for two reasons:

- The distributions are based on small-scale laboratory data, which represents a cm-scale but is applied at the dm-scale.
- There is no variance reduction from dm to meter.

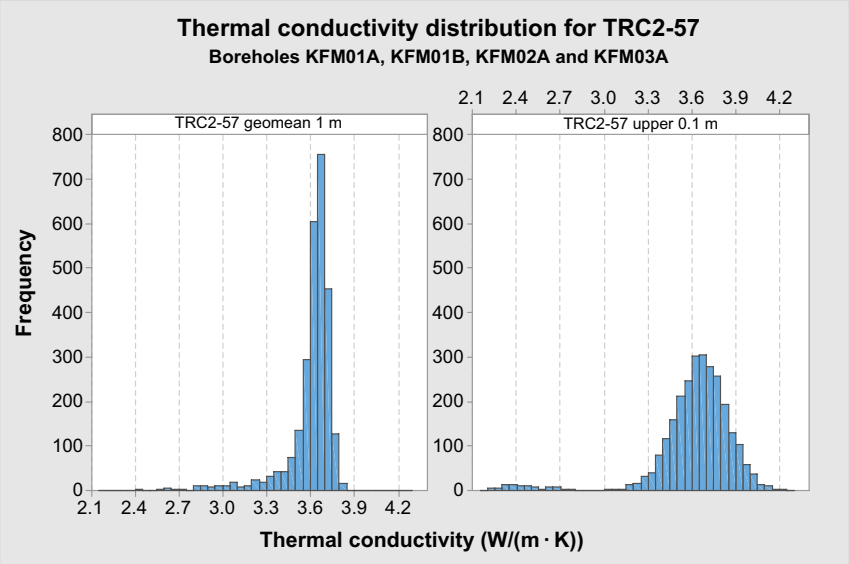


Figure 3-11. *Resulting thermal conductivity distributions for the new TRC-57.*

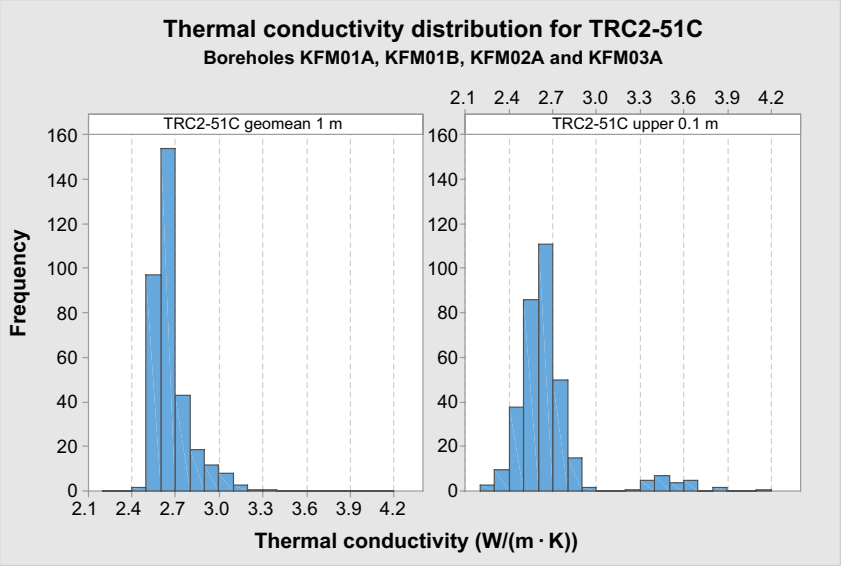


Figure 3-12. *Resulting thermal conductivity distributions for the new TRC-51C.*

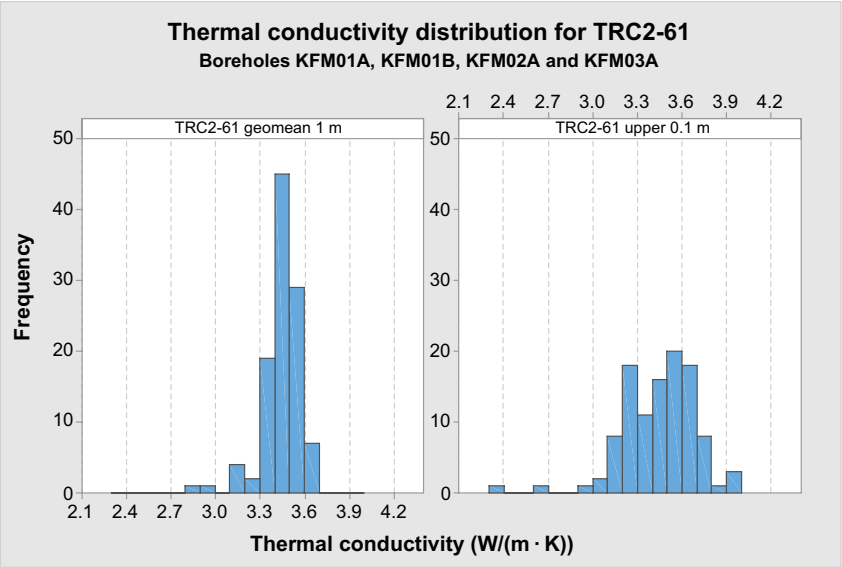


Figure 3-13. *Resulting thermal conductivity distributions for the new TRC-61.*

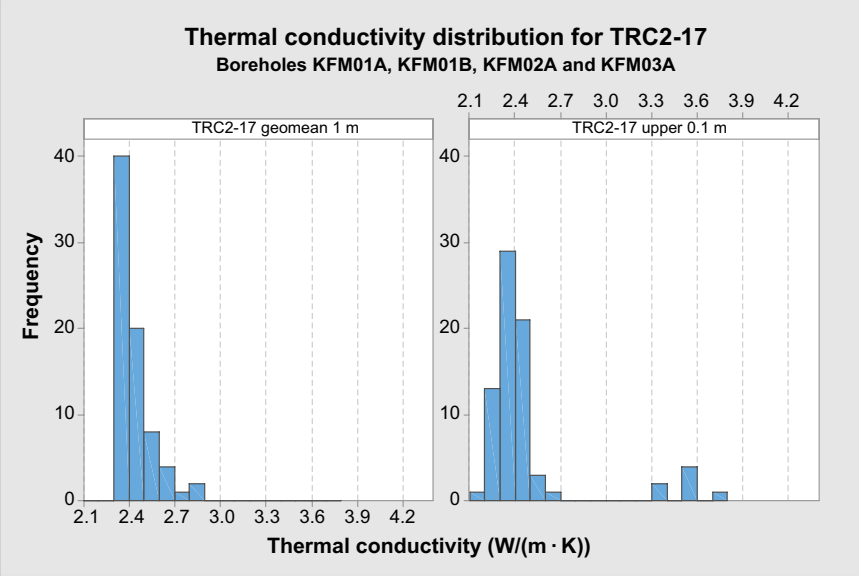


Figure 3-14. *Resulting thermal conductivity distributions for the new TRC-17.*

Table 3-7. Statistics for the derived TRC-distributions.

	Total count	Mean	StDev	Variance	Minimum	Maximum
TRC-57, geomean 1 m	2757	3.600	0.1807	0.0327	2.400	3.881
TRC-51C, geomean 1 m	340	2.673	0.1294	0.0168	2.485	3.323
TRC-61, geomean 1 m	108	3.447	0.1251	0.0156	2.872	3.661
TRC-17, geomean 1 m	75	2.441	0.1118	0.0125	2.317	2.846
TRC-57, scale 0.1 m	2757	3.614	0.2940	0.0864	2.189	4.262
TRC-51C, scale 0.1 m	340	2.680	0.2732	0.0746	2.275	4.173
TRC-61, scale 0.1 m	108	3.445	0.246	0.0603	2.328	3.971
TRC-17, scale 0.1 m	75	2.479	0.3416	0.1167	2.183	3.701

It is more difficult to assess whether the geometric mean distributions underestimates or overestimates the variance because there are two contradicting effects. The first is the same as above: Small-scale laboratory data are assumed to represent the dm-scale, which results in a variance that is too large. On the contrary, by not considering spatial autocorrelation within each meter section, the variance is underestimated (the variance reduction within a meter section will be too large). It is the combined effect of these two effects that determine whether or not the variance is underestimated.

In order to study the effect of autocorrelation, a smaller test was made where this was taken into account, according to the approach presented in Section 3.6.2. The result is illustrated in Figure 3-15. The figure illustrates how the standard deviation increases with increasing correlation. It can be concluded that autocorrelation will have an effect on the TRC-distribution of thermal conductivity. How large the effect is depends on the correlation coefficient, which is not fully known for such small separation distances. A correlation coefficient of 0 corresponds to no correlation at all and 1 represents maximum correlation. A reasonable value is expected in the region of 0.5 to 0.8 but the value will also depend on the laboratory data. Roughly, the standard deviation of the distribution increases 10–15 % when autocorrelation is considered.

The share of rock occurrences within each 1 m-section is illustrated in Figure 3-16 for TRC-57. Almost 60 % of the meter sections did not contain rock occurrences. A vast majority of the meter sections have shares smaller than 0.2, which is equivalent with lengths less than 0.2 m. The statistics in Table 3-8 confirm this; the mean share of rock occurrences is about 12 %. This indicates that the effect on thermal conductivity at the 1 m-scale is expected to be small.

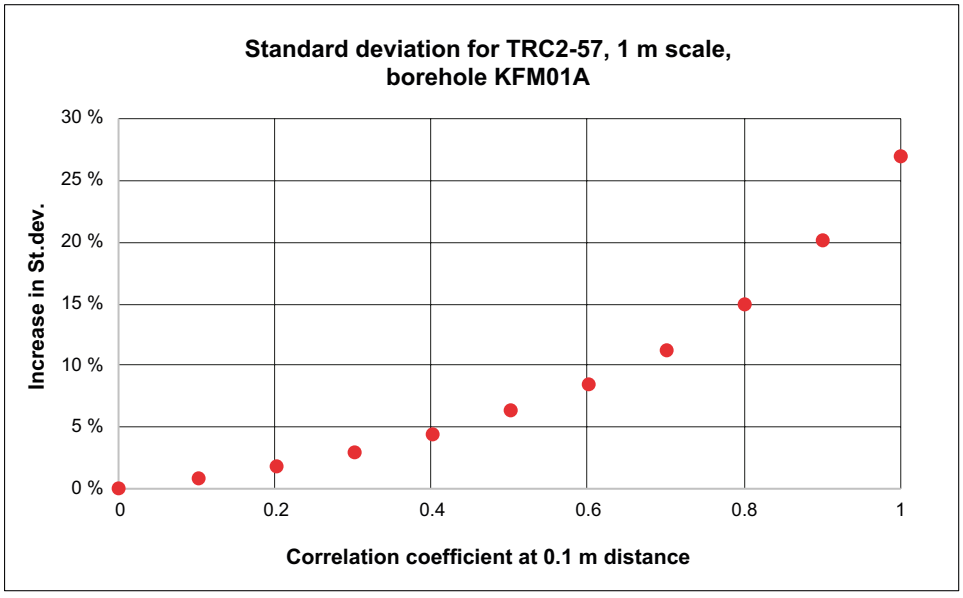


Figure 3-15. *The increase in standard deviation from spatial autocorrelation within each 1 m-section for TRC57. The result is based on borehole KFM01A.*

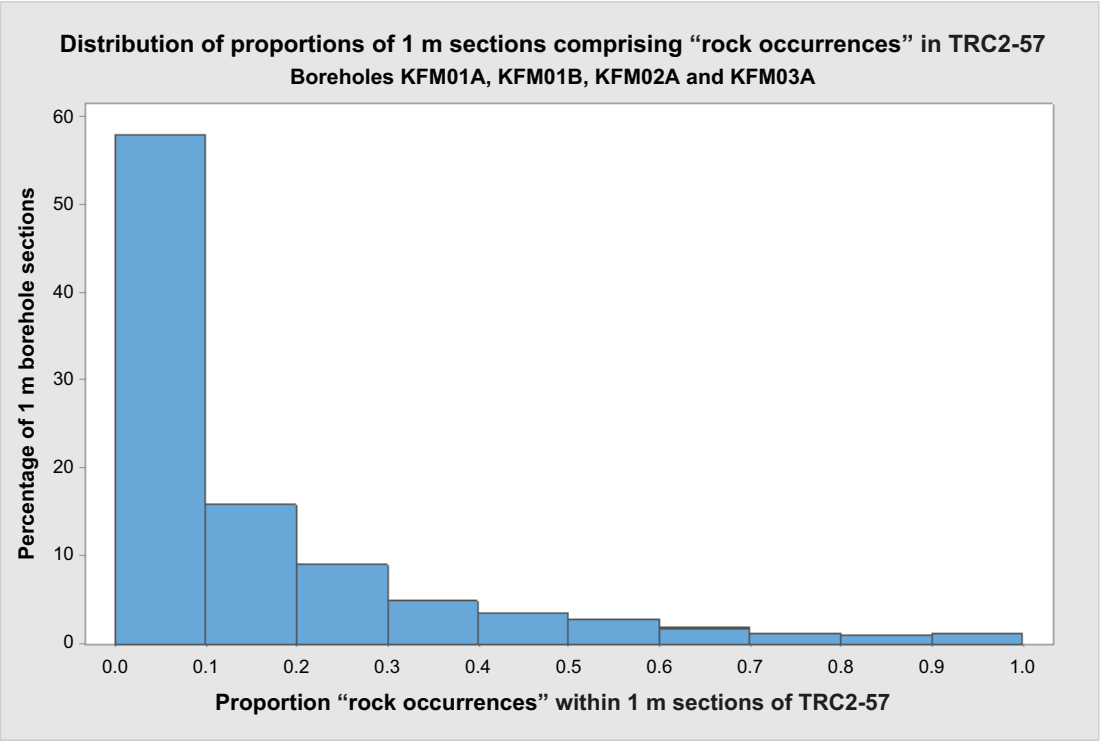


Figure 3-16. *The length distribution of rock occurrences in TRC-57 for boreholes KFM01A, KFM01B, KFM02A, and KFM03A.*

Statistics for each subordinate rock type is summarized in Table 3-8. The shares were calculated for a total of 2 757 meter sections in TRC-57. Pegmatite (101061) is by far the most common rock occurrence, whereas amphibolite (102017) and fine- to medium-grained granite (111058) each have shares of 2–3 %. One thing to be noted is that some meter sections have a share of rock occurrences equal to 1. The reason is that in some of the boreholes rock occurrences have lengths longer than 1 m, which is not in accordance with the definition of a rock occurrence.

Table 3-8. Statistics of the shares of rock occurrences within each meter section in TRC-57.

	Share of rock occurrences	Share of 101051	Share of 101061	Share of 102017	Share of 111058
N	2757	2757	2757	2757	2757
Mean	0.119	0.013	0.048	0.027	0.026
St.dev.	0.2005	0.0777	0.1261	0.1000	0.0955
Minimum	0.000	0.000	0.000	0.000	0.000
Maximum	1.000	1.000	1.000	0.900	1.000

3.7.4 Investigation of possible anisotropy

As mentioned above, introducing rock occurrence in TRCs may cause less thermal anisotropy since the rock occurrence in TRC loses its directional dependence. A simple way to investigate this is to use the test results in Section 3.7.3 (see Table 3-8) and calculate the potential anisotropy.

In an extremely anisotropic material, all subordinate rock types occur as veins or dykes in the same direction. The harmonic and arithmetic mean can then be used to calculate the thermal conductivities in the principal directions in 2D, representing the maximum conductivity parallel and minimum conductivity perpendicular to the dykes/veins respectively (Sundberg 1988). If the difference between these directions is large there is an increased risk to underestimate the anisotropy when introducing the new TRC definition.

Table 3-9 shows the mean shares and thermal conductivities for the different rock types/TRCs. Using the sub TRC51C is a conservative approach, since it has the lowest thermal conductivity of the three TRC51s, and thus the largest influence on increased anisotropy.

Table 3-9. Mean of the shares of rock types within each meter section in TRC-57 and thermal conductivity for each corresponding TRC. The sub TRC51C has been chosen to represent TRC51. Thermal conductivities from Back et al. (2007), representing the TRC definition from SDM.

	Share of 101057	Share of 101051	Share of 101061	Share of 102017	Share of 111058
N	2757	2757	2757	2757	2757
Mean shares	0.886	0.013	0.048	0.027	0.026
Corresponding TRC	TRC57	TRC51C	TRC61	TRC17	TRC58
Thermal conductivity of TRCs	3.68	2.63	3.47	2.38	3.79

Table 3-10 shows the results of the calculations. The calculated 2D extreme mean anisotropy is negligible.

Table 3-10. Calculation of the mean 2D anisotropy in thermal conductivity for the main TRC-57 (new definition) in an extreme configuration of veins and minor dykes.

Mean thermal conductivity based on geometric mean, W/(m·K)	3.614
Maximum thermal conductivity based on arithmetic mean, W/(m·K)	3.624
Minimum thermal conductivity based on harmonic mean, W/(m·K)	3.601
Mean anisotropy ($\lambda_{max}/\lambda_{min}$)	1.006

3.7.5 Conclusions of the test

The thermal conductivity distributions calculated with the laboratory data approach are not normal distributions, but neither are they markedly skewed. This means that there is no major concern for using them as distribution models in the thermal simulations. However, it is advisable to apply transformation prior to simulation and back transformation of the simulation result, as was performed in the site descriptive modelling. Such transformation algorithms are generally available in simulation software.

The conclusion that the laboratory data approach works can be made regardless of the size of an appropriate correlation coefficient. However, as indicated in Section 3.6.2, the approach is not the first-hand choice of developing TRC distribution models because of the uncertainties involved. Much uncertainty could be eliminated if the field measurement approach (Section 3.6.3) is applied instead.

One thing noted during the performed test was that some sections of rock occurrences actually have lengths exceeding 1.0 meter, although such long sections should have been classified as rock type in Boremap. A total of 27 such sections were identified in the boreholes included in the test (only borehole KFM01A and KFM01B contained such sections). This implies that rock occurrences had a

slightly larger and unexpected influence on the calculated TRC-distributions. If such misclassification is avoided in new boreholes, the impact of rock occurrences could be slightly lower than this test indicates. This is a positive conclusion that furthermore supports the conclusion that the approach works for developing thermal TRC distribution models.

The risk of underestimating anisotropy when considering rock occurrence within the main rock type is small. The calculated extreme anisotropy is negligible on average.

3.8 Conclusions regarding new TRC definition

3.8.1 Lithological implications

Pegmatites

- The results of the size distribution of pegmatites are not in conflict with a simulation scale of 1 m and the use of the new TRC definition.
- There is a much larger number of pegmatite veins compared to dykes and they are largely differentiated by their length. A substantially large proportion of the pegmatites is not categorized as veins or dykes, but their short occurrences along the borehole cores indicate that they may have originated mostly from veins or small pegmatites.
- The pegmatite veins are in large extent concordant to the foliation.
- The pegmatite dykes can also be concordant to the foliation but are mostly discordant.
- The orientation of pegmatites strengthens the approach of creating new TRCs that include rock occurrences in the main rock type and enables separate orientations for larger dykes and smaller veins, since the latter is included in the dominant TRC.

Amphibolites

- A substantially large proportion of the amphibolites are thin with true thickness less than 0.2 m.
- The results support the approach of including rock occurrences of amphibolites in the actual rock type to form new TRCs.

Granodiorite-tonalite

- Density loggings can be used to distinguish between the different subtypes of 101051.
- Each body of 101051 seems to be rather homogeneous from density interpretations and supported by information from site geologist (Petersson, 2020, personal communication).

3.8.2 Thermal implications

The main conclusion about the new TRC definition is:

- The performed test of using new definitions of TRCs showed that thermal distribution models can be developed for each TRC. Inclusion of rock occurrences will make the thermal conductivity distributions somewhat skewed, but the effect is small. It is believed that the distribution models can be used in the stochastic simulations without problems.

Other conclusions are:

- The field measurement approach has significant advantages for both distribution modelling and variogram modelling: no upscaling is needed, data will consider rock occurrences in a TRC, and data may be continuous in boreholes.
- These important advantages make the field measurement approach the first of choice, both for developing thermal distribution models and variogram models.
- The laboratory data approach can be used for deriving distribution and variogram models of thermal conductivity for TRCs, however with several uncertainties and shortcomings.

- A method for considering spatial autocorrelation over small distances (< 1 m), when using laboratory data as a basis for thermal distribution models, was illustrated in the test in Section 3.7. This approach could be used in combination with the laboratory data approach if large scale data is not available.
- Using the laboratory data approach in combination with the density log approach will reduce the uncertainty, especially for the variogram modelling and the possibility to divide TRC-51 into sub types.
- The risk of underestimating anisotropy when considering rock occurrence within the main rock type is small. The calculated extreme anisotropy is negligible on average.

3.8.3 Data requirements

- Similar Boremap data is needed for new TRCs as during the site investigations and site descriptive modelling, as a base for the thermal modelling. It would be a large advantage to use Boremap data of rock type to assign TRCs to data at a simulation scale of 1 m. This facilitates the modelling.

Field measurement (DTS-H) has significant advantages for both distribution modelling and variogram modelling and is the first-hand choice of data for conditioned modelling using the new TRC definition. However, the method needs development in order to fulfil SKBs demands (Sundberg 2019).

4 Lithology simulation software

4.1 Previous work

In previous work on modelling of thermal properties in crystalline rock adjacent to canister positions in a repository for spent nuclear fuel, 3D stochastic geostatistical modelling was used to model both the spatial distribution of lithological units and the thermal properties within the rock mass made up by these units (see e.g. Back et al. 2007). The lithological modelling was made using categorical data (lithological unit classification), whereas the modelling of thermal properties was made using continuous data (thermal conductivity). The stochastic lithological modelling was made using the Transition Probability Geostatistical Software (T-PROGS) (Carle 1999).

As a basis for future lithological modelling as input to modelling of thermal properties, this chapter provides:

- An overview of geostatistical modelling of categorical data.
- Comparison of Multiple Point Statistical (MPS) methods to the two-point statistical approach of T-PROGS.
- A description of possible continued use of T-PROGS for lithological modelling.
- Results from testing T-PROGS on different model sizes and configurations.

4.2 Geostatistical modelling of categorical data

There are several different mathematical approaches to geostatistical modelling. One sub-division is (e.g. Journel 2003):

- Object-based approaches.
- Pixel-based approaches.

Kriging is the dominating geostatistical approach, where interpolation across an area or volume is based on geospatial properties (e.g. of thermal conductivity) that can be summarized by a variogram model. Kriging as an interpolation method (e.g. Matheron 1969) has a major drawback in that, due to its inherent smoothing, produces maps that do not properly represent structures and patterns observed in reality. It provides a local most likely (best) estimate at every point across the modelled area or volume, but not a globally correct or realistic pattern of the studied phenomenon. Simulation can then be used to correct for the smoothing effect of kriging to more realistically reproduce properties that can be summarized by a variogram model. An example being the spatial distribution of thermal conductivity within a homogeneous rock facies.

To represent categorical data, indicator variograms can be developed and used for simulation of the spatial distribution of e.g. lithological units. However, variogram-based simulation algorithms have difficulties in representing certain complex geological patterns of spatial variability, such as complex folding structures and other curvilinear features. To overcome this, Boolean object-based algorithms were introduced in the late 1980's to simulate random geometry (Stoyan et al. 1987, Haldorsen and Damsleth 1990). Parametric shapes, such as sinusoidal channels or ellipsoidal lenses, are dropped onto the volume to be simulated, then displaced or removed, their shapes changed, to fit conditioning statistics and local data through an iterative process. This can produce simulated maps that look as expected.

However, object-based methods proved to have practical limitations of different types. One important drawback is that they lack a central feature of the pixel-based approach, where the simulated field, its shapes and patterns, are constructed one point or one grid node at-a-time. This is particularly convenient for conditioning to data of various volume supports and resolutions: it suffices to freeze the hard data at their locations and then build around them.

Object-based approaches have been further developed since the 1980's (see e.g. Chilès and Delfiner 2012), but still suffer from challenges in conditioning. In addition, calculations can be very CPU demanding (see e.g. Hashemi et al. 2014). Most applications on geostatistical modelling of categorical data therefore seem to use pixel-based approaches. They can be categorized into two main types:

- Two-point geostatistical approaches.
- Multi-point geostatistical approaches.

Variogram measures the degree of similarity and dissimilarity of a given variable (i.e. facies) by computing the correlation between two spatial points at a time. Therefore, variogram-based techniques are called two-point statistics (Liu 2006, Zhang 2008) or traditional two-point geostatistics (Deutsch and Journel 1998). As noted earlier, using variogram as a measure of linear continuity, made variogram-based algorithms unable to reproduce geometrically complex geological features (curvilinear geobodies) in simulated realizations. However, two-point geostatistical approaches are widely used for simulations of categorical data in many fields. For categorical data, the sequential indicator simulation (SISIM) approach (see e.g. Deutsch and Journel 1998) is frequently used.

Applying variogram analysis to categorical data involving several categories, e.g. lithological units, is sometimes perceived as complicated and not in line with geological interpretation. To overcome this drawback the use of Markov transition probabilities to describe spatial properties of categorical data was explored by Carle and Fogg (1997), leading to the development of the Transition Probability Geostatistical Software (T-PROGS) (Carle 1999). The idea behind the T-PROGS approach to describe spatial properties is to enable implementation of a transition probability/Markov approach to geostatistical simulation of categorical variables. In comparison to traditional variogram-based geostatistical methods, the transition probability/Markov approach improves consideration of spatial cross-correlations and facilitates the integration of geologic interpretation of facies architecture into the model development process. The T-PROGS code has been implemented in major software packages for geological and hydrogeological simulations, e.g. the Groundwater Modeling System (GMS) by Aquaevo.

In the T-PROGS approach, just like in variogram-based approaches, correlation is calculated between two spatial points at a time and is thus to be considered as a two-point geostatistical method. However, it uses a Markov approach to represent spatial correlation rather than a Gaussian variogram approach. In T-PROGS, the user can use both hard data and subjective interpretations based on geological expert knowledge to develop the spatial correlation structure, with respect to transitions between different categories (e.g. lithological units) and the typical sizes and orientations of these categories. The method then:

1. generates an ''initial configuration'' using a cokriging-based version of the sequential indicator simulation (SIS) algorithm by Deutsch and Journel (1998), and
2. iteratively improves the conditional simulation in terms of matching simulated and modeled transition probabilities by applying the simulated quenching (zero-temperature annealing).

Although the T-PROGS approach provides an approach for modelling spatial correlation more in-line with the thinking of practitioners and improves consideration of spatial cross-correlations, it still has the limitation of two-point statistical methods in reproducing complex curvilinear geobodies.

To overcome the limitations of two-point geostatistical approaches in situations where complex curvilinear geological structures need to be modeled Guardiano and Srivastava (1993) proposed a technique to infer multiple-point conditional distributions directly from a training image. This led to the development of the single normal equation simulation (SNESIM), which has become the most popular multi-point statistical (MPS) algorithm in use to date (Daly and Caers 2010). The training image is in fact a numerical database that depicts all the available geological information in a study area. This data base has to capture all the facies structures including shapes, patterns, and facies distribution (Caers and Zhang 2004, Liu 2006, Zhang 2008) to make up a relevant basis for the simulations. A wide variety of data sources can be used to construct a training image, including core analysis, sequence stratigraphic analysis, outcrops, and modern occurrences.

Although the SNESIM algorithm has been improved since its invention, it still has some limitations. One of these limitations occurs when working with more than four classes of facies that makes the algorithm impractical, especially when dealing with a large study area (Tahmasebi et al. 2012). The most recent pattern-based MPS algorithm is the cross correlation simulation (CCSIM) algorithm (Tahmasebi et al. 2012), which has significantly improved the overall computational time and is much less CPU and memory demanding compared to other MPS algorithms (Tahmasebi et al. 2012). These features are especially improved in a more recent version, which is the multiscale cross correlation (MS-CCSIM) algorithm (Tahmasebi et al. 2014). However, this method is still reported to need some improvement, especially when the simulations are required to be conditioned with hard data (Hashemi et al. 2014).

Considering the available approaches to simulation of categorical data the choice seems to be between the two-point geostatistical approach, represented by methods such as sequential indicator simulation and transition probability approach, and multiple-point geostatistical (MPS) approaches represented by e.g. the multiscale cross-correlation approach.

4.3 Comparison of MPS and T-PROGS approaches

Both two-point and multiple-point geostatistical (MPS) methods have advantages and disadvantages. As described above, one of the bottlenecks in the two-point based geostatistical simulations is their inability in dealing with complex and heterogeneous spatial structures. In particular, these methods cannot convey the connectivity and variability when the considered phenomenon contains definite patterns or structures. For example, models containing regular structures cannot be adequately reproduced using the two-point geostatistical methods. On the other hand, these approaches can produce the conditioning point data exactly, which is not possible with MPS methods.

The approach of using Markov analysis for describing the spatial statistical properties of categorical information has proven useful also to practitioners. The T-PROGS model is implemented in commercial software (GMS) and has proven to be robust. It was also able to properly reproduce rock mass properties in Swedish crystalline rock in previously performed simulations (e.g. Back et al. 2007). It can be set up to handle fairly large models (up to approximately 4 million cells), is able to reproduce conditioning data exactly, and has few or no limitations in handling large amounts of conditioning data. The major drawback is that it is a two-point geostatistical method that is not capable of reproducing complex structural patterns and curvilinear features.

The MPS methods intend to reproduce the physics in natural phenomena and they all are based on a set of training images. While they can reproduce complex curvilinear structures, conditioning in these methods to borehole data and soft data require intensive computation (Tahmasebi 2018). Software for running MPS is available (e.g. the ISATIS MPS developed by Geovariances[4] and the SNESIM developed by Liu 2006), making practical applications possible. Tahmasebi (2018) describes the advantages and limitations of MPS and shows examples on applications to different geological settings. However, while the capability of accurately reproducing complex curvilinear structures, Tahmasebi (2018) outlines some critical challenges in MPS that require more research, such as:

- **Conditioning the model to dense point/secondary datasets** (e.g. well/borehole and seismic data): This issue is one of the main challenges in the pattern-based techniques. In fact, such methods require calculating distance function between the patterns of heterogeneities in the training image (TI) and the model and between the conditioning data that must be honored exactly in conditional simulation.
- **Lack of similarity between the internal patterns in the TI**: Producing a comprehensive and diverse TI is very challenging. Thus, serious discontinuity and unrealistic structures are generated using the current MPS techniques when the TI is not large enough. On the other hand, using very large or multiple TIs are costly for most of such methods.

[4] ISATIS Multiple Point Statistics (MPS) algorithm, www.geovariances.com

- **Deficiency of the current similarity functions for quantitative modelling**: Most of the current distance functions are based on some simple and two-point criteria by which the optimal pattern cannot be identified. Such distance functions are very limited in conveying the information. Thus, more informative similarity functions are required.
- **Better and more realistic validation methods**: There are not so many methods that can be used to evaluate the performance of new developed MPS algorithms. For example, the realizations can show a considerable difference between each other and the TI, while such a variability cannot be quantified using the current methods. Thus, visual comparison is still one of the popular methods for verifying the performance of the MPS methods.
- **Large uncertainties are associated with every TI**: Thus, if several TIs are available, it is necessary to design methods that can determine which TI(s) to use in a given context.

To summarize the comparison between two-point geostatistical approaches (represented by the T-PROGS method) and multiple-point geostatistical methods, it can be concluded that the T-PROGS approach has been successfully applied to crystalline rock conditions representative of the future Swedish repository area. The method has proven to represent the spatial distribution of lithological units reasonably well (Back et al. 2007).

MPS methods have the capability of representing complex curvilinear geological structures in a more realistic way than two-point geostatistical approaches. However, the crystalline rock mass is not in general expected to exhibit curvilinear features, such as folding or cyclic patterns, in the scale relevant for modelling thermal properties in the repository. Given the challenges of MPS outlined by e.g. Tahmasebi (2018) it is recommended that future geostatistical simulation of the spatial distribution of lithological units is performed using T-PROGS.

4.4 Description of T-PROGS

4.4.1 Approach

Stochastic three-dimensional (3D) simulations of lithology are used as input into the thermal modelling. Before the lithology can be simulated, the spatial structure for the different thermal rock units (TRCs) must be modelled. Traditionally, in spatial statistical analysis for geological applications, the following approach is used:

1. Calculate values of a spatial statistic (usually the semivariogram) at regularly spaced lags (separation vectors).
2. Fit a mathematical function (e.g. spherical, exponential) through the variogram measurements.
3. Implement various estimation (e.g. kriging) or simulation (e.g. sequential simulation, simulated annealing) procedures.

Geologic or “subjective” knowledge does not necessarily enter directly into this procedure. Another approach to modelling spatial conditions in geological systems is Markov chain analysis, where the transitional trends between geological materials are analysed. Spatial modelling using Markov chain analysis makes it possible to more directly and explicitly consider factors with geological meanings, such as:

- volumetric proportions of rock categories,
- mean lengths (e.g. mean thickness in the vertical direction),
- juxtapositional tendencies (how one category tends to locate in space relative to another),
- anisotropy directions,
- spatial variations of the above.

Because of the importance of fully acknowledging both hard and soft geological information for the lithological modelling, the Markov chain analysis approach is recommended here. Markov analysis methods have previously been applied in modelling lithological units in Swedish crystalline bedrock (see e.g. Norberg et al. 2002, Rosén and Gustafson 1996).

4.4.2 Transition probabilities and spatial variability

The modelling consists of calculating transition probabilities followed by expert adjustments based on geological interpretations. Carle and Fogg (1997) describe how transition probabilities can be used to model the spatial structure, using Markov chains. An example is given in Figure 4-1.

The probability of a transition from one TRC to another is described in each of the 16 graphs. The example illustrates four different TRCs, i.e. there are 16 different possible transitions. One such transition is from class A to class A, i.e. the TRC is the same when we move one grid cell in the strike direction (left graph in upper row). Another such transition is from class A to class B (second left graph in upper row) and from class B to class A (left graph in second upper row). The probability of a transition from one TRC to another is represented on the y-axis in each graph and the distance between two points is represented on the x-axis.

The transition probability example in Figure 4-1 is for one-dimensional (1D) simulation in the x-direction. For 3D simulations, similar sets of graphs are required also for the y- and z-directions. This results in a total of 3 × 16 = 48 transition probability graphs when four TRCs are present. Five TRCs requires a total of 75 graphs, which is a maximum for this approach in practice. The hard input data for the transition probability modelling will consist of lithological data, mainly from boreholes. The resolution in input data should be the same as the simulation scale.

In practice, software is required for creating the transition probability plots in Figure 4-1. One such software is described below.

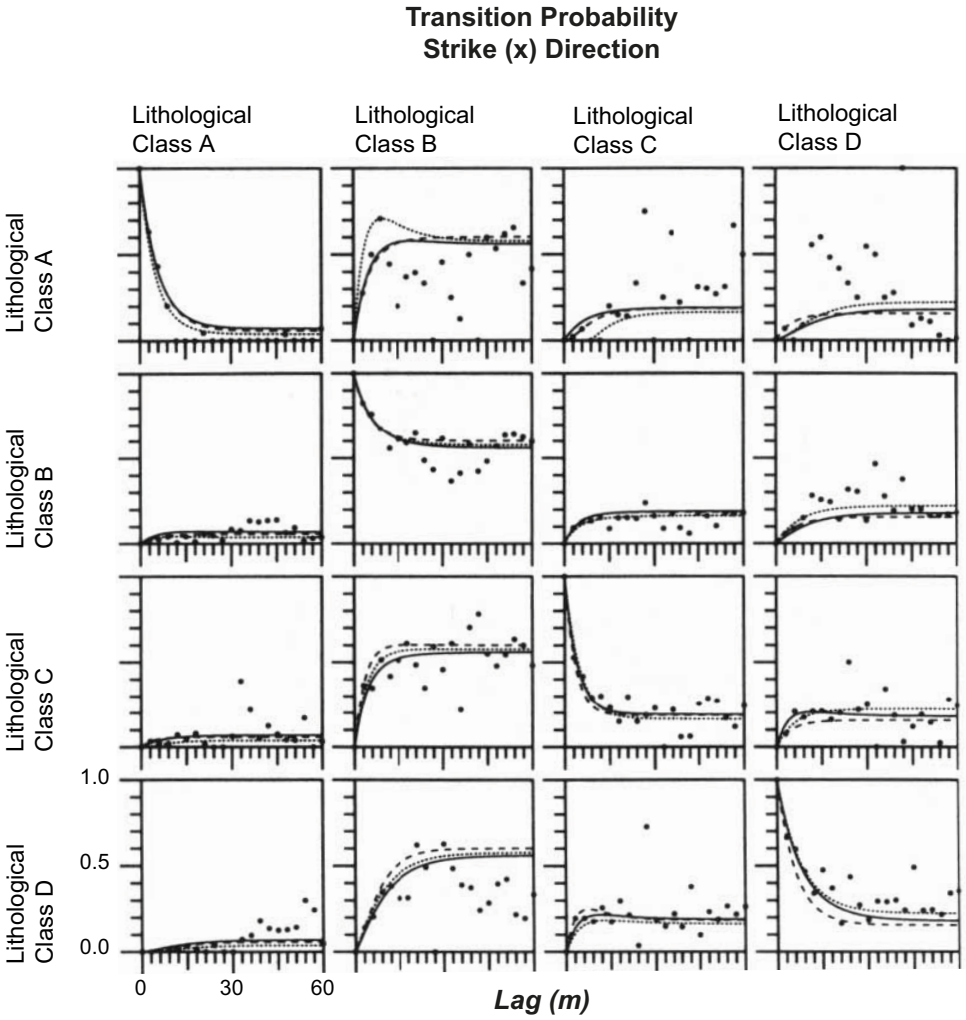


Figure 4-1. *Principle for transition probabilities using Markov chains for 1D simulation (after Carle and Fogg 1997). The graphs present the probability of transition from one thermal rock class to another when moving from one cell to the next in the x-direction.*

4.4.3 Simulation in T-PROGS

The T-PROGS (Transition PRObability GeoStatistics) (Carle 1999, GMS 2019) is used for the transition probability analysis and the stochastic simulations. This software utilizes a transition probability-based geostatistical approach to:

1. model spatial variability of categorical data, e.g. rock classes, by 3D Markov Chains;
2. set up indicator co-kriging equations for predicting rock categories at positions where observations have not been made; and
3. formulate the objective function for simulated annealing for finding the global maximum of the predicted model, i.e. for finding the optimal spatial configuration given the selected input parameters.

T-PROGS facilitates stochastic modelling of lithology based on Markov chain analysis for describing the spatial conditions of lithological categories. The major steps in T-PROGS are:

1. Calculation of transition probability measurements.
2. Modelling spatial variability with Markov chains.
3. Stochastic simulation.

These steps are performed using five different modules; Figure 4-2.

The spatial dependency of material categories is estimated by classical Markov chain analysis. The analysis starts with calculations of the number of transitions between material categories at equally spaced distances (lags). For our case, the material classes are equal to the TRCs. The frequencies are transformed to transition probabilities for the specific lag distance. The probability calculations are displayed in a transition probability matrix, such as:

$$\mathbf{T}\,(\Delta h_z = 1.5\text{ m}) = \begin{bmatrix} t_{11}(\Delta h_z) & \cdots & t_{1K}(\Delta h_z) \\ & \ddots & \vdots \\ t_{K1}(\Delta h_z) & \cdots & t_{KK}(\Delta h_z) \end{bmatrix} = \begin{bmatrix} 0.63 & 0.11 & 0.0 & 0.04 & 0.22 \\ 0.16 & 0.48 & 0.04 & 0.0 & 0.32 \\ 1.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.14 & 0 & 0.14 & 0.57 & 0.14 \\ 0.05 & 0.11 & 0.03 & 0.0 & 0.81 \end{bmatrix}$$

Transition probability matrices are calculated for different lag distances. For each transition category, these "experimental" probabilities are plotted against the lag distance. For each transition category, a curve is fitted to the experimental data to model the spatial dependency. Transition probability matrices are estimated for each principal direction in the modelling volume to facilitate three-dimensional modelling of spatial dependency. Anisotropy is accounted for by using different transition probability matrices for different directions.

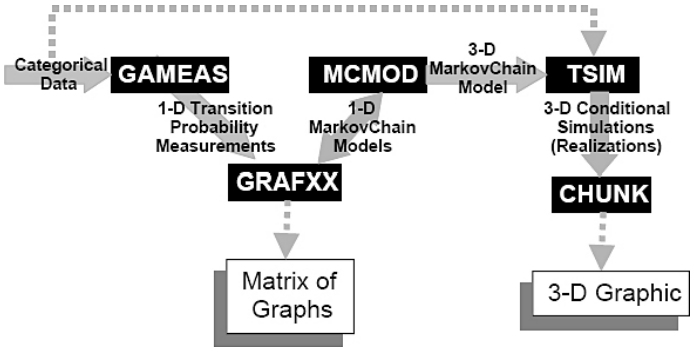


Figure 4-2. *The T-PROGS modelling procedure (GMS 2019).*

The transition probability analysis thus provides the spatial dependency structure between categories (e.g. TRCs) in the system. The proportion and the typical length of each category are derived from the transition probability structure; see Carle (1999).

Most Markov chain analyses in geological applications have been performed in the form of so-called embedded analyses (Carle 1999), in which transition probabilities of occurrences from one discrete occurrence of a category to another, is considered, irrespective of the lag distance. The embedded analysis thus provides the probabilities of what other category to enter when leaving a specific category and does not directly give information about the spatial dependencies of the categories. T-PROGS links the embedded Markov chain analysis to the development of *continuous-lag* (spatially dependent) Markov chain models. The reason this is important is that geologists are more inclined to think and work in the embedded framework (Carle 1999).

4.4.4 Stochastic simulation of TRCs (lithology)

An overview of the T-PROGS modelling procedure for TRCs is given here. For details, the reader is referred to the cited references.

Based on the transitional properties analysed by Markov chain analysis, stochastic simulations (conditional or unconditional) of categorical configurations, such as TRCs, are made through a two-step procedure of:

1. Generating an "initial configuration" using a co-kriging-based version of the sequential indicator simulation (SIS) algorithm (Deutsch and Journel 1998) where a transition probability-based indicator co-kriging estimate is used to approximate local conditional probabilities by:

$$\Pr\{k \text{ occurs at } X_0 \mid i_j(X_\alpha);\ \alpha = 1, ..., N;\ j = 1, ..., K\} \approx \sum_{\alpha=1}^{N}\sum_{j=1}^{K} i_j(X\alpha) w_{jk},\alpha$$

where N is the number of data, K is the number of categories, $w_{jk,a}$ represent a weighting coefficient, and $i_j(x_\alpha)$ represents the value of an indicator variable:

$$i_j(x_\alpha)\{\begin{array}{l} 1\ \textit{if category } j \textit{ occurs at } x_\alpha, \quad j = 1,...,K \\ 0\ \textit{otherwise} \end{array}$$

Use of co-kriging instead of the traditional indicator kriging approach improves consideration of spatial cross-correlations.

2. Iteratively improving the conditional simulation in terms of matching simulated and modelled transition probabilities by applying the simulated quenching (zero-temperature annealing) algorithm:

$$\min\left\{ O = \sum_{l=1}^{M}\sum_{j=1}^{K}\sum_{k=1}^{K}[t_{jk}(h_l)_{MEAS} - t_{jk}(h_l)_{MOD}]^2 \right\}$$

where O denotes an objective function, the h_l denotes $l = 1,\ldots, M$ specified lag vectors, and *MEAS* and *MOD* distinguish measured and simulated (measured from the realization) transition probabilities, respectively (Aarts and Korst 1989, Deutsch and Journel 1998, Deutsch and Cockerham 1994, Carle 1997). The simulated quenching algorithm is implemented by repeatedly cycling through each nodal location of the conditional simulation and inquiring whether a change to another category will reduce O; if so, the change is accepted. This iterative improvement procedure continues until O is minimized, or a limit on the number of iterations is reached. Conditioning is maintained by not allowing changes of categories at conditioning locations. "Artifact discontinuities" (Deutsch and Cockerham 1994) are avoided by generation of the initial configuration and including consideration for anisotropy and limiting the number of lags in formulation of the objective function (Carle 1997).

This procedure is used for each equally probable stochastic realization being performed. T-PROGS can produce up to 999 realizations in one batch of simulation. T-PROGS assumes statistical homogeneity throughout the modelling volume. The proportions of the material categories calculated in the Markov chain analysis is inherently in the method kept stationary in all realizations.

4.5 Borehole configuration, anisotropy and bias

4.5.1 Accounting for anisotropy in lithological modelling

The spatial dependency of the rock mass properties is modelled from borehole data. Hence, the representativeness of the borehole data regarding the spatial dependency of rock mass properties is a crucial factor in the lithological modelling.

Using T-PROGS, the typical lens length and interactions of TRCs are calculated through transition probability analysis of borehole data. Borehole data are often vertical or sub-vertical, but in future modelling work, performed during ongoing construction of the repository, horizontal boreholes will also be available.

In previous modelling work at Äspö and in Forsmark (e.g. Back et al. 2007), vertical or sub-vertical boreholes drilled from the land surface were used for the transition probability analysis, thus giving a better representation of the spatial dependencies in the vertical direction than in horizontal directions.

The spatial distribution of TRCs is known to be anisotropic. This means that subordinate rock types have extensions in a particular direction that are different to another direction. For example, an amphibolite body may have a much longer extension in one direction, perhaps because of its original mode of emplacement as a dyke, or due to subsequent deformation which may have stretched, compressed or flattened rock bodies.

The anisotropy of the TRCs (with stationary spatial statistical properties) cannot be properly estimated based on borehole information from one direction only. Therefore, anisotropy in previous modelling work at Forsmark was considered based on geological interpretations by the site geologists (see Back et al. 2007).

A stepwise procedure was developed for assessing the spatial dependencies and accounting for anisotropy:

1. Initial transition probability analysis of TRCs from borehole data, oriented in the vertical direction. This initial analysis is based on all available borehole data and without any consideration of anisotropy.
2. Transformation of borehole data into the anisotropy orientation of the TRCs. The simulations need to be orientated in the principal directions of anisotropy to properly represent the spatial properties of the rock mass. Existing boreholes are typically not oriented in the anisotropy directions, and a transformation of borehole data to the orientation of the anisotropy of the system is therefore necessary. Geological information describing anisotropy are: (1) the trend and plunge of the mineral stretching orientation, and (2) the strike and dip of the foliation. The orientation of rock units is a function of these.

   A local coordinate system (x’’’, y’’’, z’’’) is developed for the rock mass, governed by the principal directions of anisotropy and an origo defined by minimum easting, minimum northing, and maximum elevation from positions in borehole records. The local coordinate system is obtained through rotations:

   1. To the trend direction of the mineral orientation, i.e. rotation of the x-y plane around the z-axis. This produces the principal axes x’, y’ and z’.
   2. To the plunge of the mineral orientation, i.e. rotation of the x’-z’ plane around the y’-axis. This produces the principal axes x’’, y’’ and z’’.
   3. To the foliation of the rocks, i.e. rotation of the y’’-z’’ plane around the x’’-axis. This produces the principal axes x’’’, y’’’ and z’’’.

   This results in a transformed coordinate system with main axis (x’’’) parallel to the principal direction of anisotropy, see Figure 4-3.

   A detailed mathematical description of the transformation of borehole data to a coordinate system orientated in the principal directions of anisotropy as a function of the mineral stretching and foliation plane is given in Appendix B.

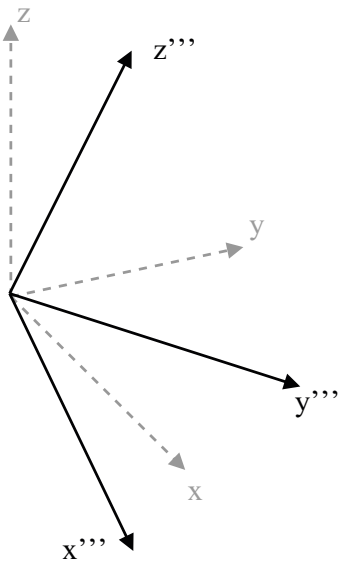


Figure 4-3. *Principal directions x’’’, y’’’ and z’’’ of the local transformed coordinate system, where x’’’ coincides with the principal directions of anisotropy.*

3. Calculation of typical lengths of TRCs for the x’’’, y’’’ or z’’’ directions of anisotropy. These calculations are made on the transformed borehole data. For minimizing errors due to borehole deviations from anisotropy directions only the boreholes with orientation close to one of the axes x’’’, y’’’ or z’’’ are used. The typical lengths in remaining directions are obtained from geometry relationship information for TRCs, given by the geological interpretations of the rock mass.
4. Calculation of transition probability structure with respect to the typical lengths calculated for x’’’, y’’’ and z’’’. By adjusting the typical lengths, the transition probability matrix was updated for each direction x’’’, y’’’ and z’’’. The proportions of TRCs were assumed to be identical to those calculated in Step 1.

The strategy developed thus makes it possible to assign typical lengths of lithological units (and transition probabilities between units) for the three directions in the transformed coordinate system oriented in accordance with anisotropy directions. It should be noted that the developed procedure described above, and the T-PROGS software allows for categories (e.g. TRCs) to have different major anisotropy directions. It should also be noted that, as indicated above, the assignments of the spatial statistical properties (e.g. anisotropy directions, transition probabilities and lens lengths) require statistical homogeneity and that the information used to estimate the properties is not biased.

4.5.2 Bias of data due to orientation of sub-ordinate rock types

In a situation where borehole information is mainly oriented in one direction, e.g. the pre-investigation stage with most boreholes steeply dipping, estimations of typical lengths of TRCs will primarily be based on information from this direction. Typical lengths in other directions will be calculated based on assumptions on anisotropy from geological interpretations of existing information and the transformation procedure described above. As the construction of the repository advances, borehole information also in other directions than the vertical will become available, and thus more information on the anisotropy and typical lens lengths of TRCs in anisotropy directions will become available.

However, as borehole orientations will likely never be totally random, there will always be a potential bias of the information due to the orientation of subordinate rocks relative to the borehole directions. Non-randomized borehole positions and directions may cause bias with respect to both the proportions of specific TRCs as well as to their spatial properties.

He et al. (2014) address the issue of bias due to borehole positions in T-PROGS modelling of sedimentary environments. They show how bias results in erroneous estimations of category (e.g. TRCs) proportions and thereby also typical lengths and transition probabilities. Thus, a correction of the

category proportions is a key to reduce bias in the input data and the outcome model realizations. He et al. (2014) used geophysical data to complement borehole data and to develop bias-adjusted histograms for category proportions.

Barton and Zoback (1990), among many others, address the issue of correcting for sampling bias of fracture apertures. They correct for bias by transformation of the investigated rock volume with respect to the dip angle of fracture zones. This transformation is similar in scope to the approach for transformation with respect to anisotropy directions described in Section 4.5.1 above.

As the number of vertical boreholes increases in an area it would be expected that the uncertainty regarding anisotropy directions should be reduced to a greater extent than the uncertainty regarding the properties of categories (e.g. the lens lengths of TRCs). However, depending on the number of boreholes and the complexity of the rock mass, there will be uncertainty also for anisotropy directions.

To manage the issue of uncertainty regarding anisotropy directions and spatial properties of categories it is suggested that anisotropy and category properties are treated as stochastic variables. The uncertainty of these variables will be reduced by increased borehole information, but as mentioned above uncertainties regarding category properties may be substantial in initial stages of the repository construction.

Borehole information and geological interpretation must be combined to achieve the most information from the combination of hard borehole data and soft expert knowledge. There are different approaches for elicitation of expert knowledge, see e.g. Bedford and Cooke (2001), O'Hagan (2019), and Oakley and O'Hagan (2016).

The Sheffield Elicitation Framework (SHELF), developed at the University of Sheffield, UK (O'Hagan 2019) is an increasingly used approach for performing structured expert elicitation of uncertainty distributions for variables. The objective of the elicitation is to find a probability distribution to represent, as accurately as possible, the knowledge and beliefs of an expert or group of experts regarding a variable of interest. A SHELF elicitation always involves two rounds of judgements:

1. Each expert first makes their own individual judgements.
2. After discussion of these judgements a consensus probability distribution is constructed from group judgements.

Depending on the nature of the quantity or quantities of interest, there is a variety of different judgements that the experts may be asked to make. There is a software developed to assist the use of SHELF and to provide visual feedback to experts during elicitation; see example in Figure 4-4.

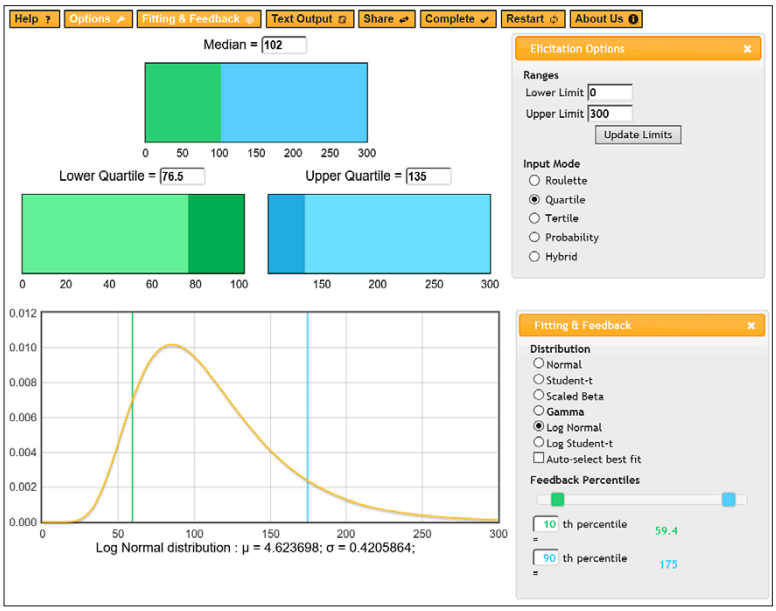


Figure 4-4. *Example of visual feedback using the MATCH software for SHELF expert elicitation. The expert is given feedback regarding the shape of the probability distribution given made assessment of distribution parameters, here min, lower quartile, median, upper quartile, and max.*

The approach has been used in projects of different character and it would likely be useful in the lithological modelling procedure to describe uncertainties in spatial properties of TRCs. It remains to be investigated which variables that are relevant to include in the uncertainty analysis and for which of those to apply expert elicitation.

As mentioned above, randomized borehole directions would be an effective way of reducing bias and uncertainty of spatial property parameters, but this is not realistic in the early stages of a repository construction. Therefore, it is assumed that uncertainty analysis involving expert elicitation will be most valuable in early stages.

Performing uncertainty analysis of e.g. category (TRC) proportions or lens lengths and incorporating these uncertainties in the stochastic T-PROGS-modelling is straightforward by:

1. Constructing probability distributions for spatial parameters e.g. TRC proportions or lens lengths.
2. Random sampling of n values from each constructed uncertainty distribution, giving N sets of values.
3. Using each set of values for N iterations in the stochastic T-PROGS simulation.

4.6 Test of T-PROGS

Previous modelling efforts using T-PROGS for modelling of lithological properties as input to modelling of thermal properties were mainly limited to a model size of 50 × 50 × 50 = 125 000 cells and unconditional simulation (Back et al. 2007). This model size was chosen to provide relevant results while keeping simulation times practical. To investigate the possibilities for increasing the model sizes as well as changing the geometry of the simulation volume, tests have been performed using the following:

- Data for rock domain 45 of Forsmark.
- Four categories Thermal Rock Classes (TRCs):
  - TRC17
  - TRC51
  - TRC58
  - TRC61
- Similar correlation structure as used in 2007, with a relatively high anisotropy. The anisotropies in X-, Y-directions (i.e. horizontal directions) relative to the Z-direction (i.e. the vertical direction) are given by the following factors:
  - X: 3–4
  - Y: 2–3
- Numerical criteria:
  - Number of quenching: 4.
  - Tolerance: 0.00001.

Unconditional simulation was performed for the following model configurations, all using 1 × 1 × 1 meter large grid cells:

I. 50 × 50 × 50 = 125 000 cells

II. 100 × 100 × 100 = 1 000 000 cells

III. 125 × 125 × 125 = 1 953 125 cells

IV. 720 × 160 × 20 = 2 304 000 cells

V. 150 × 150 × 150 = 3 375 000 cells

Models I–III and V represented a cubic simulation volume, whereas model IV represented a cuboid volume, facilitating a geometric representation of the future layout of the repository. Initial attempts to run a cuboid model of size 720 × 160 × 40 = 4 608 000 cells failed due to insufficient computer RAM-memory, see below.

The unconditional simulations were performed according to Table 4-1. The simulations were performed on a conventional laptop computer: HP EliteBook (model year 2017) with 16 GB of RAM using an Intel Core i7-6600U CPU @ 2.60 GHz with a 64-bit operating system (Windows 10). Given the model sizes that were possible to run on this computer, it appears that the maximum model size is about 4 000 000 cells on this type of ordinary PC.

The results from the simulations are shown in graphical form in Figure 4-5.

Table 4-1. Model size, number of cells and realization times in T-PROGS.

Model size (cells)	No of cells in the model	Time (minutes)		
		10 realizations	*50 realizations*	*100 realizations*
50 × 50 × 50	125 000	1	5	10
100 × 100 × 100	1 000 000	8	40	82
125 × 125 × 125	1 953 125	16	78	159
720 × 160 × 20	2 304 000	36	157	304
150 × 150 × 150	3 375 000	30	147	297

As can be seen from Table 4-1 and Figure 4-5 simulation times increase linearly with the number of cells for models I–III and V for all studied numbers of realizations. For the cuboid model IV the simulation time is longer than what would have been predicted for a similarly large cubic model. The reason for this is likely due to increased time for convergence due to the greater asymmetry of the model.

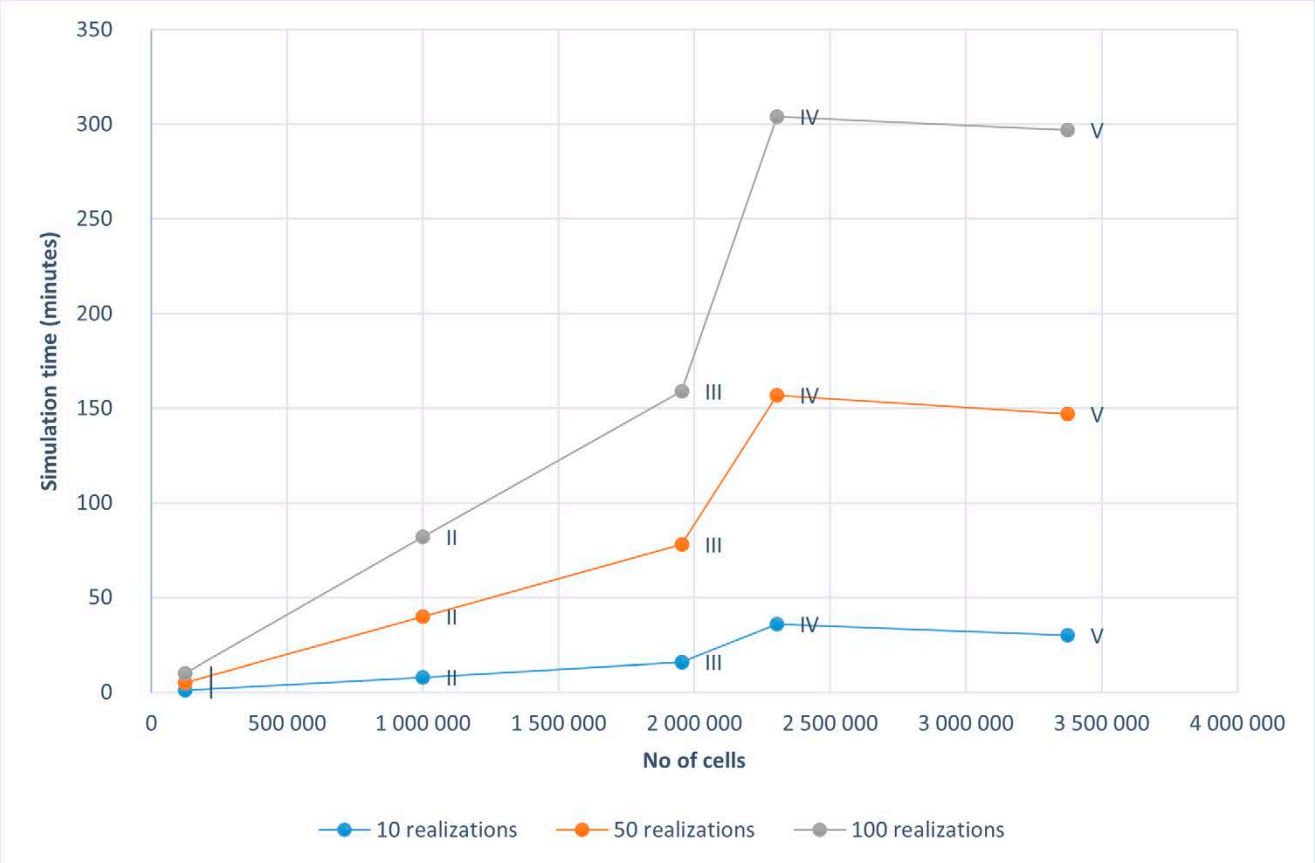


Figure 4-5. *Simulation times for different number of realizations and different model size (number of cells). Models I–III and V have a cubic model geometry whereas model IV has a distinct cuboid geometry.*

Conditional simulation was performed for the 100 × 100 × 100 = 1 000 000 cell cubic model, using one borehole. Simulation times were identical to the simulation times for the unconditional simulations with the same model. This indicates that the simulation times in T-PROGS, at least for conditional simulations including few boreholes, are to a much larger extent governed by the model size and geometry than on the conditioning on borehole data. However, conditioning has not been evaluated for situations with many boreholes located close to each other and simulation times may be affected by complex borehole configurations.

Graphical examples of simulation results are shown in Figure 4-6 to Figure 4-10.

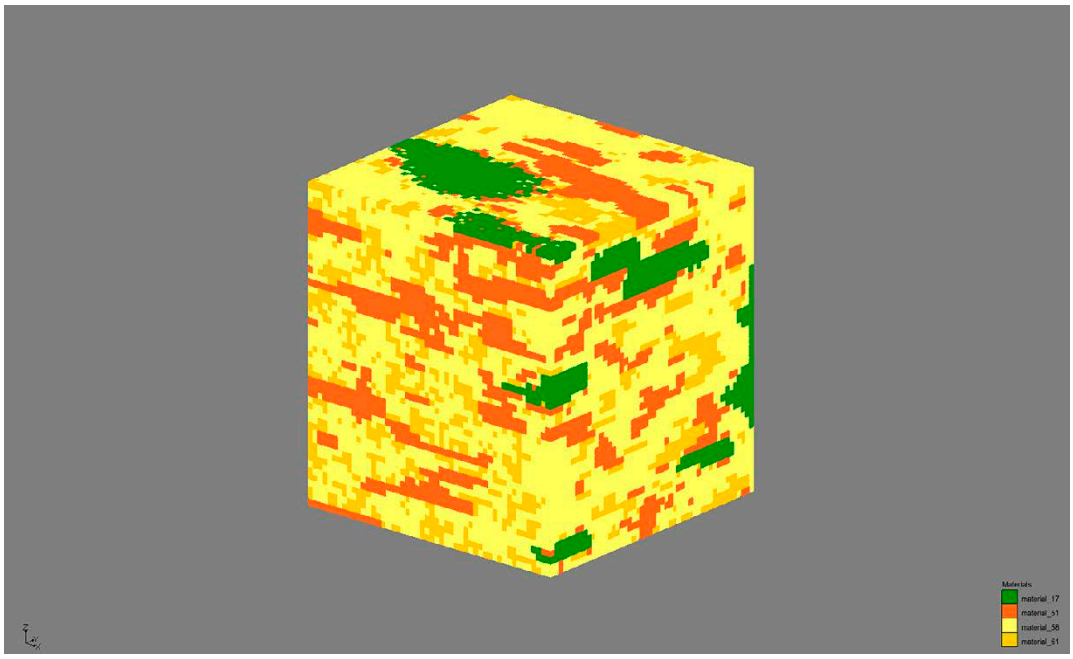


Figure 4-6. *Example of T-PROGS realization for a cubic model geometry of 50 × 50 × 50 meters with a cell size of 1 × 1 × 1 meter. Unconditional simulation with a total model size of 125 000 cells.*

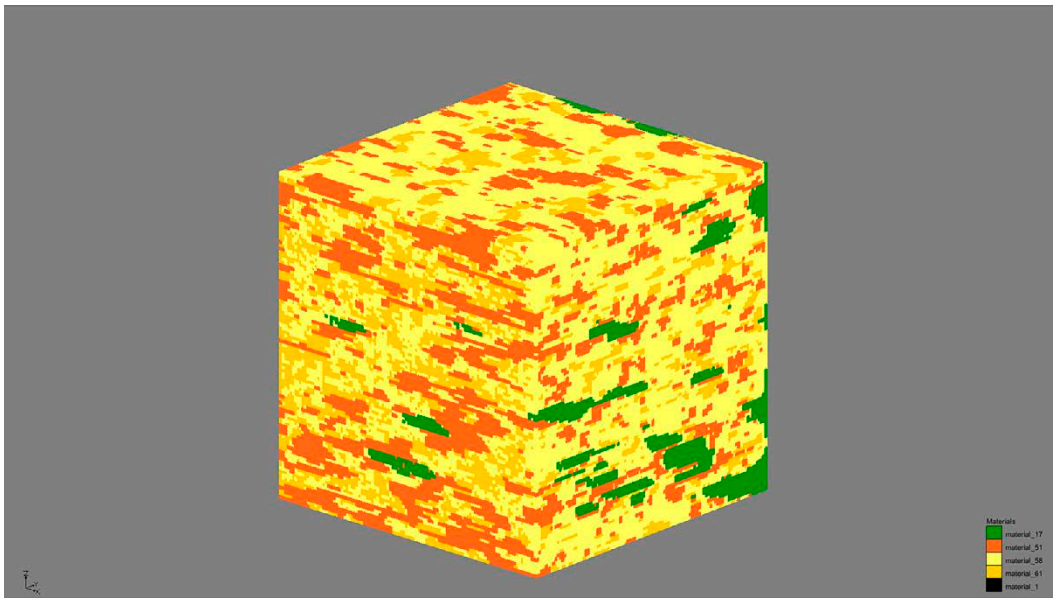


Figure 4-7. *Example of T-PROGS realization for a cubic model geometry of 100 × 100 × 100 meters with a cell size of 1 × 1 × 1 meter. Unconditional simulation with a total model size of 1 000 000 cells.*

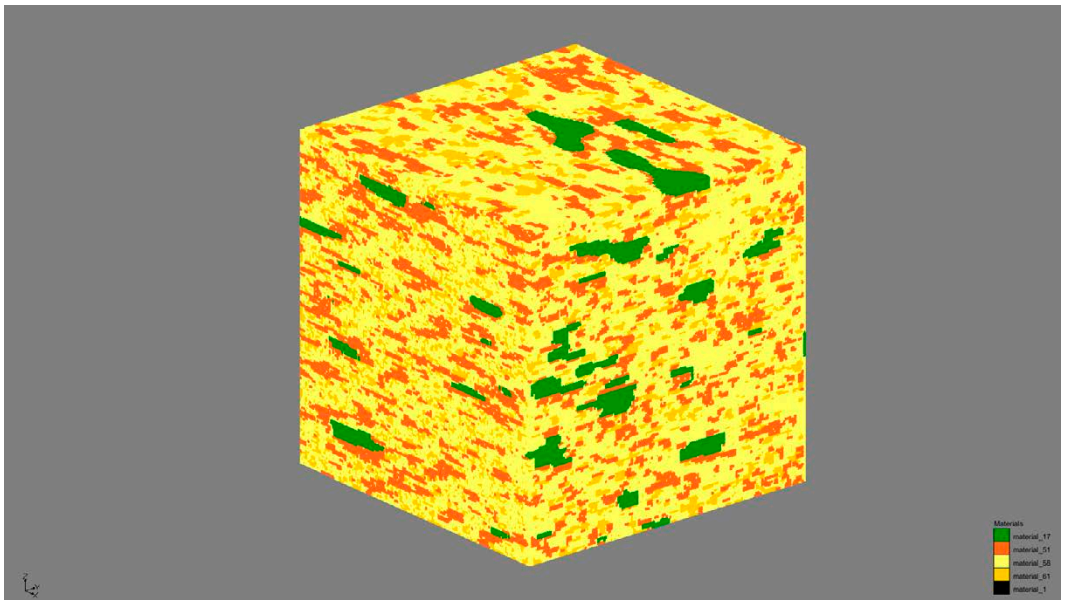


Figure 4-8. *Example of T-PROGS realization for a cubic model geometry of 150 × 150 × 150 meters with a cell size of 1 × 1 × 1 meter. Unconditional simulation with a total model size of 3 375 000 cells.*

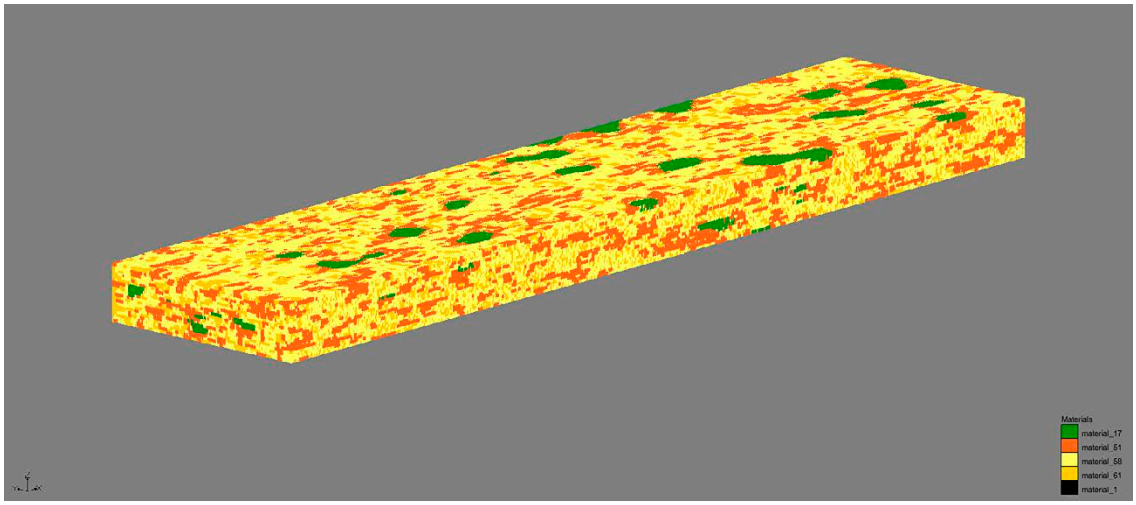


Figure 4-9. *Example of T-PROGS realization for a cuboid model geometry of dimensions 720 × 160 × 20 meters with a cell size of 1 × 1 × 1 meter. Unconditional simulation with a total model size of 2 304 000 cells.*

Additional testing with an increased number of material classes (TRCs) to 5 (maximum number of classes in T-PROGS) was made to investigate the impact on simulation times. For upcoming geological simulations of the repository area, the maximum number of material classes (TRCs) may be needed.

In addition, it was also investigated how different anisotropy directions for a fifth TRC would impact simulation times. The following scenarios were modelled with 5 TRCs:

A. Fifth TRC with the same relative anisotropies as for the other TRCs.

B. Fifth TRC with relative anisotropies representing a discordant orientation of the TRC:
   - X-direction: 2
   - Y-direction: 5

The simulations were made (50 realizations in each) for the cuboid model geometry of dimensions 720 × 160 × 20 meters with a cell size of 1 × 1 × 1 meter, i.e. a model size of 2 304 000 cells, using unconditional simulation.

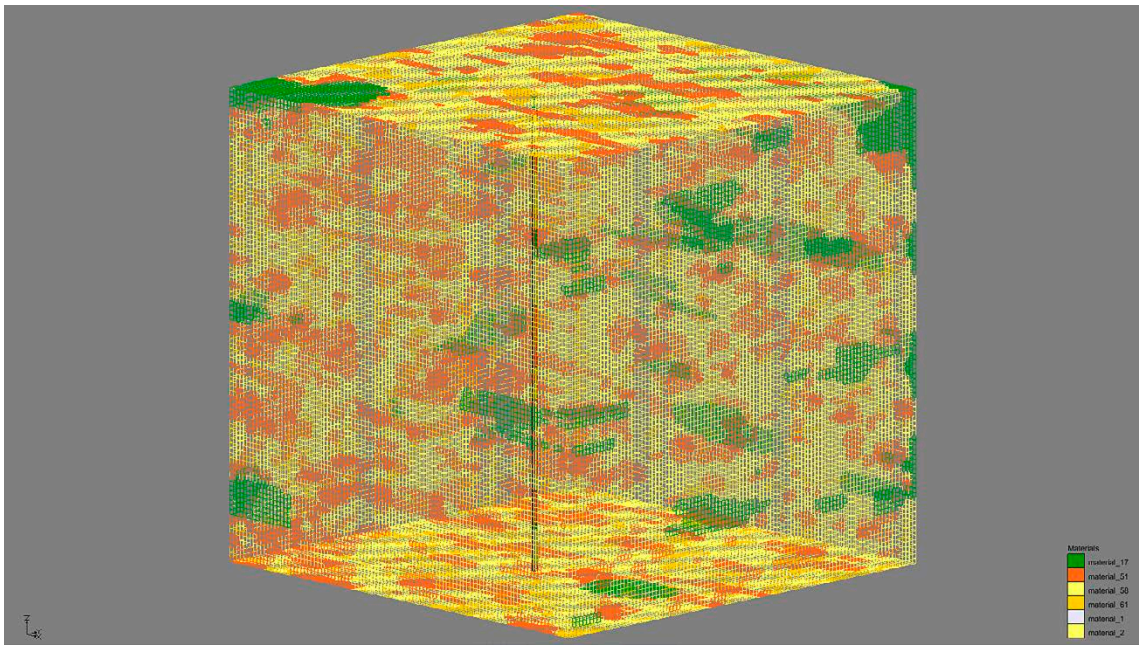


Figure 4-10. *Example of T-PROGS realization for a cubic model geometry of 100 × 100 × 100 meters with a cell size of 1 × 1 × 1 meter. Conditional simulation using 1 borehole at the center of the simulation volume. Total model size: 1 000 000 cells.*

The reason for choosing these anisotropies is that it may be needed to introduce pegmatite as a TRC and that according to the latest geological studies pegmatites occur both as concordant veins but also as discordant dikes, cutting through all other rocks.

The results of the simulation times are shown in Table 4-2. An example of a realization is shown in Figure 4-11.

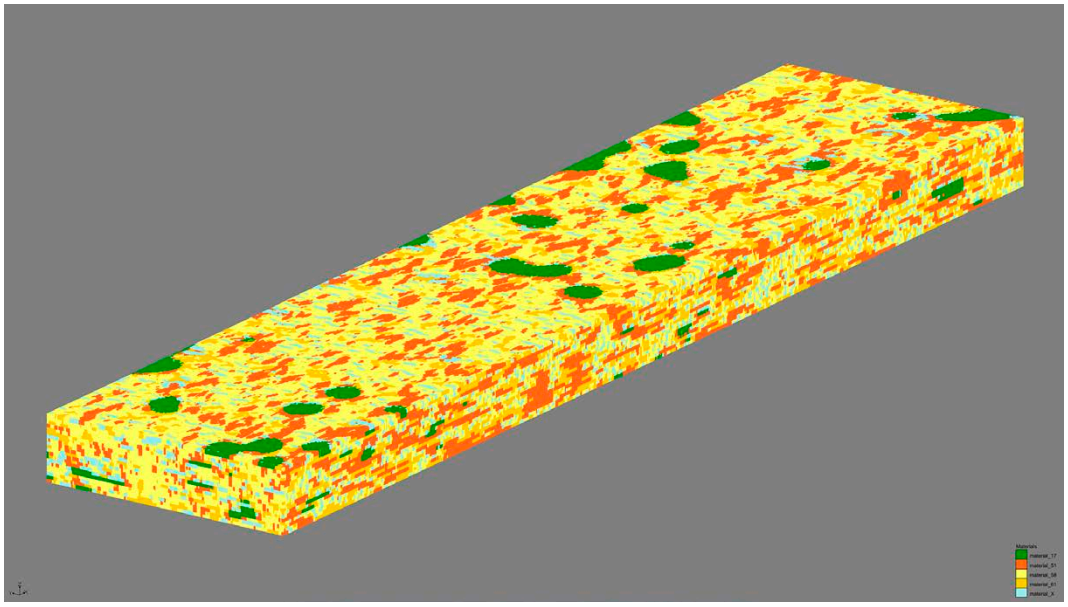


Figure 4-11. *Example of T-PROGS realization for a cuboid model geometry of dimensions 720 × 160 × 40 meters with a cell size of 1 × 1 × 1 meter. Unconditional simulation. Total model size: 2 304 000 cells. Five material classes (TRCs). Anisotropy for fifth material (material_x) in this realization: X = 2 and Y = 5 (see Table 4-2).*

Table 4-2. Simulation times for cuboid model with dimensions 720 × 160 × 40 meters with cell size 1 × 1 × 1 meter, including four and five material classes (TRCs) with different anisotropy directions.

Model size (cells)	No of cells in the model	*No of TRCs*	*Anisotropy for a fifth TRC (material_X)*	*Simulation time 50 realizations (minutes)*
720 × 160 × 40	2 304 000	4	-	157
720 × 160 × 40	2 304 000	5	X: 2 Y: 5	220
720 × 160 × 40	2 304 000	5	X: 5 Y: 2	234

Finally, a set of simulations were made for the 100 × 100 × 100 cubic model representing uncertainty in the spatial property parameters. For this example, uncertainty in the proportions of TRCs was assigned according to Table 4-3. The three different scenarios were chosen to represent three different possible sets of TRC proportions based on a hypothetical uncertainty analysis.

Table 4-3. Possible hypothetical sets of proportions for TRCs.

TRC	Proportion set N_1	Proportion set N_2	Proportion set N_3
TRC17	0.073	0.14	0.05
TRC51	0.189	0.11	0.22
TRC58	0.577	0.56	0.65
TRC61	0.161	0.19	0.08

Results of simulations are shown in Figure 4-12 to Figure 4-14.

It can be seen from the simulations that relatively limited changes in the proportions of the TRCs can give different character to the rock volumes. The assigned property differences are hypothetical and may be smaller in situations of extensive borehole information. However, in the initial stages with sparse borehole information the uncertainties in spatial parameters may lead to substantial variation in lithological properties of the rock mass, and hence thermal properties. It is therefore recommended that the uncertainties of spatial property parameters are carefully analysed and considered in the stochastic lithological simulations of the Forsmark repository area.

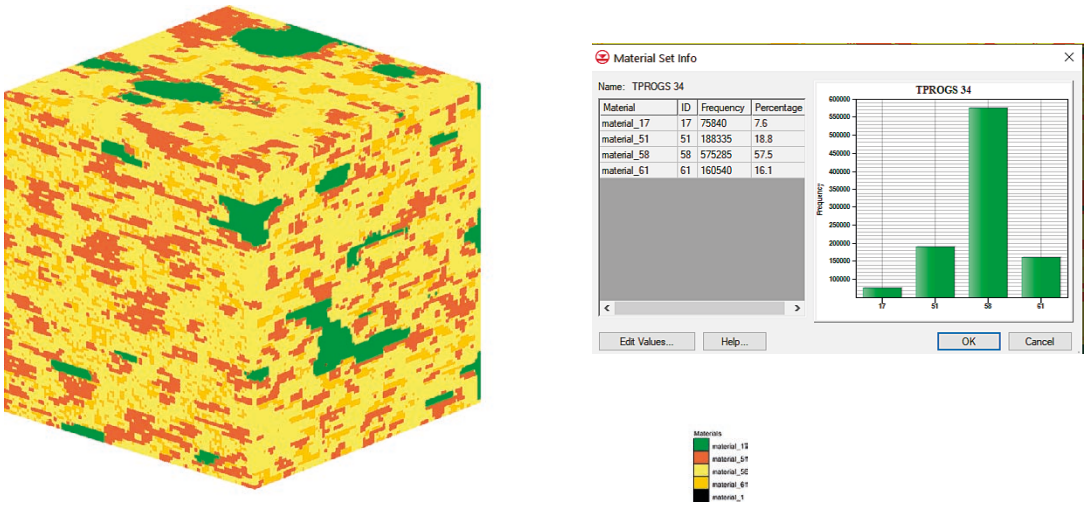


Figure 4-12. *Example stochastic simulation of TRCs with proportions N_1.*

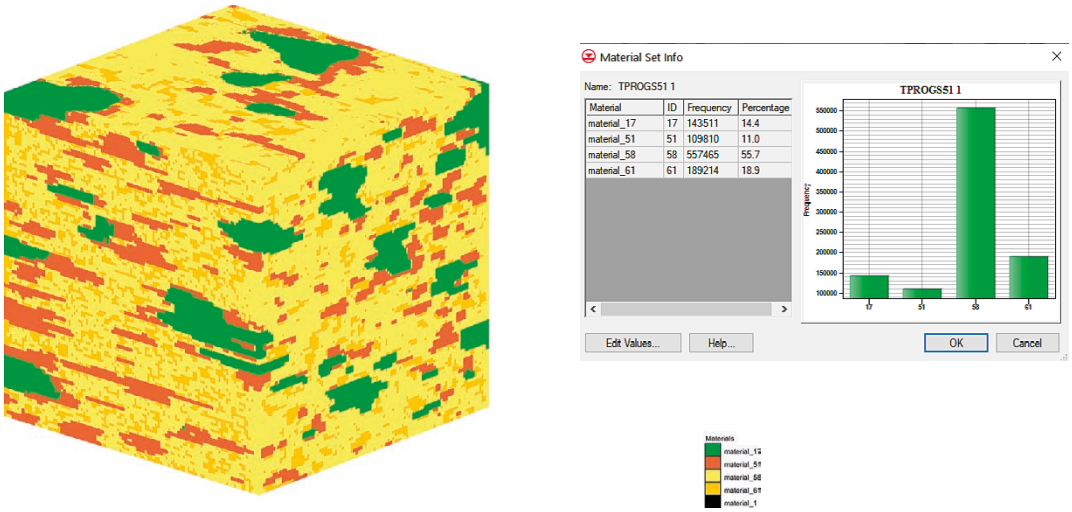


Figure 4-13. *Example stochastic simulation of TRCs with proportions N_2.*

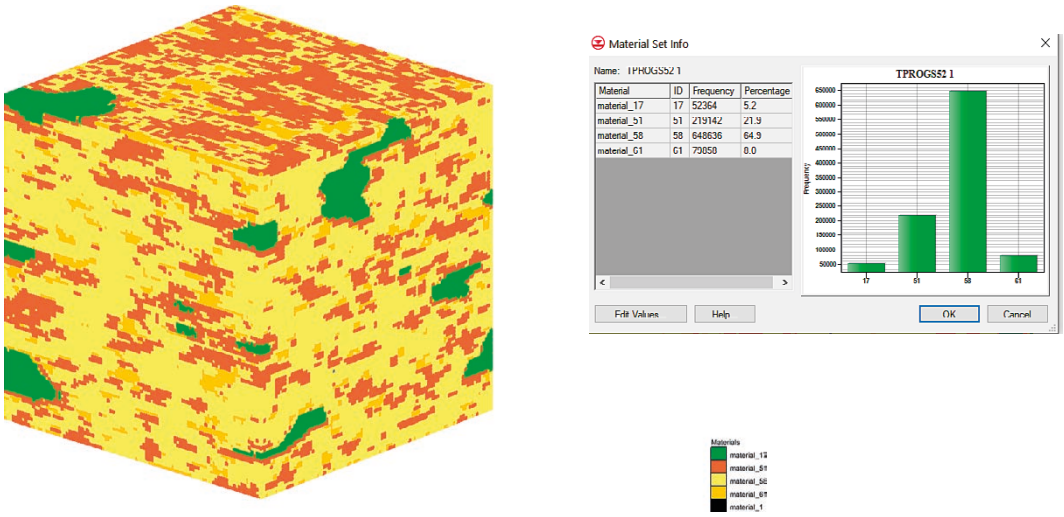


Figure 4-14. *Example stochastic simulation of TRCs with proportions N_3.*

4.7 Conclusions on lithological simulation

The main conclusions from this work are:

- The T-PROGS model is proposed to be used for modelling the lithology when modelling thermal properties in connection with the development of deposition areas of a spent fuel repository at the Forsmark site.
- Running time for the simulations are reasonable and T-PROGS can handle model sizes up to 4 million cells in various geometric layouts on an ordinary laptop computer.
- Lithological data must be combined with expert judgements to avoid potential bias in the conditioned modelling during the development of deposition areas. It is suggested that the SHELF method is considered to use for expert elicitation.

Motives and other conclusions are described below.

4.7.1 Conclusions on comparison of different models

When comparing two-point geostatistical approaches (here represented by the T-PROGS method) and multiple-point geostatistical methods (MPS), MPS methods have the capability of representing complex curvilinear geological structures in a more realistic way than two-point geostatistical approaches. However, the crystalline rock mass in the repository area is not expected to exhibit curvilinear features, such as folding or cyclic patterns, in the scale relevant for modelling thermal properties in the repository.

Given the challenges of MPS regarding e.g. conditional simulations outlined by several authors it is recommended that future geostatistical simulation of the spatial distribution of lithological units is performed using a two-point geostatistical approach.

The T-PROGS approach has previously been successfully applied to crystalline rock conditions representative of the future Swedish repository area. The method has proven to represent the spatial distribution of lithological units reasonably well and the approach is recommended for future stochastic lithological simulation for modelling of thermal properties.

The T-PROGS method facilitates a straightforward approach for integrating hard data from borehole information with geological interpretations of e.g. anisotropy, size and relative positions of lithological units. As geological interpretations will by necessity play an important role, especially in the initial stages of the repository construction, using an approach that is intuitive and in agreement with how geological interpretations are made is an advantageous method property.

4.7.2 Conclusions on stability and running time

The T-PROGS model, implemented in the Groundwater Modeling System (GMS) software interface, has been tested for different model configurations and sizes, different numbers of categories (Thermal Rock Classes, TRCs), different spatial statistical properties, and for unconditional and unconditional simulations. The conclusion is that the model code is stable and that it can handle model sized up to 4 million cells in various geometric layouts, e.g. cubic and cuboid models, on an ordinary laptop computer. The running times for such models with up to a maximum of 5 categories (TRCs) is around 5 hours for 100 realizations. In case of larger models, computers with stronger CPU performance is needed.

4.7.3 Handling of bias data due to direction of boreholes

It is described how bias in the information used for estimating spatial statistical properties of lithological categories (TRCs) due to boreholes deviating from anisotropy directions can be corrected for by transformation of the model domain to anisotropy directions. However, there will still be bias needed to be corrected for since more information will by necessity be available from the vertical or sub-vertical directions than the horizontal directions, especially in the initial stages of repository construction. The bias is expected to decrease when the number of horizontal boreholes increase, as the construction of the repository progresses. It is suggested that expert elicitation is explored for improving the information and decreasing bias regarding spatial statistical properties such as lengths, orientations and proportions of lithological units. It is suggested that the SHELF method is considered to use for expert elicitation.

5 Thermal anisotropy

5.1 Anisotropy in the rock mass

The rock mass at Forsmark have anisotropic thermal properties. There are three types of thermal anisotropy to consider:

1. Anisotropy due to foliation/lineation within a rock type.
2. Anisotropy due to orientation of subordinate rock bodies.
3. Anisotropy due to heterogeneity within a rock type.

The first type is a structural anisotropy caused by foliation and lineation which occur within a rock type. The foliation and lineation imply a directional orientation of the minerals in the rock mass. The thermal conductivity is generally higher parallel with the mineral foliation and lower perpendicular to the foliation plane.

This is because conductive minerals will control the heat flow parallel to the foliation; the minerals extend longer in this plane and are not interrupted to the same extent by less conductive minerals. Perpendicular to the foliation there is a higher density of transitions between different minerals, resulting in less conductive minerals having greater influence. This is accentuated by the crystallographic orientation of the commonly occurring minerals in a rock, such as quartz and biotite.

The second type of anisotropy is a result of the spatial orientation of magmatic rock bodies, primarily subordinate rocks. These bodies may have preferential directions in space, resulting in anisotropy of the thermal properties. Amphibolites parallel to the foliation at Forsmark are typical examples of this anisotropy (SKB 2005).

In addition to these types of anisotropy there are also other types that could occur, at least theoretically. Anisotropy may be caused by heterogeneity within a rock type, i.e. by different spatial trends in the composition of a rock type in different directions. This anisotropy could be modelled by directional variograms, i.e. separate variograms for the three principal directions.

The first type is described and evaluated below in this chapter. The second type is taken into account in the lithological modelling. However, depending on the method to assign TRCs at the simulation scale (including the assignment of lithology to each TRC), the thermal anisotropy may be influenced. This may be the case when sub-ordinate and main rock types are included in the same TRC and the sub-ordinate rocks bodies have a predominant orientation. The impact is accentuated in conditioned modelling.

There is no geological evidence to indicate that the third type of anisotropy is significant in Forsmark. Therefore, it is not been further considered in the thermal modelling approach.

5.2 Anisotropy in thermal conductivity

All rock types in rock domains 29 and 45 at Forsmark have been subjected to some degree of ductile deformation, both lineation and foliation. The preferred alignment of mineral grains produced by this deformation may produce anisotropy in thermal transport properties. Anisotropy of thermal properties in the rock mass may impact the design of a deep repository. For this reason, the directional dependence of thermal properties caused by foliation within deformed granite at the Forsmark site has been investigated in a large-scale field experiment, combined with laboratory and field testing at a smaller scale (Sundberg et al. 2007). The investigations were carried in the vicinity of drill site 7, situated in the outer margins of the Forsmark tectonic lens. The rock at this site is characterised by relatively intense tectonic foliation, striking NNW and steeply dipping.

Laboratory measurements of anisotropy at the centimetre scale, using a modified TPS method, indicate that thermal conductivity parallel with the foliation is, on average, a factor of 1.4 larger than conductivity perpendicular to the foliation. Field measurements, including the large-scale experiment which measure larger volumes of rock, have yielded anisotropy factors of approximately 1.15, considerably lower than for the laboratory measurements.

A distinct scale dependence for the anisotropy factor is indicated. A plausible explanation for the scale dependence is that the foliation is more clearly defined at the cm scale compared to larger scale. However, the variability in anisotropy still appears to be quite large at the dm to m scale.

The results of the investigations, with regard to the scale dependence of anisotropy of thermal conductivity, are summarised in Table 5-1.

Table 5-1. Results of investigation with regard to the scale dependence of thermal conductivity (Sundberg et al. 2007). Factor of anisotropy and the effective thermal conductivity (geometric mean of the two principal directions).

Scale	Thermal conductivity Mean factor of anisotropy $\lambda_{pa}/\lambda_{pe}$	Geometric mean W/(m·K)
Centimetre scale	1.40	3.45
Decimetre to metre scale	1.15	3.42

$\lambda_{pa}/\lambda_{pe}$ = thermal conductivity parallel to the foliation divided by thermal conductivity perpendicular to the foliation.

Given that most of the sampled rock displays less intense deformation than what is shown by the rock close to drill site 7, lower anisotropy factors may be expected for the larger part of the target volume.

5.3 Evaluation and conclusions

Introducing anisotropy in the thermal modelling implies a more complex modelling. Three different data sets of thermal conductivity are required for each rock type, one for each principal direction. This puts demand on both the thermal anisotropy data for all significant rock types, the thermal modelling strategy, and the computer capacity.

The thermal data on anisotropic properties are very sparse at Forsmark. In order to consider the introduction of anisotropy in the thermal modelling, much more data must be obtained.

The following conclusion and recommendations can be made:

- There is no reason to introduce anisotropy in the thermal conductivity modelling since available data is very sparse.
- To introduce anisotropy in the thermal modelling, the following requirements should be met:
  - Much more anisotropy data for each rock type or TRC, must be available.
  - Significant anisotropy in thermal conductivity in the supplementary data.
  - Significant variation of anisotropy in different parts of the rock mass in the supplementary data.
- Until these requirements are met, it is recommended that the anisotropy, in a comprehensive manner, is handled by the recipients of the thermal model, such as an application for temperature and heat transport calculations. Since data are very sparse, and with significant variability, it is recommended that more data be obtained even for a schematic use.

In addition, it should be noted that the normal thermal conductivity measurements by the TPS method were made based on the assumption of isotropy. The values represent the effective thermal conductivity of the measured volume, without regard to any particular direction. Thus, there may be some uncertainty in the TPS data due to anisotropy.

The strategy to create TRCs, and the assignment of TRCs at the simulation scale, may influence the thermal anisotropy of each TRC and the modelling results (see anisotropy due to orientation of subordinate rock bodies in Section 5.1). Therefore, it is recommended to investigate this when dealing with the issue of new TRC definitions (Chapter 3) and the assignment of thermal properties to each TRC.

6 Thermal simulation software

6.1 Previous work

A strategy for describing, predicting and visualizing the thermal properties of rock volumes has previously been developed for use in site descriptive modelling (Back and Sundberg 2007, Back et al. 2007). For the construction and operational phases of the repository, the main objective will be prediction of thermal conductivity in a specific rock volume. Of special interest is prediction of thermal conductivities in the deposition tunnels and around the deposition holes, although other objectives cannot be ruled out. Prediction requires conditional stochastic simulation, i.e. data from specific spatial locations are honoured during the simulation.

The methodology for thermal modelling is outlined in Section 1.3. There are several steps in the thermal modelling of a rock volume. A key part of the thermal modelling is to model the spatial statistical structure of thermal conductivity (or other thermal property) and perform stochastic simulation to produce a spatial distribution of thermal conductivity that is representative of the modelled rock type or rock volume. Related steps are change of support (inference about the values of a variable in rock volumes different from those at which it has been observed) and merging of realisations of thermal conductivity with realizations of lithology. Adaption of the thermal modelling strategy for conditional simulation may have implications for the functionality requirements of software used for stochastic simulation.

In previous modelling work, geostatistical analysis and simulation were performed by the software WinGslib and supplemented by Fortran programs written by the modelling team.

This short report sets out to investigate if new geostatistical software packages and/or faster processors can contribute to better and/or faster modelling results. This discussion stems from a need for increased efficiency in modelling work as well as new requirements related to conditional simulation for prediction purposes. In relation to previously implemented methods using WinGslib as well as our own Fortran programs, there is a need for improvements as regards interfacing with other software, data management, automation of certain modelling tasks and post processing of simulation results. In addition, tools for data processing, automatization and visualization are discussed. It is desirable that the 3D visualisations can be updated and reinterpreted as the rock volumes of interest are better and better investigated.

6.2 Modelling tools/software

6.2.1 Alternatives

There is a wide range of geostatistical software available, varying considerably as regards price, operating systems, user-friendliness, functionalities, as well as graphical and visualization capabilities.

The following criteria have been considered when compiling a short-list of potentially usable software for thermal modelling:

- capable of handling data collected in 3D,
- performs stochastic simulation in 3D,
- flexible – either allows access to source code in order to revise existing algorithms and add new algorithms, or allows addition (plug-in) of new algorithms to augment functionality, and
- variogram modelling – allows the user to customize variogram models.

Software that meets these requirements are listed in Table 6-1. In addition, a software package (EVS) that does not meet the simulation criteria is included. Although simulation cannot be performed with EVS, it is flexible enough to interface with other programs that can execute simulation. Three of the software are freeware or can be obtained at low cost. The other two are commercial products with expensive licences.

Table 6-1. Selected geostatistical software with the corresponding reference.

Name of software	Code	Open source/ Access to the source code	Allows addition of new geostatistical algorithms	Cost	Reference
WinGslib	Fortran	Yes	Yes	Freeware/Low	Deutsch and Journel (1998). http://www.statios.com/WinGslib/
S-GeMS	C++	Yes	Yes (Python)	Freeware/Low	Remy et al. (2009). http://sgems.sourceforge.net/
Gstat	C, R	Yes	Yes	Freeware/Low	Pebesma and Wesseling (1998), Pebesma (2004). http://www.gstat.org/
ISATIS	?	No	Yes (Python)	Commercial/High	https://www.geovariances.com/en/software/isatis-neo-geostatistics-software/
EVS (C-Tech)	?	No	Yes (Python)	Commercial/High	Earth Volumetric Studio, C Tech Development Corporation. http://www.ctech.com/products/earth-volumetric-studio/

GSLIB is a collection of geostatistical programs and utilities originating at Stanford University. GSLIB, from which WinGslib developed, was first released in 1992 and appears not to have undergone any updating and maintenance in recent years. A method for accelerating GSLIB routines by (Fortran 77) code optimization and multi-core programming has been developed by Peredo et al. (2015). Using all CPU resources, GSLIB algorithms can handle large datasets and grids, which is especially useful for computing and memory intensive applications.

S-GeMS, also emerging from Stanford, was developed circa 2005. S-GeMS incorporates most of the algorithms from its precursor GSLIB. S-GeMS has been widely used by researchers in a variety of geological investigations (e.g. Li and Caers 2008).

Development of Gstat began in 1993 and the software is currently maintained by one of the original authors/developers. It does not, however, appear to have been used extensively to tackle geology-related problems.

ISATIS is a commercial software which is actively maintained and supported. Most of the algorithms originate from the Center of Geostatistics at Mines ParisTech in Paris. ISATIS is used extensively in the mining and petroleum industries.

C-Tech's software EVS (Earth Volumetric Studio) was developed in 2006 from older C-Tech programs. It is a commercial software which is actively maintained and supported.

6.2.2 Open source code versus black-box software

Table 6-1 above lists the programming language whenever the source code is known.

Both ISATIS and EVS software can be considered as black-box software; the code cannot, generally speaking, be examined or modified. However, the latest versions of these programs support a plug-in mechanism to expand their functionalities, allowing for the addition of new geostatistical algorithms (using Python) in order to customise the modelling workflows. S-GeMS also facilitates the use of plug-ins to add new geostatistical algorithms.

6.2.3 Main functionalities

Most geostatistical investigations share a similar sequence of steps:

- exploratory data analysis to get familiar with the data,
- characterization and modelling of the pattern of spatial variation,
- interpolation or simulation to the nodes of a grid or over blocks (upscaling), and
- modelling of local and spatial uncertainty.

A single software product may not comprise all the functionalities required by a specific application. The functionalities of the software on the short-list are outlined in the following sections.

Data management and exploratory data analysis

The software in Table 6-1 differ in their ability to handle 3-dimensional databases, as well as in their capacity to explore and visualise the data. EVS and ISATIS allow sophisticated visualisation of three-dimensional data. S-GeMS has visualisation tools for 3D datasets but lack some of the more advanced capabilities of the EVS and ISATIS, such as data queries and other GIS capabilities. WinGslib and Gstat are considerably more limited as regards exploratory data analysis and visualisation.

Variogram estimation and modelling

Modelling of the spatial structure of data is the basis of any geostatistical analysis. Options vary from fully automated computation and modelling of variograms to highly interactive programs that allow the detection and elimination of spatial outliers, the exploration of spatial anisotropy through variogram maps or surfaces, and the manual fitting of variograms. The commercial products EVS and ISATIS provide an automatic 3D modelling procedure. ISATIS is probably the most flexible software since it allows identification of directions and scales of continuity using 3D variogram maps. In contrast, public-domain software requires both a lot of data manipulation and expert knowledge in the modelling of variograms. S-GeMS software can compute variograms in three directions, but only visual fitting is implemented.

Upscaling of the variogram model, from measurement scale to larger blocks, can be performed by the ISATIS program.

Stochastic simulation

By performing stochastic simulation one can generate a set of equiprobable representations (realizations) of the spatial distribution of attribute values (e.g. thermal conductivity) and use differences among simulated realizations as a measure of uncertainty.

Software may differ in their ability to handle irregular simulation grids or uneven prediction supports. S-GeMS and Gstat allows the specification of user-defined simulation grids (the size of a cell can be defined in the X, Y and Z directions) instead of the traditional regular grids permitted by WinGslib.

The most complete range of simulation methods, covering both continuous and categorical variables, is provided by ISATIS. S-GeMS implements the most common algorithms (i.e. sequential indicator and Gaussian simulations), as well as recent methods based on multiple point statistics. Sequential Gaussian simulation in WinGslib has been used in the previous thermal modelling work and, except for EVS, can be implemented by the other software treated here.

Simulation of continuous variables generally use Gaussian simulation algorithms. This requires normal score transformation and subsequent back-transformation. Direct sequential simulation (DSSIM) is also designed for continuous variables. DSSIM does not require any Gaussian assumption so it can be applied to the original variable without any preliminary normal score transform (Remy et al. 2009). There may be some disadvantages however; see Remy et al. (2009) for details. DSSIM can be implemented by S-GeMS and ISATIS but not by the other software on the shortlist.

EVS does not have simulation capabilities. It does, however, have a module that was designed to allow users to export their grids (including stratigraphic and/or lithologic geology) and their analytical data into S-GeMS which can then perform the simulations. The same module will import the S-GeMS results and then allow those results to be used by EVS as if they were created internally in EVS (pers. comm. Reed Copsey, C-Tech 2019).

Change of support problems are complicated, especially when only point-support data (very small support) are available. In many applications, for example mining, one needs to simulate the mean value within blocks rather than point values. The same is true for thermal modelling of a rock volume. The simplest method for simulating a block is to discretize it into a fine enough grid and simulate the nodes by point (small cells) simulation. The simulated block values are the means of their simulated discretization points. Disadvantages of this procedure are that simulation of a large model volume is time consuming and requires a large amount of computer memory.

Lower-upper (LU) Decomposition is a 3D simulation technique that was used for the purpose of change of support (upscaling) as a preparatory step to unconditional simulation in the site descriptive modelling (Sundberg et al. 2008). The LU simulation algorithm is provided by both WinGslib and S-Gems, but not by EVS or ISATIS. It is not clear whether Gstat has the LU simulation function. The major disadvantage of this algorithm is that it is computationally demanding (unrealistic) for simulation of large model volumes. Furthermore, it is not commonly used for conditional simulation.

To solve the problem of change of support one can perform a direct simulation of blocks. Assuming that only point-support data are available, then the general approach to creating a block-support distribution or affecting a change of support is to perform a transformation from the distribution of the point-support Z(x) to the distribution of the block-support Zv(x).

Several change-of-support models are presented in Chilès and Delfiner (2012). Direct simulation of blocks using the Gaussian Discrete model is one alternative and according to Chilès and Delfiner (2012) the most widely used method. An interesting aspect of the Gaussian Discrete model is that the position of each data point (small support samples) is not fixed but is located randomly within its block. This model allows conditioning on both point and block data. The software ISATIS performs change of support using this Gaussian Discrete model.

Block simulation incorporates a change-of-support model into the simulation process, instead of point simulation followed by scaling up. S-GeMS and ISATIS include algorithms for dealing with both point data and block data; block kriging and simulation of properties using both point data and block data are facilitated. S-GeMS provides an algorithm not available in its precursor WinGslib for simulations conditioned to both block and point data. The algorithm, called Block sequential simulation, utilizes block kriging and direct sequential simulation (Remy et al. 2009, Emery and Ortiz 2011). A key step in this simulation method is computing block-to-block, block-to-point, point-to-block and point-to-point covariances. S-GeMS includes a program for calculation of different types of covariances (Remy et al. 2009). An example of how direct block simulation has been applied can be found in Hosseini et al. (2017).

Gstat also includes algorithms for change of support and block prediction.

Avoiding point simulation of a large model volume results in a considerable reduction in memory, time, and storage space requirements. A disadvantage is that the models/realisations cannot be resized to other scales post simulation as can be done with the results of a point simulation.

Verification

Various implementation schemes (variogram models and search strategies) can be compared by cross-validation. With this approach, a partial data set is used to estimate values at actual sampled positions. “Real” and “estimated” values are then compared in such a way that the model can be accepted or rejected. Cross-validation is not part of S-GeMS, but it can be found in WinGslib. The extent to which cross-validation can be efficiently performed by the other software considered here has not been established.

Probability/uncertainty mapping

In order to assess the range of possible results, it is necessary to generate a large number of realisations. Uncertainty of stochastic simulated values is often assessed using maps of the probability of exceeding critical values, such as regulatory or critical thresholds. Algorithms for estimating uncertainties are available in S-GeMS as well as ISATIS. In the previous thermal modelling work, an algorithm was written in Fortran specifically for the purpose of estimating percentiles and probabilities. The merged realisations of thermal conductivity were used as input to these estimations.

Postprocessing

The software packages investigated here vary in their ability to post-process the set of realisations generated by stochastic simulation. All (with the exception of EVS which does not perform simulation) allow statistical summaries to be extracted and histograms to be created from the realisations. Both ISATIS and S-GeMS software packages can create maps of averaged simulated values and

maps of probability of exceeding critical thresholds. S-GeMS can estimate measures of differences among realisations. ISATIS can perform Simulation Reduction to extract a representative subset of the simulations. WinGslib also has some post-processing capabilities, e.g. creating maps of averaged simulated values. It is unclear if Gstat can perform any of these functions.

Table 6-2 summarizes the main functionalities of the investigated software.

Table 6-2. Functionalities of the investigated geostatistical software packages.

Name	3D anisotropic Variography	Kriging/ Co-kriging	Simulation	Post-processing of realisations	Maximum model size based on grid nodes
WinGslib	Yes	Yes	Yes	Averages, probabilities	Information not found
S-GeMS	Yes	Yes	Yes	Averages, probabilities, variance	Information not found
Gstat	Yes	Yes	Yes	Unclear	Over 1 Million
ISATIS	Yes	Yes	Yes	Averages, probabilities, variance	No upper limit
EVS (C-Tech)	Yes	Yes (kriging), no co-kriging	No	Not relevant	Over 100 Million

6.2.4 User friendliness

ISATIS provides a large range of functionalities within a single package. Of the software presented here ISATIS probably offers the best system for data management and automatic batch procedures. It interacts with other software packages through numerous data exchange plugins and import/export interfaces. Production workflows may be set up and automated through scripting.

An important aspect of user-friendliness is the ability to automate some of the modelling steps, in particular the variogram modelling procedure which is typically very time consuming. EVS analyses the input data and constructs a multidimensional variogram which is a best fit to the dataset being analysed.

S-GeMS is a Windows-based software package with a graphical user interface. It is based on the original GSLIB suite of DOS-based software. According to the developers, user-friendliness was greatly improved compared to WinGslib. For example, it is possible to combine several commands (one for each task) and have S-GeMS execute them all sequentially. This would be advantageous when performing a sensitivity study. A Python script may also be used with S-GeMS to automate this workflow. S-GeMS can also be integrated with Matlab.

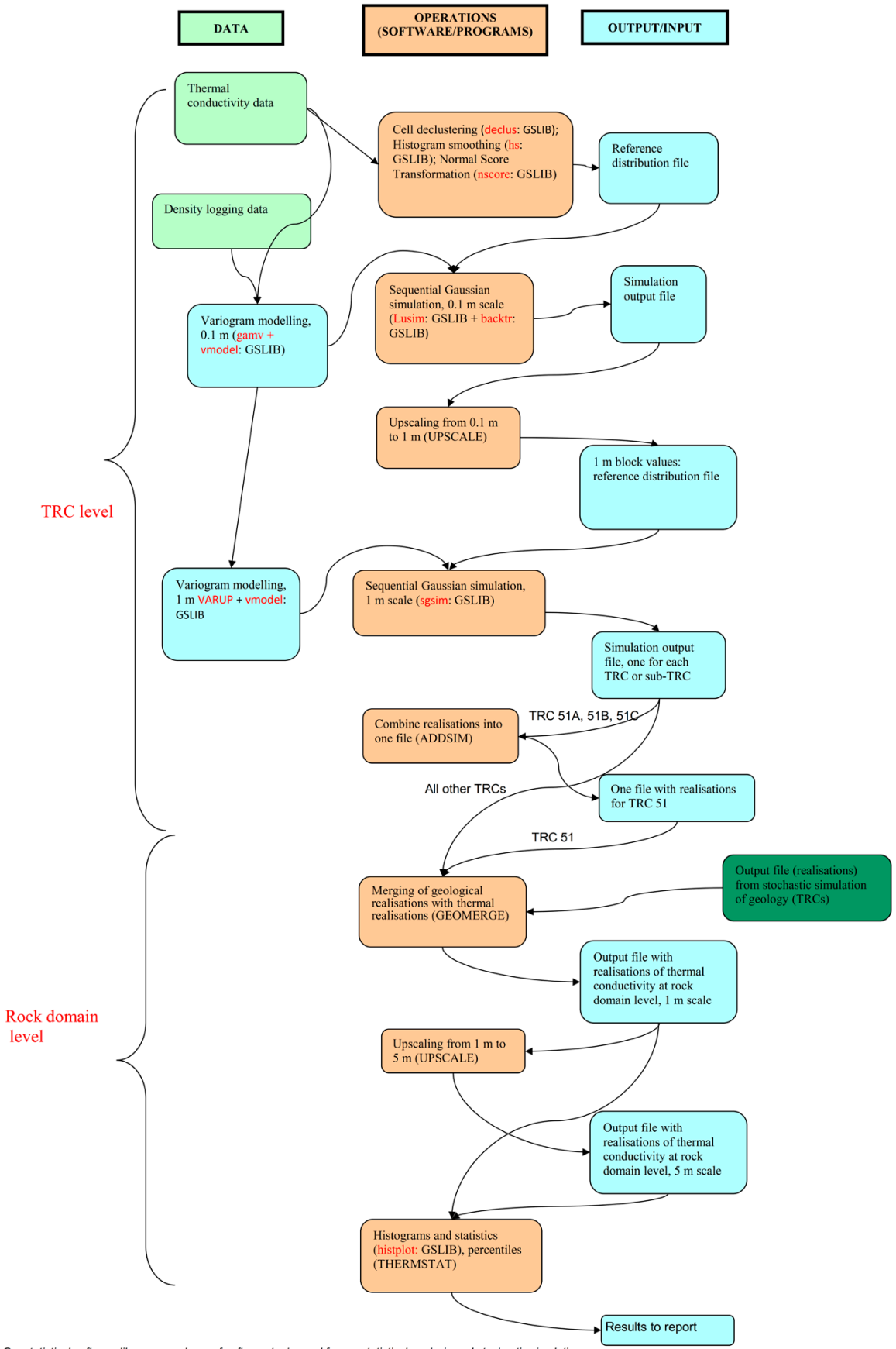
S-GeMS and Gstat offer graphic capabilities for the analysis and display of results. ISATIS and EVS allow more flexible 3D analysis and provide more sophisticated visualisation capabilities than any of the other packages. The latter two software also integrate with GIS interface.

6.2.5 Capability in thermal modelling

Thermal modelling carried out in the site descriptive modelling phase consisted of several steps (Figure 6-1). Appendix C shows the steps, as described in Back and Sundberg (2007), pertaining, directly or indirectly, to the (unconditional) stochastic simulation of thermal conductivity. Not included here are modelling steps involving lithological modelling. For each modelling step, the task(s) performed, and the program(s) used to execute the task(s), are listed. The ability of the investigated software packages to perform these tasks is also presented.

S-GeMS and ISATIS can perform most or all the tasks that WinGslib performed in the previous thermal modelling. ISATIS has greater functionality as regards postprocessing, being capable of calculating percentiles and extracting a representative subset of realisations. Programs written specifically for thermal modelling, namely VARUP, GEOMERGE and UPSCALE, are not generally incorporated by the software packages dealt with here. ISATIS does however have an algorithm corresponding to VARUP, for the upscaling of variograms.

Flow chart describing steps in stochastic simulation of thermal conductivity in thermal modelling Forsmark 2.2.



GSLIB - Geostatistical software library: a package of software tools used for geostatistical analysis and stochastic simulation
All other programs - written by the thermal modelling team

Figure 6-1. *Modelling steps and programs used in the site descriptive thermal modelling at Forsmark (after Back et al. 2007).*

For conditional stochastic simulation, some of the above steps may be omitted or require modification. Change of support of thermal data, using the procedure outlined in Appendix C, will most likely require revision because of difficulties in incorporating the conditioning data into the simulation by LU decomposition[5]. If thermal data is available at a scale (support) that is suitable for simulation without upscaling, then this step can be omitted. Alternatively, if there are large amounts of small-scale data along continuous sections of boreholes or tunnels, then change of support could be achieved by calculating means over regular sections (e.g. 1 m), thus avoiding using LU decomposition. This was done as a test in the site description modelling stage for Laxemar using density logging data from boreholes (Sundberg et al. 2008a). The thermal conductivities estimated from density were upscaled to 2 m scale before being used as conditioning data in the subsequent stochastic simulation.

6.3 Other programs

Depending on the geostatistical software used and the modelling tasks that are executed, there may arise a need for other programs that can streamline, speed up and automate the thermal modelling workflow. Spatial ETL (Extract, Transform, Load) tools may be useful in this respect. These tools may be used for 3D data management, customizing workflows, data conversion, automation of modelling workflows. Such a program can manage workflows by stringing together sequences of geoprocessing tools, feeding the output of one tool into another tool as input. They could be usefully employed, for example, to merge simulation results from thermal modelling with lithological modelling as *GEOMERGE* has done previously.

The commercial FME is one such spatial ETL software tool. Although FME has its own 3D capabilities it can communicate with ArcMap (GIS) for improved functionality. Free open source alternatives, e.g. Talend, also exist.

6.4 Concluding remarks

Two of the software packages examined above deserve further consideration; these are S-GeMS and ISATIS. The other alternatives do not fulfil the requirements of thermal modelling for the purpose of prediction during the production phase of a deep repository. Gstat was developed and used by researchers in the field of spatial analysis and geography. Although it has 3D capabilities, the software does not appear to have been used to any great extent for 3D problems in rock volumes. WinGslib is time consuming and lacks user friendliness. Recent efforts to accelerate modelling with GSLIB are interesting but do not solve the problem of poor user friendliness. EVS cannot carry out stochastic simulation. Programs that do meet the requirements are S-GeMS and ISATIS.

The pros and cons of S-GeMS and ISATIS are summarized below.

S-GeMS would allow changes/improvements to be made to the thermal modelling strategy without changing the fundamentals, since many of its algorithms have been developed for WinGslib, which was used in previous thermal modelling. S-GeMS is more user friendly, has better graphic capabilities, and in some respects has improved functionality in that it can perform some simulation methods not available in WinGslib. Of particular interest is its ability to perform block simulation. S-GeMS has even a certain level of automation allowing several tasks to be run sequentially. It can be acquired free of charge.

ISATIS is probably the most flexible and sophisticated geostatistical software available, as regards data analysis and visualization, variogram modelling, simulation methods, automation, no practical limits on model size (apart from time considerations) and interfacing with other programs including GIS. In addition, it claims to be user friendly. Although a drawback is that the source code cannot be examined or modified, S-GeMS does allow addition of new geostatistical algorithms (using Python) in order to customise the modelling. The main drawback with ISATIS is its high cost.

[5] LU decomposition is a method in numerical analysis and linear algebra for solving problems expressed in matrix form.

Given that S-GeMS is an improvement over WinGslib but yet retains many of the features of its predecessor, and can be obtained at little or no cost, it is recommended that this software be the first choice for testing the revised modelling strategy in a pilot study. After evaluating the suitability of S-GeMS, it may be necessary to test a more advanced product, in which case ISATIS is recommended. A demo version can be used initially to compare its performance with that of S-GeMS.

7 Application scenarios

7.1 Basis for scenarios

The thermal modelling methodology outlined in Section 1.3 and in Part 1 (Back and Sundberg 2022) can be applied for different scenarios. The scenarios can be based on different starting points. They can be divided into e.g. the following:

- Scenarios based on availability of data.
- Scenarios based on desired simulated volume.
- Scenarios based on planned modelling intervals in terms of time.

The first bullet may involve:

- Types of data available – thermal conductivity distributions, variogram models, lithological data.
- Scale of data acquisition – small scale, large scale.
- Data available – existing data, new data.
- Flow of data – when are new data available.

The second bullet may include the initial phase with small rock volumes involved. Subsequently, the modelled volumes grow, maybe up to the annual production of pilot holes or tunnels.

Some example scenarios are sketched in the following. The scenarios are based on different stages in the development of the repository, which provides access to successively more data. As one basis for this sketch, a draft of the operative investigation and modelling programme for deposition areas has been used.

7.2 Pilot test of developed modelling strategy

Simulations based on existing data from SDM-Site (SKB 2008) are assumed to include the following:

- Existing thermal TPS data – no new data are assumed to be available.
- Existing or developed definitions of TRCs – existing or modified thermal distributions and variograms from preceding site descriptive modelling. Potentially new data are added.
- Old and new lithological data – possible to use new lithological data from the area for the access to the repository, which have been added over the last 10 years (e.g. KFM024).

The scenario could be used for intended test of the updated strategy for thermal modelling, and also for updating of the baseline site descriptive model.

7.3 Modelling of a single main tunnel

Simulations based on existing data and new data (from accesses and central area) are assumed to be based on the following:

- Mainly existing thermal TPS data.
- Possible new thermal data acquired by the DTS-Heat method using integrated heating – gives thermal distributions in larger scale and, particularly, much better data describing the spatial correlation structure.
- Existing TRC-definitions – existing (or possibly developed) thermal distribution, developed thermal variogram models, old and new lithological data.

7.4 Modelling of a batch/set of deposition tunnels

Simulations based on existing data and new data from pilot boreholes and corresponding deposition tunnels are assumed to include the following:

- Mainly existing thermal TPS data.
- Data sets with thermal data acquired by the DTS-Heat method using integrated heating, which gives continuous profiles of thermal conductivity that provide thermal distributions at a larger scale and, particularly, much better data regarding the spatial correlation structure (for variogram modelling).
- New TRC-definitions with developed thermal distribution and variogram models.
- Old and new lithological data.

New additional data may come from pilot holes and tunnels in different directions:

- Pilot boreholes for the main tunnel.
- Pilot holes for deposition tunnel, perpendicular to the main tunnel.
- Pilot holes for the deposition holes, vertical and perpendicular to the other two.

The first two data sets may be biased regarding lithology since deposition tunnels are sub-parallel to amphibolite dykes.

Time offsets between different datasets of e.g. lithological data from one and the same tunnel can be expected (e.g. pilot holes for deposition holes are drilled after the tunnel is excavated). This may result in biased new data sets, especially during the first stages during development of repository areas.

A strategy for data supply is needed on how to add new data to the stochastic simulations in order to avoid problems of representativeness.

References

SKB's (Svensk Kärnbränslehantering AB) publications can be found at www.skb.com/publications. SKBdoc documents will be submitted upon request to document@skb.se.

Aarts E, Korst J, 1989. Simulated annealing and Boltzmann machines: a stochastic approach to combinatorial optimization and neural computing. New York: Wiley.

Back P-E, Sundberg J, 2007. Thermal site descriptive model. A strategy for the model development during site investigations – version 2. SKB R-07-42, Svensk Kärnbränslehantering AB.

Back P-E, Sundberg J, 2022. Methodology for modelling of thermal properties of the Forsmark site. Part 1 – Recommended data and interpretation methods. SKB R-20-18, Svensk Kärnbränslehantering AB.

Back P-E, Wrafter J, Sundberg J, Rosén L, 2007. Thermal properties. Site descriptive modelling Forsmark – Stage 2.2. SKB R-07-47, Svensk Kärnbränslehantering AB.

Barton C A, Zoback M D, 1990. Self-similar distributions of macroscopic fractures at depth in crystalline rocks in the Cajon Pass scientific drillhole. In Barton N, Stephansson O (eds). Rock joints: proceedings of the International Symposium on Rock Joints, Loen, Norway, 4–6 June 1990. Balkema: Rotterdam.

Bedford T, Cooke R, 2001. Probabilistic risk analysis: foundations and methods. Cambridge University Press.

Caers J, Zhang T, 2004. Multiple-point geostatistics: a quantitative vehicle for integrating geologic analogs into multiple reservoir models. In Grammer G M, Harris P M, Eberli G P (eds). Integration of outcrop and modern analogs in reservoir modeling. American Association of Petroleum Geologists. (AAPG Memoir 80), 383–394.

Carle S F, 1999. T-PROGS: Transition Probability Geostatistical Software. Version 2.1. Hydrologic Sciences Graduate Group. University of California, Davis, USA.

Carle S F, Fogg G E, 1997. Modeling spatial variability with one and multidimensional continuous-lag Markov chains. Mathematical Geology 29, 891–918.

Carslaw H S, Jaeger J C, 1959. Conduction of heat in solids. 2nd ed. Oxford: Clarendon.

Chilès J-P, Delfiner P, 2012. Geostatistics: modeling spatial uncertainty. 2nd ed. Hoboken, NJ: Wiley.

Dagan G, 1979. Models of groundwater flow in statistically homogeneous porous formations. Water Resources Research 15, 47–63.

Daly C, Caers J, 2010. Multi-point geostatistics – an introductory overview. First Break 28, 39–47.

Deutsch C V, Cockerham P W, 1994. Practical considerations in the application of simulated annealing in stochastic simulation: Mathematical Geology 26, 67–82.

Deutsch C, Journel A, 1998. GS-LIB: Geostatistical Software Library and User´s guide. 2nd ed. New York: Oxford University Press.

Emery X, Ortiz J M, 2011. Two approaches to direct block-support conditional co-simulation. Computers & Geosciences 37, 1015–1025.

GMS, 2019. Tutorials, Volume 1. Ver 6.0. Environmental Modeling Research Laboratory. Aquaevo. USA.

Guardiano F B, Srivastava R M, 1993. Multivariate Geostatistics: Beyond Bivariate Moments. Geostatistics Tróia 92 A. Soares (ed.). Kluwer Academic Publishers 1993, doi:10.1007/978-94-011-1739-5_12

Gutjahr A L, Gelhar L W, Bakr A A, MacMillan J R, 1978. Stochastic analysis of spatial variability in subsurface flows: 2. Evaluation and application. Water Resources Research 14, 953–959.

Haldorsen H, Damsleth E, 1990. Stochastic modeling. Journal of Petroleum Technology 42, 404–412.

Hashemi S, Javaherian A, Ataee-pour M, Khoshdel H, 2014. Two-point versus multiple-point geostatistics: the ability of geostatistical methods to capture complex geobodies and their facies associations – an application to a channelized carbonate reservoir, southwest Iran. Journal of Geophysics and Engineering 11, 065002. doi:10.1088/1742-2132/11/6/065002

Hashin Z, Shtrikman S, 1962. A variational approach to the theory of the effective magnetic permeability of multiphase materials. Journal of Applied Physics 33, 3125. doi:doi.org/10.1063/1.1728579

He X, Koch J, Sonnenborg T O, Jørgensen F, Schamper C, Refsgaard J C, 2014. Transition probability-based stochastic geological modelling using airborne geophysical data and borehole data. Water Resources Research 50, 3147–3169.

Hosseini S A, Asghari O, Emery X, 2017. Direct block-support simulation of grades in multi-element deposits: application to recoverable mineral resource estimation at Sungun porphyry copper-molybdenum deposit. Journal of the Southern African Institute of Mining and Metallurgy 117, 557–585.

Isaaks E H, Srivastava R M, 1989. An introduction to applied geostatistics. New York: Oxford University Press.

Journel A, 2003. Multiple-point geostatistics: A state of the art. Stanford Center for Reservoir Forecasting. Available at: https://pangea.stanford.edu/departments/ere/dropbox/scrf/documents/reports/16/SCRF2003_Report16/SCRF2003_Andre_Journel.pdf [10 May 2019].

Journel A G, Huijbregts C J, 1978. Mining geostatistics. London: Academic Press.

Li T, Caers J, 2008. Solving spatial inverse problems using the probability perturbation method: An S-GEMS implementation. Computers & Geosciences 34, 1127–1141.

Liu Y, 2006. Using the *Snesim* program for multiple-point statistical simulation. Computers & Geosciences 32, 1544–1566.

Matheron G, 1969. Le krigeage universel. Les Cahiers du Centre de Morphologie Mathématique, Fasc. 1. Fontainebleau, France: Ecole des Mines de Paris. (In French.)

Norberg T, Rosén L, Baran Á, Baran S, 2002. On modelling discrete geological structures as Markov random fields. Mathematical Geology 34, 63–77.

Oakley J E, O'Hagan A, 2016. SHELF: the Sheffield Elicitation Framework (version 3.0). School of Mathematics and Statistics, University of Sheffield, UK. Available at: http://tonyohagan.co.uk/shelf [6 May 2020].

O'Hagan A, 2019. Expert knowledge elicitation: subjective but scientific. The American Statistician 73 sup 1, 69–81.

Pebesma E J, 2004. Multivariable geostatistics in S: the gstat package. Computers & Geosciences 30, 683–691.

Pebesma E J, Wesseling C G, 1998. Gstat: a program for geostatistical modelling, prediction and simulation. Computers & Geosciences 24, 17–31.

Peredo O, Ortiz J M, Herrero J R, 2015. Acceleration of the Geostatistical Software Library (GSLIB) by code optimization and hybrid parallel programming. Computers & Geosciences 85, Part A, 210–233.

Remy N, Boucher A, Wu J, 2009. Applied geostatistics with SGeMS: a user's guide. Cambridge: Cambridge University Press.

Rosen L, Gustafson G, 1996. A Bayesian Markov geostatistical model for estimating hydrogeological properties. Groundwater 34, 865–875.

SKB, 2005. Preliminary site description. Forsmark area – version 1.2. SKB R-05-18, Svensk Kärnbränslehantering AB.

SKB, 2008. Site description of Forsmark at completion of the site investigation phase. SDM-Site Forsmark. SKB TR-08-05, Svensk Kärnbränslehantering AB.

Starks T H, 1986. Determination of support in soil sampling. Mathematical Geology 18, 529–537.

Stephens M B, Fox A, La Pointe P, Simeonov A, Isaksson H, Hermanson J, Öhman J, 2007. Geology Forsmark. Site descriptive modelling, Forsmark stage 2.2. SKB R-07-45, Svensk Kärnbränslehantering AB.

Stoyan D, Kendall W S, Mecke J, 1987. Stochastic geometry and its applications. Chichester: Wiley.

Sundberg J, 1988. Thermal properties of soils and rocks. PhD thesis. Chalmers University of Technology, Sweden.

Sundberg J, 2003. Thermal site descriptive model. A strategy for the model development during site investigations. Version 1.0. SKB R-03-10, Svensk Kärnbränslehantering AB.

Sundberg J, 2019. Fältmätning av värmeledningsförmåga. Förslag till handlingsplan. Omarbetning av handlingsplan i PIR-06-30. SKBdoc 1859805 ver 2.0, Svensk Kärnbränslehantering AB. (In Swedish.)

Sundberg J, Back P-E, Hellström G, 2005. Scale dependence and estimation of rock thermal conductivity. Analysis of upscaling, inverse thermal modelling and value of information with the Äspö HRL prototype repository as an example. SKB R-05-82, Svensk Kärnbränslehantering AB.

Sundberg J, Wrafter J, Back P-E, Ländell M, 2006. Thermal modelling. Preliminary site description Laxemar subarea – version 1.2 SKB R-06-13, Svensk Kärnbränslehantering AB.

Sundberg J, Wrafter J, Mossmark F, Sundberg A, 2007. Forsmark site investigation. Anisotropy of thermal properties in metagranite at Forsmark. Comparison between large-scale field measurements, small-scale field measurements and laboratory measurements. SKB P-07-194, Svensk Kärnbränslehantering AB.

Sundberg J, Wrafter J, Back P-E, Rosén L, 2008a. Thermal properties Laxemar. Site descriptive modelling SDM-Site Laxemar. SKB R-08-61, Svensk Kärnbränslehantering AB.

Sundberg J, Wrafter J, Ländell M, Back P-E, Rosén L, 2008b. Thermal properties Forsmark. Modelling stage 2.3. Complementary analysis and verification of the thermal bedrock model, stage 2.2. SKB R-08-65, Svensk Kärnbränslehantering AB.

Sundberg J, Back P-E, Ericsson L O, Wrafter J, 2009. Estimation of thermal conductivity and its spatial variability in igneous rocks from *in situ* density logging. International Journal of Rock Mechanics and Mining Sciences 40, 1023–1028.

Tahmasebi P, 2018. Multiple point statistics: a review. In Daya Sagar B S, Cheng Q, Agterberg F (eds). Handbook of mathematical geosciences. Cham: Springer International Publishing, 613–643.

Tahmasebi P, Hezarkhani A, Sahimi M, 2012. Multiple-point geostatistical modeling based on the cross-correlation functions. Computtational Geosciences 16, 779–797.

Tahmasebi P, Sahimi M, Caers J, 2014. MS-CCSIM: accelerating pattern-based geostatistical simulation of categorical variables using a multi-scale search in Fourier space. Computers & Geosciences 67, 75–88.

Zhang T, 2008. Incorporating geological conceptual models and interpretations into reservoir modeling using multiple-point geostatistics. Earth Science Frontiers 15, 26–35.

Appendix A

Upscaling with the SCA approach

Upscaling approaches

Isaaks and Srivastava (1989) suggest different methods for transforming one distribution to another, including the affine correction and the indirect lognormal correction. All these methods leave the mean of the distribution unchanged, while the variance is adjusted. However, such transformation methods are sufficient only for quantities that average arithmetically, such as ore grades or pollutant concentrations (Isaaks and Srivastava 1989). Thermal conductivity, on the other hand, is a transport property (like hydraulic conductivity) and does not average arithmetically. Therefore, other approaches are required. However, it should be noted that thermal conductivity data follow a much more symmetrical distribution than do hydraulic conductivity in rock, which is an advantage. Experience from the site descriptive modelling indicates that thermal conductivity follows a close to normal or slightly lognormal distribution for many rock types (Sundberg et al. 2007).

The SCA approach

Several methods have been developed for change of support related to hydrogeological applications. According to Gutjahr et al. (1978) and Dagan (1979), the effective hydraulic conductivity depends on whether the problem is 1D or 3D. Dagan (1979) presented the following general solution to the effective mean hydraulic conductivity (transformed to thermal conductivity; see Sundberg et al. (2005)):

$$\lambda_e = -(m-1) \cdot \lambda_x + \left(\int f(\lambda)\, d\lambda / (m-1) \cdot \lambda_x + \lambda \right)^{-1} \quad \text{(A-1)}$$

where m is the dimensionality (1, 2, or 3) of the problem and $f(\lambda)$ the frequency function of thermal conductivity λ. If λ_x is replaced by λ_{max} and λ_{min}, the result will be the same as Hashin's and Shtrikman's upper and lower bounds for an isotropic material (Hashin and Shtrikman 1962). If λ_x is replaced by λ_e, the self-consistent approximation (SCA) is obtained as follows:

$$\lambda_e = 1/m \cdot \left(\int f(\lambda)\, d\lambda / (m-1) \cdot \lambda_e + \lambda \right)^{-1} \quad \text{(A-2)}$$

For a lognormal distribution, the effective conductivity according to Equation A-2 for two dimensions (m = 2) coincides with the geometric mean. For three dimensions (m = 3) the effective conductivity is slightly higher. Equation A-2 is used to calculate the thermal conductivity from modal analysis by iteration (Sundberg 1988).

If the standard deviation (σ) of the natural logarithms of λ is small, then the effective thermal conductivity can be approximated as follows for a lognormal conductivity distribution (after Gutjahr et al. (1978)):

$$\text{2D: } \lambda_e = \lambda_G \quad \text{(A-3)}$$

$$\text{3D: } \lambda_e = \lambda_G \left[1 + \sigma^2/6 \right] \quad \text{(A-4)}$$

where λ_G is the geometric mean thermal conductivity.

Appendix B

Coordinate transformation for simulation of TRC lithology

Introduction

This appendix describes the coordinate transformation procedure for setting up a model for the stochastic simulation of geology in T-PROGS. As input data, T-PROGS uses transition probabilities for the principal direction of anisotropy. To facilitate realistic modelling results the model volume therefore needs to be orientated in a transformed coordinate system, obtained through rotation of the main principal directions (Figure B-1) in three steps: (x, y, z) → (x', y', z') → (x'', y'', z'') → (x''', y''', z'''); see Figures B-3 to B-5.

Geological information describing anisotropy is given as: (1) the trend and plunge of the mineral stretching orientation and (2) the strike and dip of the foliation. The orientation of rock units is a function of these. The theoretical description of the transformation to a coordinate system orientated in the principal direction of anisotropy as a function of the mineral stretching and foliation plane is given below.

Coordinate Transformation

The main principal directions are shown in Figure B-1.

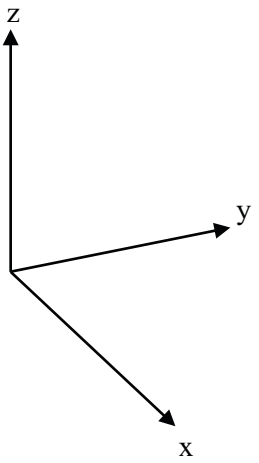


Figure B-1. *Main principal directions (x = easting direction, y = northing direction, z = elevation direction).*

Each rock mass of assumed stationary spatial statistical properties (domain) has a local coordinate system (x’’’, y’’’, z’’’) governed by the principal direction of anisotropy and an origo defined by minimum easting, minimum northing, and maximum elevation from positions in borehole records supplied. The local coordinate system is obtained through rotations in three steps, as described below. This results in a transformed coordinate system with main axis (x’’’) parallel to the principal direction of anisotropy, see Figure B-2.

Two unit vectors (x_1,y_1,z_1) and (x_2,y_2,z_2) can be defined as:

$x_1 = \cos \alpha_1 \cos \beta_1$

$y_1 = -\sin \alpha_1 \cos \beta_1$; (x_1, y_1, z_1) in principal direction of anisotropy

$z_1 = -\sin \beta_1$

and

$x_2 = \cos \alpha_2 \cos \beta_2$

$y_2 = -\sin \alpha_2 \cos \beta_2$; (x_2, y_2, z_2) in direction of foliation and TRC contacts

$z_2 = -\sin \beta_2$

where

α_1 = trend-90°; β_1 = plunge, α_2 = strike-90°, β_2 = dip.

The coordinate system is now rotated so that the x-axis points in the trend/plunge direction. The position (x_1,y_1,z_1) will then have the coordinates (1,0,0). The first rotation is thus the x-y plane α_1 degrees clockwise around the z-axis. The second rotation is the x’-z’ plane β_1 degrees clockwise around the y’-axis. The final transformation is to rotate the y’’-z’’ plane γ_1 degrees around the x’’-axis until $z_2 = 0$ for the position (x_2,y_2,z_2). The unit vectors (x_1,y_1,z_1) and (x_2,y_2,z_2) will then be in the new x’’’-y’’’ plane and z’’’ is perpendicular to this plane. The three rotations are defined as follows:

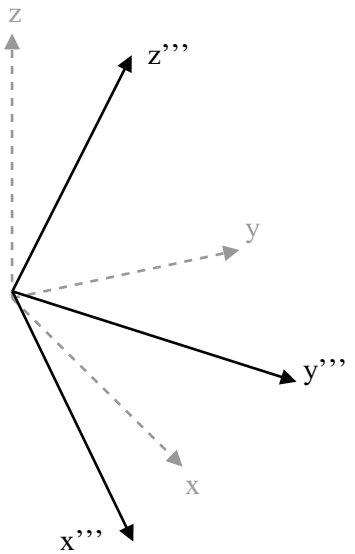


Figure B-2. *Principal directions of the local transformed coordinate system where x’’’ coincides with the principal direction of anisotropy.*

First rotation

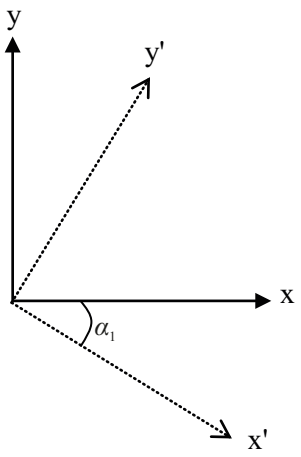


Figure B-3. *First rotation, α1 degrees around the z-axis.*

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} \cos\alpha_1 & -\sin\alpha_1 & 0 \\ \sin\alpha_1 & \cos\alpha_1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

Second rotation

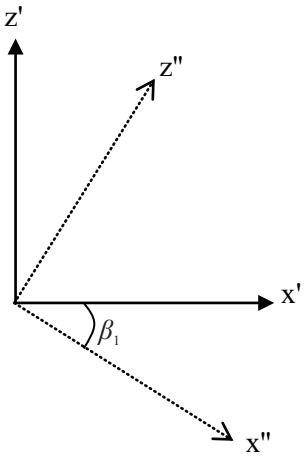


Figure B-4. *Second rotation, β1 degrees around the y'-axis.*

$$\begin{bmatrix} x'' \\ y'' \\ z'' \end{bmatrix} = \begin{bmatrix} \cos\beta_1 & 0 & -\sin\beta_1 \\ 0 & 1 & 0 \\ \sin\beta_1 & 0 & \cos\beta_1 \end{bmatrix} \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix}$$

Third rotation

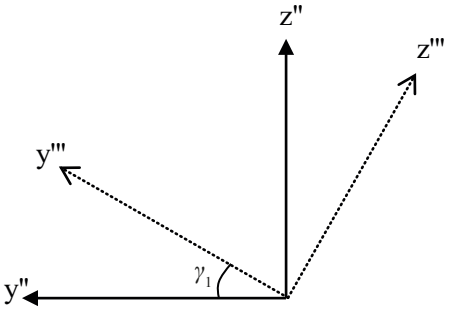


Figure B-5. *Third rotation, γ_1 degrees around the x"-axis.*

$$\begin{bmatrix} x''' \\ y''' \\ z''' \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\gamma_1 & \sin\gamma_1 \\ 0 & -\sin\gamma_1 & \cos\gamma_1 \end{bmatrix} \times \begin{bmatrix} x'' \\ y'' \\ z'' \end{bmatrix}$$

where the angle γ_1 is defined by:

$$\tan\gamma_1 = \frac{z_2''}{y_2''} \Leftrightarrow y_2'' \sin\gamma_1 = z_2'' \cos\gamma_1 \Leftrightarrow -y_2'' \sin\gamma_1 + z_2'' \cos\gamma_1 = 0$$

Verification

To verify that the unit vectors (x_1,y_1,z_1) and (x_2,y_2,z_2) are within the new x'''-y''' plane and that (x_1,y_1,z_1) will have the new coordinates (1,0,0) the following can be set up:

$$\begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} = \begin{bmatrix} \cos\alpha_1 \cos\beta_1 \\ -\sin\alpha_1 \cos\beta_1 \\ -\sin\beta_1 \end{bmatrix}$$

$$\begin{bmatrix} x_1' \\ y_1' \\ z_1' \end{bmatrix} = \begin{bmatrix} \cos\alpha_1 & -\sin\alpha_1 & 0 \\ \sin\alpha_1 & \cos\alpha_1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\alpha_1\cos\beta_1 \\ -\sin\alpha_1\cos\beta_1 \\ -\sin\beta_1 \end{bmatrix} = \begin{bmatrix} \cos^2\alpha_1\cos\beta_1 + \sin^2\alpha_1\cos\beta_1 \\ \sin\alpha_1\cos\alpha_1\cos\beta_1 - \cos\alpha_1\sin\alpha_1\cos\beta_1 \\ -\sin\beta_1 \end{bmatrix} = \begin{bmatrix} \cos\beta_1 \\ 0 \\ -\sin\beta_1 \end{bmatrix}$$

$$\begin{bmatrix} x_1'' \\ y_1'' \\ z_1'' \end{bmatrix} = \begin{bmatrix} \cos\beta_1 & 0 & -\sin\beta_1 \\ 0 & 1 & 0 \\ \sin\beta_1 & 0 & \cos\beta_1 \end{bmatrix} \begin{bmatrix} \cos\beta_1 \\ 0 \\ -\sin\beta_1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} x_1''' \\ y_1''' \\ z_1''' \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\gamma_1 & \sin\gamma_1 \\ 0 & -\sin\gamma_1 & \cos\gamma_1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$ Q.E.D.

$$\begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix} = \begin{bmatrix} \cos\alpha_2 \cos\beta_2 \\ -\sin\alpha_2 \cos\beta_2 \\ -\sin\beta_2 \end{bmatrix}$$

$$\begin{bmatrix} x_2' \\ y_2' \\ z_2' \end{bmatrix} = \begin{bmatrix} \cos\alpha_1 & -\sin\alpha_1 & 0 \\ \sin\alpha_1 & \cos\alpha_1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\alpha_2 \cos\beta_2 \\ -\sin\alpha_2 \cos\beta_2 \\ -\sin\beta_2 \end{bmatrix} = \begin{bmatrix} \cos\alpha_1 \cos\alpha_2 \cos\beta_2 + \sin\alpha_1 \sin\alpha_2 \cos\beta_2 \\ \sin\alpha_1 \cos\alpha_2 \cos\beta_2 - \cos\alpha_1 \sin\alpha_2 \cos\beta_2 \\ -\sin\beta_2 \end{bmatrix}$$

$$\begin{bmatrix} x_2'' \\ y_2'' \\ z_2'' \end{bmatrix} = \begin{bmatrix} \cos\beta_1 & 0 & -\sin\beta_1 \\ 0 & 1 & 0 \\ \sin\beta_1 & 0 & \cos\beta_1 \end{bmatrix} \begin{bmatrix} \cos\alpha_1 \cos\alpha_2 \cos\beta_2 + \sin\alpha_1 \sin\alpha_2 \cos\beta_2 \\ \sin\alpha_1 \cos\alpha_2 \cos\beta_2 - \cos\alpha_1 \sin\alpha_2 \cos\beta_2 \\ -\sin\beta_2 \end{bmatrix} =$$

$$\begin{bmatrix} \cos\beta_1 \cos\alpha_1 \cos\alpha_2 \cos\beta_2 + \cos\beta_1 \sin\alpha_1 \sin\alpha_2 \cos\beta_2 + \sin\beta_1 \sin\beta_2 \\ \sin\alpha_1 \cos\alpha_2 \cos\beta_2 - \cos\alpha_1 \sin\alpha_2 \cos\beta_2 \\ \sin\beta_1 \cos\alpha_1 \cos\alpha_2 \cos\beta_2 + \sin\beta_1 \sin\alpha_1 \sin\alpha_2 \cos\beta_2 - \cos\beta_1 \sin\beta_2 \end{bmatrix}$$

It can now be noted that

$$\tan\gamma_1 = \frac{\sin\beta_1 \cos\alpha_1 \cos\alpha_2 \cos\beta_2 + \sin\beta_1 \sin\alpha_1 \sin\alpha_2 \cos\beta_2 - \cos\beta_1 \sin\beta_2}{\sin\alpha_1 \cos\alpha_2 \cos\beta_2 - \cos\alpha_1 \sin\alpha_2 \cos\beta_2}$$

and that

$$\begin{bmatrix} x_2''' \\ y_2''' \\ z_2''' \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\gamma_1 & \sin\gamma_1 \\ 0 & -\sin\gamma_1 & \cos\gamma_1 \end{bmatrix} \begin{bmatrix} \cos\beta_1 \cos\alpha_1 \cos\alpha_2 \cos\beta_2 + \cos\beta_1 \sin\alpha_1 \sin\alpha_2 \cos\beta_2 + \sin\beta_1 \sin\beta_2 \\ \sin\alpha_1 \cos\alpha_2 \cos\beta_2 - \cos\alpha_1 \sin\alpha_2 \cos\beta_2 \\ \sin\beta_1 \cos\alpha_1 \cos\alpha_2 \cos\beta_2 + \sin\beta_1 \sin\alpha_1 \sin\alpha_2 \cos\beta_2 - \cos\beta_1 \sin\beta_2 \end{bmatrix}$$

$$= \begin{bmatrix} \cos\beta_1 \cos\alpha_1 \cos\alpha_2 \cos\beta_2 + \cos\beta_1 \sin\alpha_1 \sin\alpha_2 \cos\beta_2 + \sin\beta_1 \sin\beta_2 \\ \cos\gamma_1 \sin\alpha_1 \cos\alpha_2 \cos\beta_2 - \cos\gamma_1 \cos\alpha_1 \sin\alpha_2 \cos\beta_2 + \sin\gamma_1 \sin\beta_1 \cos\alpha_1 \cos\alpha_2 \cos\beta_2 + \\ + \sin\gamma_1 \sin\beta_1 \sin\alpha_1 \sin\alpha_2 \cos\beta_2 - \sin\gamma_1 \cos\beta_1 \sin\beta_2 \\ -\sin\gamma_1 \sin\alpha_1 \cos\alpha_2 \cos\beta_2 + \sin\gamma_1 \cos\alpha_1 \sin\alpha_2 \cos\beta_2 + \cos\gamma_1 \sin\beta_1 \cos\alpha_1 \cos\alpha_2 \cos\beta_2 + \\ + \cos\gamma_1 \sin\beta_1 \sin\alpha_1 \sin\alpha_2 \cos\beta_2 - \cos\gamma_1 \cos\beta_1 \sin\beta_2 \end{bmatrix}$$

$$z_2''' = -\sin\gamma_1(\sin\alpha_1 \cos\alpha_2 \cos\beta_2 - \cos\alpha_1 \sin\alpha_2 \cos\beta_2) + \cos\gamma_1(\sin\beta_1 \cos\alpha_1 \cos\alpha_2 \cos\beta_2 +$$
$$+ \sin\beta_1 \sin\alpha_1 \sin\alpha_2 \cos\beta_2 - \cos\beta_1 \sin\beta_2) =$$
$$= -\sin\gamma_1(\sin\alpha_1 \cos\alpha_2 \cos\beta_2 - \cos\alpha_1 \sin\alpha_2 \cos\beta_2) +$$
$$+ \cos\gamma_1 \tan\gamma_1(\sin\alpha_1 \cos\alpha_2 \cos\beta_2 - \cos\alpha_1 \sin\alpha_2 \cos\beta_2) = 0$$

Q.E.D.

From z1''' = z2''' = 0 it follows that both vectors (x_1,y_1,z_1) and (x_2,y_2,z_2) are within the new x'''-y''' plane.

Back transformation

When transforming back original coordinates, the procedure described above is performed in reverse order, α, β and γ angles represented with negative signs. The procedure is as described below:

Rotation 1:

$-\gamma_1$ degrees around the x’’’-axis transforms (x’’’,y’’’,z’’’) back to (x’’,y’’,z’’)

Rotation 2:

$-\beta_1$ degrees around the y’’-axis transforms (x’’,y’’,z’’) back to (x’,y’,z’)

Rotation 3:

$-\alpha_1$ degrees around the x’-axis transforms (x’,y’,z’) back to (x,y,z)

Appendix C

Geostatistical software and steps in the modelling

Table C-1. Specific capabilities of the investigated geostatistical software packages in relation to the steps in the thermal modelling strategy (unconditional stochastic simulation) as described in Figure 1-1 in Section 1.3 and in Back and Sundberg (2007) for SDM.

MODELLING STEP Number relates to Figure 1-1 (step in SDM strategy in parenthesis)	Description	Task/Program used in original strategy	Software						Comment on revised strategy for conditional simulation
			WinGslib	S-GeMS	Gstat	ISATIS	EVS	Own programs	
2. Preparation of thermal data (step 4 in SDM)		a) cell declustering (*declus*: GSLIB) b) histogram smoothing (*hs*: GSLIB) c) normal Score Transformation (*nscore*: GSLIB)	Yes	Yes	?	Yes	?		Depending on measurement support(s) and simulation method this step may require revision. Not a critical issue for software.
2. Change of support for thermal data (step 5 in SDM)	Change of support is required for the thermal data derived from small-scale measurements.	a) variogram modelling (*gamv* + *vmodel*, both GSLIB) b) LU decomposition (*lusim*: GSLIB) c) back transformation (*backtr*: GSLIB) d) *UPSCALE*	Yes, except *UPSCALE*	Yes, except *UPSCALE*	Yes (but different to GSLIB/S-GeMS), except *UPSCALE*	Yes (but different to GSLIB/S-GeMS), except *UPSCALE*	?	*UPSCALE*	The measurement scale (support) will determine if change of support is required. If required, a different approach than used in previous modelling will probably be applied. S-GeMS and ISATIS provide different algorithms for solving the change of support problem (by block simulation).
5. Statistical modelling of thermal properties (step 9 in SDM)	A spatial statistical thermal model is constructed for each TRC comprising model distribution and variogram. Upscaling of variogram required in this step.	a) distribution model from *UPSCALE* in step 5 above. b) variogram upscaling (*VARUP*) c) variogram model: (*vmodel*: GSLIB)	Yes, except VARUP	Yes, except VARUP	Yes, except VARUP	Yes (unsure about upscaling of variogram)	Yes (unsure about upscaling of variogram)	*VARUP*	Again, depending on measurement support(s), upscaling of the variogram may or may not be required. If needed our own program VARUP can be used. ISATIS provides an algorithm for modification of the variogram to block support. Neither WinGslib nor S-GeMS perform this step.

MODELLING STEP Number relates to Figure 1-1 (step in SDM strategy in parenthesis)	Description	Task/Program used in original strategy	Software						Comment on revised strategy for conditional simulation
			WinGslib	S-GeMS	Gstat	ISATIS	EVS	Own programs	
8. Stochastic simulation of thermal properties (step 10 in SDM)	Result is a number of equally probable realisations of thermal conductivity	Sequential Gaussian simulation (*sgsim*: GSLIB)	Yes	Yes	Yes	Yes	No		S-GeMS, Gstat and ISATIS software packages allow other simulation algorithms to be utilised if required.
9. Realisations of thermal properties (step 11 in SDM)	The realisations of TRCs (geology) and thermal conductivity are merged, i.e. each realisation of geology is filled with simulated thermal conductivity values.	*GEOMERGE*	No	No	No	?	?	*GEOMERGE*	Probably required by the revised strategy. Unclear if ISATIS has this functionality. The other software packages do not.
10. Evaluation of results – Upscaling (step 12 in SDM)	If the result is desired in a scale different from the simulation scale upscaling is required.	*UPSCALE*	No	No	No	?	?	*UPSCALE*	Probably required by the revised strategy. Unlikely that any of the software have this functionality.
10. Evaluation of results – Presentation (step 13 in SDM)	Histograms, statistical parameters, uncertainties, 2D and 3D illustrations based on a set of realisations of thermal conductivity.	a) histograms and statistics (*histplot*: GSLIB) b) percentiles (*THERMSTAT*) c) extract a representative subset of realisations (*REDUCE*)	a) Yes b) No c) No	a) Yes b) Yes c) No	?	a) Yes b) Yes c) Yes	?	*THERMSTAT, REDUCE*	In addition to algorithms used in site descriptive thermal modelling, the calculation of probabilities of exceeding a fixed cut-off (threshold value) may be required. This function can be performed by THERMSTAT (own program) but most likely also by S-GeMS and ISATIS.

Arkitektkopia AB, Bromma, 2023

SKB is responsible for managing spent nuclear fuel and radioactive waste produced by the Swedish nuclear power plants such that man and the environment are protected in the near and distant future.

skb.se