

Geostatistical Simulations of Geothermal Reservoirs: Two-And Multiple-Point Statistic Models

Ariel Vidal and Rosalind Archer

Department of Engineering Science, The University of Auckland, Private Bag 92019, Auckland 1142, New Zealand.

avid587@aucklanduni.ac.nz

Keywords: Reservoir modelling, geostatistics, uncertainty, multiple-point statistics

ABSTRACT

Reservoir engineers and geoscientists working on subsurface process are commonly faced with a lack of information related with geological structures, rock properties and sometimes a lack of comprehension of the underlying process and the physics that govern it, hence subsurface models are built from limited information with its own resolution and quality. In the light of these facts is that data uncertainty appears as one of the major obstacles in subsurface reservoir modelling, analysis and forecasting. The uncertainty of a numerical simulation of a natural reservoir will depend on the uncertainties associated with different reservoir parameters related to subsurface rock properties, their spatial distribution and state variables. This uncertainty is also directly linked with the availability of data and its spatial coverage, so a maximum benefit from data must be considered for any model under development.

Stochastic simulation has become a popular methodology in areas such as groundwater and hydrocarbon reservoir modelling for simulating unknowable reservoir properties and geometries, and quantifying reservoir heterogeneity. However, these techniques are not being commonly used in geothermal reservoir modelling. Additionally, stochastic approaches have an extensive risk analysis capability. In this sense, geostatistical based-models are a robust methodology to include available data into simulations in a stochastic framework, and also allow an easier link between simulations and geological knowledge. Variogram based-models, like kriging or sequential gaussian simulations, are based on two-point statistical models which don't necessarily allow a satisfactory reproduction of complex, curvilinear geological features. Multiple-point statistics (MPS) models are a novel technique of simulation which is based on patterns composed of several points, so a better reproduction of complex geometries may be achieved. Another key characteristic of MPS models is the absence of any variographic analysis since all the conditional probabilities are extracted from a Training Image (TI). A TI basically is a repository of the spatial patterns and their likelihood of occurrence in the model.

This work presents some of the theoretical aspects of MPS simulations compared to classical two-point models and presents some results for the simulation of a synthetic geothermal reservoir. The set of realizations generated in the previous step are used as input parameters for setting up a numerical flow model.

1. INTRODUCTION

The use of numerical simulations for understanding subsurface processes is quite a normal practice today mainly due to the solid theoretical background of the physics behind them and the extensive computational capabilities available. Sometimes, strategic decisions are based on the results of these numerical models so high confidence in their outputs is required. Mostly all numerical models of subsurface fluid flow processes are based on previous knowledge about the geological structure of the zone under study or on properties of the subsurface rocks through which fluids circulate. This knowledge derives from a variety of disciplines; specifically the conceptualization of geothermal systems must consider data coming from diverse fields of research in a multidisciplinary approach (geology, geophysics, geochemistry, drilling, well testing, etc.). No matter how much effort we place in this task it is almost impossible to get a whole picture, without uncertainty, of the reservoir structure and their properties in every single zone being modelled. This lack of geological knowledge is even more pronounced in fields under exploration or with a small number of deep wells, which normally are concentrated in production and reinjection zones. The inherent uncertainty about subsurface structure and rock properties must also be considered, and as far as possible quantified, for any interpretation or decision based on a numerical simulation relying on a geological model.

Stochastic simulations have become a popular methodology for simulating unknowable reservoir properties and geometries and quantifying reservoir heterogeneity in a variety of problems related with subsurface fluid flow in the oil industry, hydrogeology and CO₂ sequestration (Deutsch, 2002; Zinn and Harvey, 2003; Finsterle and Kowalsky, 2008; Strebelle and Levy, 2008; Hou et al., 2013). Additionally, some attempts have been published as an approach to the inverse problem constrained by geological uncertainties (Kitanidis, 1996; Larocque et al., 2003; Wellman et al., 2014). For its part, just in the last years some attention has been put in the geothermal community on the uncertain nature of geological and conceptual models and their impact on strategic decisions based on numerical modelling (Watanabe et al., 2010; Vogt et al., 2010, 2012, 2013). Geostatistics is a robust family of methodologies to include available data into different realizations in a stochastic framework, and, at the same time, allow an easier link between simulations and geological knowledge.

In this paper we will describe and compare some classical geostatistical techniques, like kriging, with a new family of algorithms termed multiple-point statistics models (MPS), which is a novel technique of simulation based on the reproduction of complex spatial patterns. Both families of simulations will be used on the modelling of a synthetic scenario and a methodology which includes these results on numerical flow models will be introduced.

2. GEOSTATISTICAL MODELLING – GENERAL CONCEPTS

As a starting point it is possible to mention that geostatistics aims at providing quantitative descriptions of natural variables distributed in space or in time and space and provides a methodology to quantify spatial uncertainty (Chiles, 2011). Matheron (1965) coined the term *regionalized variable* to designate a numerical function $z(x)$ depending on a continuous space index x and combining high irregularity of detail with spatial correlation. So, geostatistics can then be defined as “the application of probabilistic methods to regionalized variables”. It associates randomness with the regionalized variable itself, by using a *stochastic model* in which the regionalized variable is regarded as one among many possible *realizations* of a *random function*. A deterministic variable takes only one outcome; conversely a *random variable* (RV) is a variable that can take a series of possible outcomes, each with a certain probability or frequency of occurrence (Goovaerts, 1997). In the discrete case, to each outcome z_i is attached a probability value

$$p_i = \text{Prob} \{ Z = z_i \} \in [0, 1], \text{ with } \sum_{i=1}^n p_i = 1 \quad (1)$$

In the continuous case, the distribution of probability values can take the form of:

a cumulative distribution function (cdf) providing the probability for the RV not to exceed a given threshold value z :

$$F(z) = \text{Prob} \{ Z \leq z \} \in [0, 1] \quad (2)$$

a probability density function (pdf) defined as the derivative or the slope of the cdf at z -values where F is differentiable.

The key to a probabilistic interpretation of a variable z is the modelling of the distribution function, cdf or pdf, of the corresponding random variable Z . From the distribution of Z , specific moments can be derived such as the mean, the median, the variance, etc.

A random function (RF), denoted $Z(u)$, is a set of dependent random variables $\{Z(u), u \in S\}$, each marked with a coordinate vector u spanning a field or study area S . Just like a single random variable a random function Z is characterized by a distribution function and can be defined by a finite set of simulated realizations, a RF $Z(u)$ is displayed and used through its realizations $\{z^l(u), u \in S\}, l = 1, \dots, L$ (Goovaerts, 1997). A realization can be seen as a numerical model of the possible distribution in space of the z -values.

Following the previous definitions a conditional cumulative distribution function (ccdf) of a RV can be defined as the probability that the unknown $Z(u)$ be valued no greater than a threshold value z conditional to a data set $n(u)$.

$$F(u; z \mid n(u)) = \text{Prob} \{ Z(u) \leq z \mid n(u) \} \quad (3)$$

The last expression makes explicit the dependence of that cdf on the location u and the relation of that location with the $n(u)$ data retained. For practical purposes just the ccdf has interest and only those should be used for estimation (Remy et al., 2009).

A critical step in any geostatistical modelling and sequential simulation consists of estimating at each location u along the simulation path the conditional distribution given a specific conditioning data set $n(u)$. Two different approaches will be discussed:

- 1) Traditional two-point statistics, which consists of evaluating the relation of the single unknown $Z(u)$ with one datum $Z(u_\alpha)$ at a time. Such relation normally takes the form of a covariance, or *variogram*.
- 2) Multiple-point statistics (MPS), this allows considering the $n(u)$ data as a whole or a set of neighbour data. The necessary multiple-point statistics are taken from a training image (Guardiano and Srivastava, 1993; Strebelle, 2000, 2002).

This paper does not intend to present a deep review of geostatistics and its different algorithms. For a more detailed description please refer to (Goovaerts, 1997; Deutsch and Journel, 1998 and Chiles, 2011). For the purposes of this paper geostatistical algorithms have been separated in two classes: variogram based-models, including kriging and sequential gaussian simulations, which are based on two-point statistical models and which do not necessarily allow a satisfactory reproduction of complex, curvilinear geological features. Secondly, multiple-point statistics (MPS) models are a novel technique of simulation which is based on patterns composed of several points distributed in space, so a better reproduction of complex geometries may be achieved. Another key characteristic of MPS models is the absence of any variographic analysis since all the conditional probabilities are extracted from a training image (TI). A TI basically represents a repository of the spatial patterns and their likelihood of occurrence in the model (Boucher, 2013) and can be seen as an unconditional realization of a RF model $Z(u)$, that is a prior conceptual pattern of the spatial distribution of the variable (Remy et al., 2009). The TI could be based on previous knowledge derived from geological or geophysical data that must be considered for the construction of new simulations.

2.1 Variogram-based models

The most commonly used tool in geostatistics for a measurement of the relation between two random variables is the variogram and its equivalent the covariance. All two-point based methods (kriging, cokriging) rely on this indicator. Let us define $Z(u)$ as a stationary random function, and any two of its random variables $Z(u)$ and $Z(u + h)$, separated by a vector h . The relation between these two RVs could be characterized by any one of the following indicators:

- The covariance:

$$C(h) = E \{ [Z(u) - m] [Z(u + h) - m] \} \quad (4)$$

- The correlogram:

$$\rho(h) = C(h) / C(0) \in [-1, 1] \quad (5)$$

- The variogram:

$$\begin{aligned} 2\gamma(h) &= E \left\{ [Z(u + h) - Z(u)]^2 \right\} \\ &= 2 [C(h) - C(0)] \end{aligned} \quad (6)$$

where $m = E \{ Z(u) \}$, $C(0) = \sigma^2 = \text{Var} \{ Z(u) \}$ are the stationary statistics of the random variable.

A key concept in any geostatistical approach is the assumption of stationarity, which is defined as invariance by translation of spatial statistics (Goovaerts, 1997) and is required for the inference of a probabilistic model. This is a modelling decision, not a hypothesis or property of the data which could be tested (Journel, 1986). Stationarity can be assumed for a whole reservoir volume or just for one stratigraphic unit. Again, this will depend of the characteristics of each problem under study and the availability of data.

In practice an experimental variogram (or semivariogram) can be calculated from $n(h)$ pairs of data $z(u_\alpha + h), z(u_\alpha)$, $\alpha = 1, \dots, n(h)$ distant of h :

$$2\gamma^*(h) = \frac{1}{n(h)} \sum_{\alpha=1}^{n(h)} [z(u_\alpha + h) - z(u_\alpha)]^2 \quad (7)$$

The experimental variogram is an unbiased estimator of the theoretical variogram. After this value has been calculated for a variety of distances $|h|$, according to data availability, it is needed to extrapolate/interpolate the values of $\gamma^*(h)$ for all vectors h into a model $\gamma(h)$ available for all h . That model will represent the spatial variability of the random variable under study.

2.1.1 Local Estimation – Kriging

Estimation aims to predict the value of a regionalized variable over an unsampled site or volume of the field under study based on available data. Kriging is a popular tool of interpolation/extrapolation and has been historically at the source of acceptance of geostatistics (Krige, 1951; Journel and Huijbregts, 1978). It is a fundamental tool and is used in most geostatistical estimation and simulation algorithms. Kriging is in essence a generalized linear regression algorithm, extending the data-to-unknown correlation from data-to-data correlation through a non-diagonal kriging matrix (Remy et al., 2009). This allows the estimation of a regionalized variable by means of a weighted linear combination of available data.

It is beyond the scope of this paper to present a detailed description of Kriging and all its variations and for more details please refer to Goovaerts (1997), Deutsch and Journel (1998) and Chiles (2011). In the next paragraphs simple kriging (SK) will be presented as it will help us to introduce the main methodology behind kriging and how it is constructed. Simple kriging considers within a stationary field s the estimation of an unsampled value $z(u)$ from $n(u)$ neighbouring data values $z(u_\alpha)$, $\alpha = 1, \dots, n(u)$. The estimation $z_{SK}^*(u)$ is defined as:

$$z_{SK}^*(u) - m = \sum_{\alpha=1}^{n(u)} \lambda_\alpha^{SK}(u) [z(u_\alpha) - m] = \lambda^t \times D \quad (8)$$

where λ is the column vector of the $n(u)$ kriging weights $\lambda_\alpha^{SK}(u)$ and D is the column vector of the $n(u)$ residual data values $[z(u_\alpha) - m]$ built from the stationary and assumed known and constant mean value m . The consideration of residual variables rather than the original Z -variables ensures unbiasedness defined as zero expected error:

$$E \{ z_{SK}^*(u) - z(u) \} = 0 \quad (9)$$

Ideally, unbiasedness should be conditional to both data configuration and data values, that is:

$$E \{ z_{SK}^*(u) - z(u) \mid z(u_\alpha) = z(u_\alpha), \alpha = 1, \dots, n(u) \} = 0 \quad (10)$$

Unfortunately simple kriging, as most other linear estimators, ensures unbiasedness, not conditional unbiasedness, which is the source of many disappointments. On the other hand, a great advantage of kriging over traditional interpolators is that it is not convex, it means it need not to be valued in the interval of the data values available (Goovaerts, 1997).

The weights $\lambda_a^{SK}(u)$ in equation (8) can be calculated by a kriging system of linear equations built from a covariance model (Goovaerts, 1997):

$$K \cdot \lambda = k \quad (11)$$

where $k^t = [C(u - u_\alpha), \alpha = 1, \dots, n(u)]$ is the data-to-unknown row covariance vector, and $K = [C(u_\alpha - u_\beta), \alpha, \beta = 1, \dots, n(u)]$ is the data-to-data square covariance matrix; both matrices are built from the prior stationary covariance model:

$$C(h) = \text{Cov}\{Z(u), Z(u + h)\} = C(0) - \gamma(h) \quad (12)$$

$C(0) = \text{Var}\{Z(u)\}$ is the stationary variance, and $2\gamma(h) = \text{Var}\{Z(u) - Z(u + h)\}$ is the stationary variogram model for the regionalized variable.

According to Journel and Huijbregts (1978), the two most important contributions of kriging to estimation are:

1. The utilization of a variogram distance $\gamma(h)$ specific to the variable $Z(u)$ and the stationary zone s under study.
2. The consideration of the data-to-data covariance matrix K allows data declustering, which leads to giving less weight to redundant data within a cluster of data as opposed to isolated data.

The kriging variance is a secondary product derived from the kriging process, it is the expected squared error whose minimization led to the kriging system (12) and is defined as:

$$\sigma_{SK}^2(u) = \text{Var}\{Z(u) - Z_{SK}^*(u)\} = C(0) - \lambda^t \times k \quad (13)$$

This expression is independent of the used data, so, it only depends on the spatial configuration of the data set $n(u)$ available and the covariance/variogram model adopted. It is a valuable index for comparison of diverse data configuration, not a measure of estimation accuracy (Journel and Rossi, 1989).

Sometimes the restriction of stationarity could be relaxed and a locally variable mean or covariance model can be useful. Variants of the simple kriging algorithm have been developed to take into account such flexibility. In ordinary kriging (OK) the mean is locally re-estimated from local data inside the search neighbourhood, while the covariance model is kept stationary. This estimator is used more often than SK thanks to its robustness against local departures from the original decision of stationarity (Remy et al., 2009). Kriging with a trend (KT) uses a function for the mean which is a function of the coordinates and the unknown locally variable parameters in the estimation neighbourhood; generally it takes the form of a polynomial function of the coordinates. Kriging with a local varying mean (LVM) uses auxiliary information which provides the local varying mean at all locations. All these variants amount to relaxing the decision of stationarity initially necessary to define the random function model and infer its constitutive statistics. All previous kriging variants mentioned work with continuous data but sometimes it is required to model categorical variables such as lithology, alteration mineralogy or the presence of a continuous variable with a value above or below a limit threshold. For those cases indicator kriging has been developed which works with binary indicators of occurrence of an event (Goovaerts, 1997; Deutsch and Journel, 1998).

There is nothing in the theory of kriging that constrains the unknown $Z(u)$ and the data $Z(u_\alpha)$ to relate to the same attribute. Cokriging is the extension of the kriging estimator to estimation of one attribute using data related to other attributes (Goovaerts, 1997). An important assumption that cokriging requires is that random functions associated must be jointly stationary. The main complexity with this method is the necessity to infer and model jointly the cross-covariance/variogram models, which depends strongly on the total number of attributes. Independent of the cross-covariance/variogram approach used, cokriging shares all the strengths and limitations of kriging. All the weights and the error variance only depend on the geometric configuration of data, the location of the unsampled site and variograms involved, but don't depend on the values of data. Since kriging is a linear estimator, cokriging could ignore any non-linear relation between two different attributes.

2.1.2 Sequential Gaussian Simulations

The principle behind a geostatistical simulation consist in building a fictitious variable that reproduces the spatial continuity of the regionalized variable $z = \{z(x), x \in S\}$. This construction is based on the assumption that z is only one particular realization of a random function $Z = \{Z(x), x \in S\}$. The idea is generate a set of realizations of this random function. While kriging produces a smooth image of variability, different realizations of Z presents the same spatial variability than the actual regionalized variable, particularly the same histogram and variance. Sequential Gaussian Simulation (SGSIM; Goovaerts, 1997; Deutsch and Journel, 1998) uses the convenient properties of the Gaussian RF model. A Gaussian RF is fully characterized by its mean and covariance. In the SGSIM algorithm the mean and variance of the Gaussian distribution at any location along the simulation path is estimated

by the kriging estimate and the kriging variance. The value obtained from that distribution is then used as conditioning data. A previous transform of the original data into a Gaussian distribution could be necessary (Table 1).

Table1. Sequential Gaussian Simulation (SGSIM) algorithm (Remy et al., 2009)

```

1: Transform the data into normal score space.
2: Define a random path visiting each node of the grid
3: for Each node u along the path do
4:   Get the conditioning data consisting of neighbouring original hard data (n) and previously simulated values
5:   Estimate the local conditional cdf as a Gaussian distribution with mean given by kriging and variance by the kriging
   variance
6:   Draw a value from that Gaussian cdf and add the simulated value to the data set
7: end for
8: Repeat for another realization

```

One limitation of SGSIM, and all variogram-based simulations, is that it cannot deliver a simulation with definite patterns or structures involving more than two locations at a time. A simulation can be conditioned or non-conditioned. A non-conditioned simulation looks for building realizations of the random function that represents the regionalized variable, but without taking into account available data. So, although it reproduces the original variability of the regionalized variable (same histogram and variogram) it doesn't interpolate data. By contrast, a conditioned simulation uses the available data by mean of a conditioned distribution. A conditioned distribution describes the local uncertainty on the values of the regionalized variable taking into account the actual values present in the neighbourhood. For a detailed description of this, and others methods of simulation based on kriging please refer to Goovaerts (1997) or Deutsch and Journel (1998).

2.2 Multiple-point statistics models

Before the introduction of MPS models, two large families of simulation algorithms for facies modelling were available: pixel-variogram-based (Deutsch and Journel, 1998) and object-based or Boolean algorithms (Leeder, 1978; Mackey and Bridge, 1992). Pixel-based algorithms build the simulated realizations one pixel/node at a time, thus providing great flexibility for conditioning to data of diverse support volumes and diverse types. However, pixel-based algorithms have difficulty reproducing complex geometric shapes, particularly if simulation of these pixel values is constrained only by two-point statistics, such as variogram or covariance. Object-based algorithms build realizations by dropping items onto the simulation grid one object or pattern at a time; hence they can be fast and faithful to the geometry of the object. However, they are difficult to condition to local data of different support volumes, particularly when these data are of small support volume, numerous and of diverse type (Remy et al., 2009).

The MPS concept proposed by Journel (1993) combines the strengths of the previous two classes of simulation algorithms. It operates pixel-wise with the conditional probabilities for each pixel value being lifted as conditional proportions from a training image depicting the geometry and distribution of objects deemed to prevail in the actual field. In essence, the training image provides all the necessary multiple point covariance values. The decision of stationarity allows scanning a specific training image for replicates of the single multiple-point conditioning data event. By so doing one would sidestep any variogram/covariance modelling and also any kriging.

Currently there are several MPS algorithms being used in real applications which have shown reliable results e.g. SNESIM, FILTERSIM, DisPAT, IMPALA, Direct Sampling (Strebelle, 2002; Wu et al., 2008; Honarkhak, 2011; Straubhaar et al., 2010; Mariethoz et al., 2010). It is not the scope of this paper neither to present a comparison between these algorithms nor a detailed description, but instead to show some practical considerations for its implementation and to compare results with two-point models. The worked examples in this paper are based on the FILTERSIM algorithm implemented in the SGeMS software (Remy et al., 2009), but could have been done using any other.

2.2.1 Training Images

As was previously discussed in kriging-based algorithms the variability of the random variable is implicitly done when choosing a variogram model. In actual cases, and if the data are numerous enough, the variogram model can be directly inferred from measured data. The equivalence with MPS algorithms is the selection of a training image. A training image is a repository of the patterns and their respective likelihoods. A frequent pattern appears more often in the training image than a rare one. An important point is that the actual position of a pattern in the training image is (at first) irrelevant, what matters is its presence and its frequency (Boucher, 2013).

As was discussed by Boucher (2013) to correctly choose a training image it is important to understand why the actual positions of patterns in the training image do not matter. In general, MPS algorithms follow a sequential process; they simulate one, or a group of points at a time conditional to previously simulated points and hard data present on the grid. As a starting point, the grid only contains hard data and it is filled as the simulation progresses. At every uninformed grid node, the MPS algorithm will extract the existing data on the grid, both hard and previously simulated, given a search neighbourhood and then it will try to find some matching patterns in the training image. The data points found in this neighbourhood is called a data event. The definition of a good match between the data event and the patterns in the training images differs with each algorithm. Once matching patterns are found, the simulated values are chosen either by taking it from the best matching pattern or through a Monte Carlo simulation of the set of reasonably matching patterns. Some confusion with the actual role of a training image could be generated if it is seen as an analogue of the phenomenon under study and not as a repository of prevailing patterns as the algorithm sees it. Anyway, MPS algorithms are a great communication tool between geoscientists and modellers since the geological concepts are passed as an image instead of a mathematical function. In this sense the training image is a bridge between the geological knowledge about the reservoir and the numerical model.

2.2.2 FILTERSIM algorithm

Implemented by Wu et al. (2008) this algorithm can handle both categorical and continuous variables and proceeds in three major steps: (1) filter score calculation, (2) pattern classification and (3) pattern simulation. FILTERSIM first applies a set of filters to template data obtained from scanning the TI. This creates a set of filter score maps, with each training pattern represented by a vector of score values. This significantly reduces the pattern data dimension from the template size to a small number of filter scores. Next, similar training patterns are clustered into a so-called prototype class, each such class being identified by a point-wise averaged pattern. During the sequential simulation process, the conditioning data event is retrieved with a search template of same size than that used for scanning the TI. The prototype class closest to that conditioning data event is selected using some distance function. Finally, a training pattern is sampled from that pattern class, and pasted onto the simulation grid. The simulation is thus one of the patterns based on pattern similarity.

3. RESULTS AND DISCUSSION

3.1 Synthetic Reservoir: Model setup

With the purpose of testing some of the methodologies described in the previous chapter a simple synthetic model of a rectangular reservoir was used (Figure 1). This reservoir consists of three different horizontal layers, each one populated with random values of permeability taken from a normal distribution. The lower layer distribution is characterized by a mean of 30 mD with a variance of 5, the intermediate layer has a mean of 1 mD and a variance of 0.2 and finally the upper layer has a mean of 200 mD and a variance of 10. The modelled reservoir size is 25x20x4 km and the grid used for simulation is a regular one composed of 25x20x20 blocks in the x-y-z directions with a size of 400x400x200 m per block, comprising a total of 10,000 blocks. For calculation purposes all the permeability values are transformed to $\log_{10}(m^2)$.

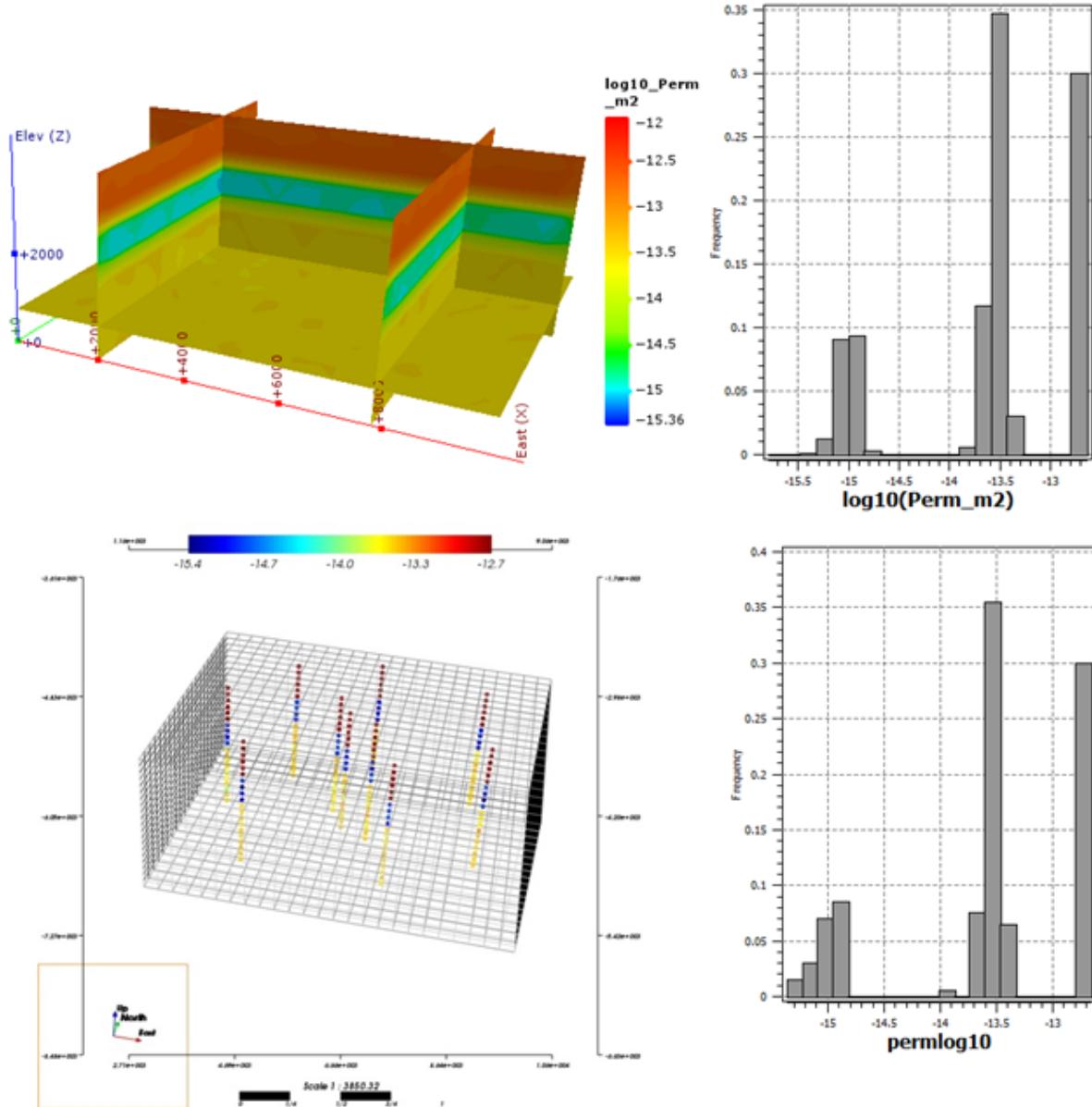


Figure 1: Architecture of the synthetic reservoir under study, geometry of the simulation grid and the location of 10 wells used as hard data. Histograms with the values of permeability of the whole reservoir and for the hard data are also showed (units in log(square-meters)).

3.2 Geostatistical Modelling

As shown in Figure 1 permeability values from the reference reservoir were selected for 10 vertical wells. The location of these wells was chosen to get a good cover of the reservoir. With these wells, used as hard data, a series of two-point-based statistics permeability models were constructed using the estimation technique of ordinary kriging and simulations based on sequential Gaussian models. MPS models based on training images were also created. Permeability is considered one of the main parameters that control fluid flow in subsurface systems; additionally, permeability is directly related to the geological structure of the reservoir and rock properties, so all the available geological knowledge may be taken into account for permeability modelling. In this paper we will treat permeability as a continuous variable but an alternative way could be first model geological bodies as categorical variables and subsequently populated each region with permeability values taken from a probability distribution or some geostatistical model. All the next simulations were generated using the SGeMS software (Remy et al., 2009).

3.2.1 Two-point models

Different kriging models were generated with the aim of getting a realistic representation of a horizontal stratified reservoir conditioned to well data. Ordinary kriging models were preferred over simple kriging models due to the ability of ordinary kriging of using a varying mean, calculated for every neighbourhood zone in the simulation pathway. Figure 2 show some obtained ordinary kriging-based models and their respective histograms.

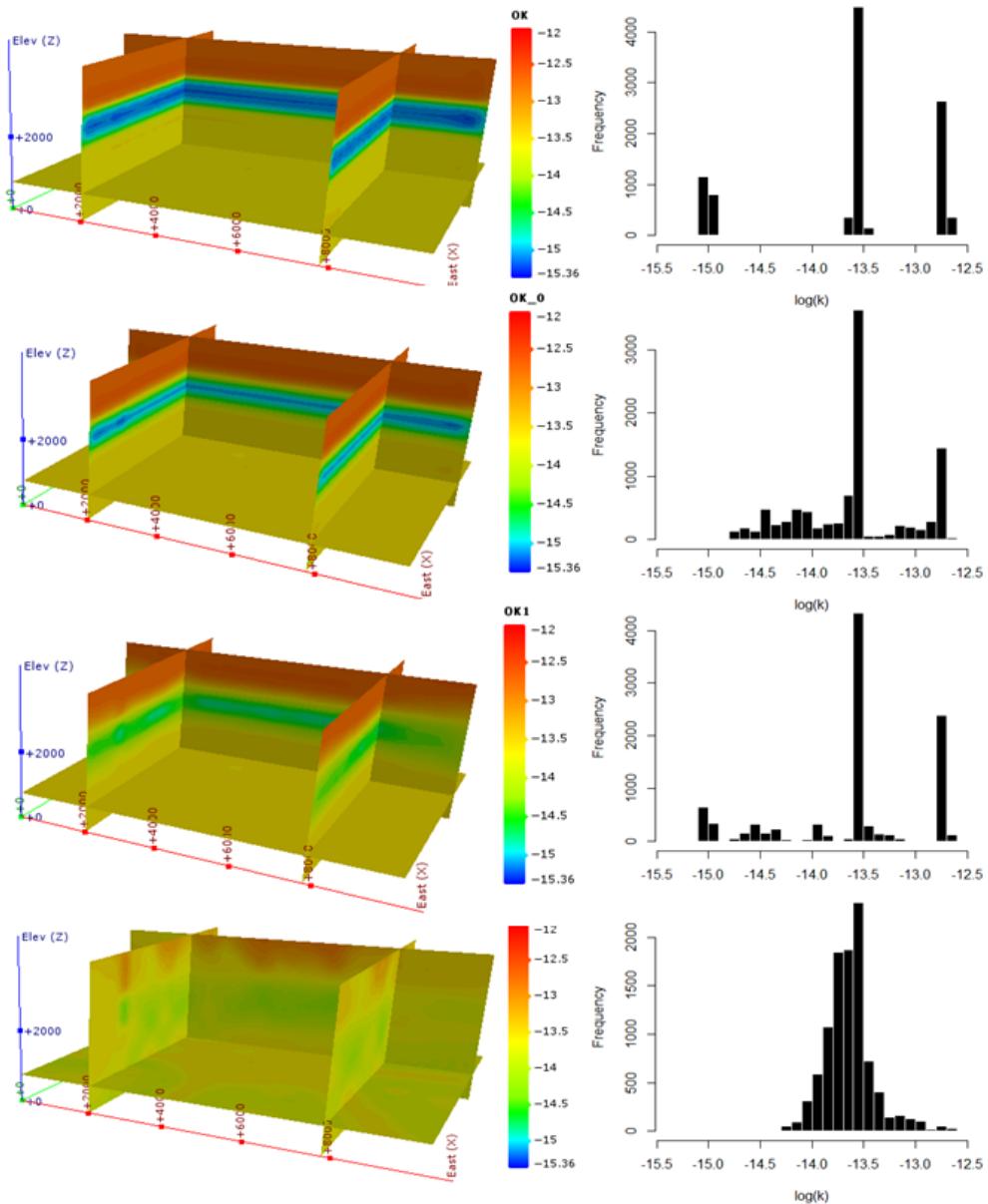


Figure 2: Four different ordinary kriging models of the permeability structure in the reference reservoir. Additionally, the histogram of the different models is showed (Please refer to text for more details and histograms in Figure 1 for a comparison with the reference model and hard data).

The variographic analysis (not shown here) revealed an anisotropic character of the data with specific directions showing lower variogram values and very low values for the nugget effect, which is coincident with the stratified structure of the reference

reservoir and the horizontal continuity of geological units. Interestingly, neither the choice of a particular kind of variogram model (spherical, power, Gaussian, etc.) nor the number of nested structures exerted too much influence on the generated models. For this case, the main parameters controlling this set of realizations were the dimensions of the search neighbourhood and the ranges for the variogram model, specifically in the vertical direction. Both of these sets of parameters were calibrated using geological restrictions such as the horizontal continuity of geological units and thickness of units based in well data. Models in Figure 2 share the same set of parameters with exception of the vertical range of the variogram which varies between 200 (upper left model) to 4000 m (lower right model). Beyond the graphical similarities that can be found between generated models and the reference reservoir (Figure 1), another important point for comparison is the resulting histogram of different models. As shown in Figure 1, the permeability histogram in well data is composed of three main peaks with a small dispersion in the lower values in agreement with the normal distributions from which the values were taken. However, it is important to keep in mind that the hard data-based histogram is not the same as the reference reservoir histogram shown in Figure 1, as some variability is lost with the selection of the ten wells used as hard data. In the kriging models presented in Figure 2 a progression between a distribution with 3 peaks towards a single-peak distribution can be seen as far as the vertical range increases in the models. In the same sense, it may be seen that the 3-peaks distribution, for the case with the lowest value of vertical range (upper left figure), shows a lower dispersion of values in agreement with the fact that kriging models tend to smooth the resulting models with a decrease in the number of extreme values and an overestimation of mean values.

Sequential Gaussian simulations (SGSIM) overcome some of the former obstacles and show a better match of histograms between well data, new realizations and the reference reservoir (Figure 3). Sequential Gaussian simulations produce more heterogeneous and variable geological scenarios which generate a better reflect of the nature of the original geological structure. In the same way that some of the kriging estimations, SGSIM simulations were able to reproduce the horizontal stratified nature of the reference reservoir. For this end, in the process of calibration of the model some geological knowledge (constraints) was imposed like long horizontal correlations and a constant thickness of geological units based on well data. Nevertheless, as with kriging models the SGSIM simulations fail to reproduce all the variability observed in the reference reservoir (Figure 1), which could be expected as these models are constrained by variability in hard data.

3.2.2 Multiple-point statistics (MPS) models

Something that all different MPS algorithms share is the use of a training image (TI) which contains all the needed statistics for the modelling process in the shape of patterns involving several points simultaneously, unlike classic variogram models based only in two point statistics. So, the choice of a proper training image becomes a central decision in the modelling process. A good practice is to use more than one training image, and hence adds more variability to the calculated response. With the purpose of testing the FILTERSIM algorithm different training images were used for the modelling of the permeability structure of the reference reservoir. Figure 4 shows four different realizations obtained using two different training images. The two lower realizations were obtained using the ordinary kriging model with a vertical range of 200m (upper left model in Figure 2) and the two upper realizations used one SGSIM simulation from Figure 3 as a training image. Both set of realizations were also conditioned by hard well data.

As can be seen in Figure 4, the choice of a particular training image exerts strong effects on the resulting realizations and so, in the aim of quantifying uncertainty. The selection of the ordinary kriging model with a vertical range of 200m, which is the best visual representation of the reference model as a TI produced new realizations that mimic almost completely the original TI, without any variability in the resulting realizations as can be seen in the generated histograms. More interestingly, the realizations based on a SGSIM simulation shows a wider variability among realizations keeping the horizontal structure of the original image. Unlike ordinary kriging and SGSIM simulations FILTERSIM requires a bigger number of input parameters which make the calibration process a harder and crucial point. In the examples aforementioned the dimensions of the Search Template and the Inner Patch proved to be determinant of the resulting outputs and in the capacity of translating geological restrictions into parameter space. Again, as with previous models, a reproduction of the original variability of the reference reservoir was not reached. More simulations based on different training images with more variability could be required.

4. FUTURE WORK

As was shown in previous sections geostatistical methods, two- and multiple-point statistics, were able to reproduce linear features expected in stratified reservoirs. However, more work is still needed for evaluate its capabilities in more complex scenarios, with non-continuous units and curvilinear patterns, as could be expected in actual cases. Variogram-based models are easy to implement and an extensive literature with examples from several fields of research is available. The derivation of a representative variogram in the absence of enough data still represents a major drawback, since geothermal reservoirs normally are developed in specific zones leaving vast areas without any information available. In this sense, the use of training images for extracting statistical information represents an appealing approach for overcoming the lack of hard data. Nevertheless, the selection and construction of a representative training image is not an easy task and examples for actual geothermal reservoirs training images are not yet available.

In the aforementioned examples a regular grid was used for the simulation process, but in some instances it may not be the optimal approach. Irregular or unstructured grids are normally used to represent more detailed information in areas of great interest or with a higher availability of hard data, as could represent a production or reinjection zone. Irregular grids on MPS simulations are a complex task as statistics/patterns extracted from training images are considered as representative of the whole domain (stationarity principle) and the algorithm doesn't know about different kinds of nodes/blocks with different shapes and sizes, and so different scales of variability could be mixed in the simulation process.

Secondary or soft data is normally available after an exploration program. Geophysical surveys, like magneto-tellurics (MT) or seismic imaging, offer a larger scale view of some reservoir properties and structures, which are complementary to some finer scale data as well logging or well drill cores. Inclusion of this big-scale data into simulations of finer-scale properties is something that must be taking into account.

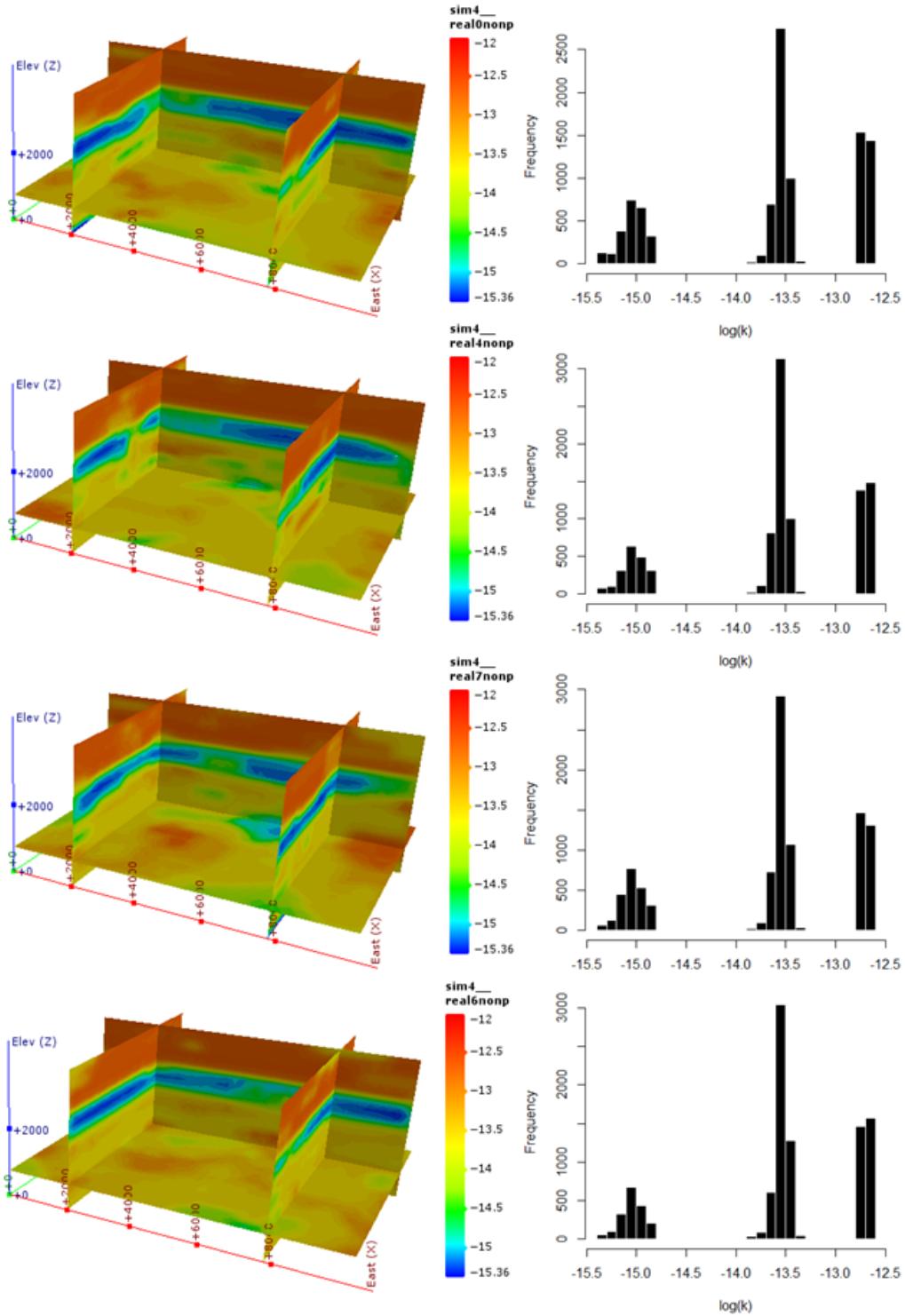


Figure 3: Four different sequential Gaussian realizations of the permeability structure using SGSIM algorithm and their respective histograms.

As have been mentioned, geostatistical models for reservoir simulation have been used mainly with the aim of quantifying geological uncertainty in natural reservoirs. But, an important point is that in a big number of applications the different geostatistical realizations are not an end-goal; instead, multiple realizations are normally used as inputs for flow numerical simulators. In this study the AUTOUGH2 numerical simulator (Yeh et al., 2012) is used for the coupled modelling of the heat and flow processes of the synthetic geothermal reservoir. In our current research we are working on a Monte Carlo approach of the inverse problem based on geostatistical simulations for geothermal flow models. We aim to simulate a geologically meaningful reservoir and at the same time generate flow model outputs that may be comparable to surface measurements. For example in Figures 5 and 6 some Temperature vs Depth profiles are shown. These profiles are based on flow simulations constructed with geostatistical realizations of the permeability field generated by using SGSIM and FILTERSIM algorithms and then compared with temperature profiles available for the ten wells in the reference model used as hard data for geostatistical modelling. As can be seen

in these figures, a close match between observations and predicted temperatures may be achieved for some models. An important point for future work is on the variability of the output of the simulator with changing input parameters, and how it is related with surface measurements for an inversion process as this data is not free of error and the actual value of some state variables may not be exactly known.

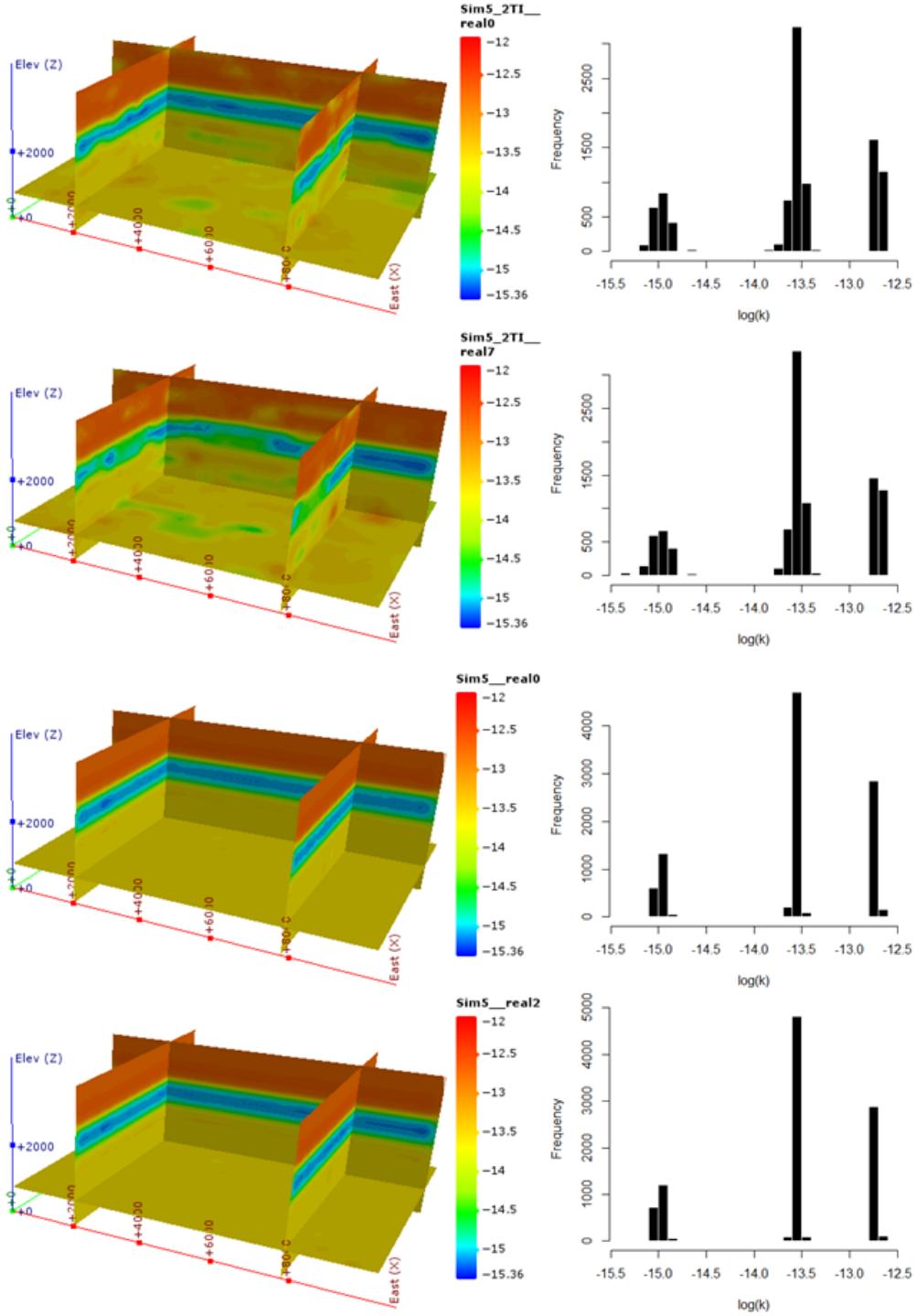


Figure 4: Four MPS-based realizations of the permeability structure using FILTERSIM algorithm and their respective histograms.

5. CONCLUSIONS

As in other disciplines based on subsurface fluid extraction, geothermal reservoir modelling requires to include a quantification of geological and model uncertainty. Reservoir simulation based on geostatistical methods helps to quantify and reduce the uncertainty based on available hard data and in numerical representations of the expected spatial variability (variogram or training image) by means of a set of equiprobable realizations of the variable under study (geological facies, permeability, porosity, etc.).

Two-point models offer an easy and reliable method for interpolation and simulation. One important limitation with this technique arises from data confidence, as much of the inferences and posterior analysis depends on the spatial distribution of hard data. In this sense interpretations on peripheral zones must be taken carefully. For its part, MPS models represent a novel approach which doesn't rely on variogram models based on the availability and distribution of hard data, but instead in a numerical representation of patterns deemed to prevail in the actual scenario. This numerical representation, or training image, also presents the advantage that is easy to understand by others, non-modeller, specialists as the spatial variability is expressed as an image instead of a mathematical function like the variogram. This training image could incorporate all the previous knowledge available for the phenomenon under study. Whether realizations are based on two- or multiple-point statistics the generated variability in each realization or among realizations is constrained by hard data and the variogram/TI selected as no more variability is added in the simulation process.

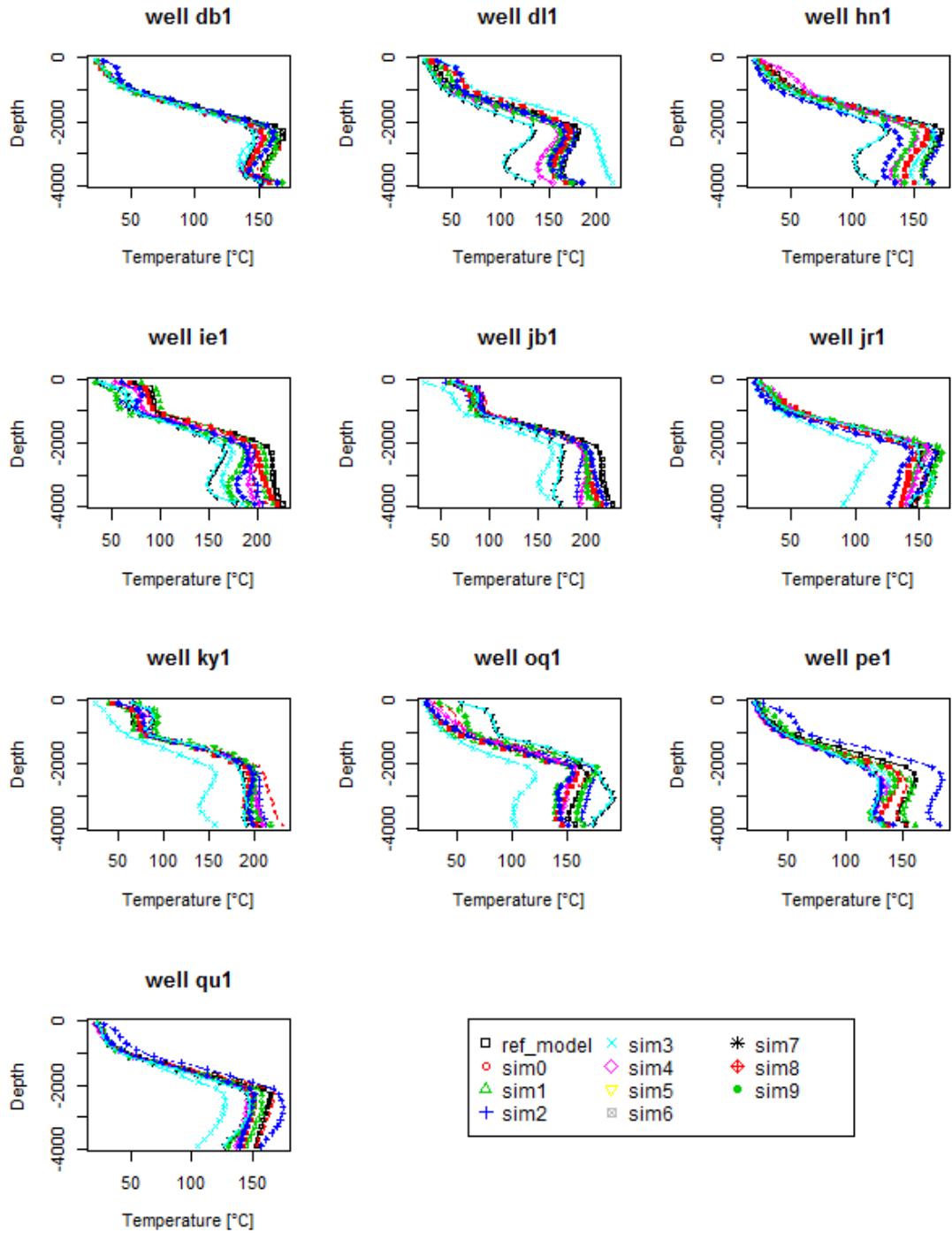


Figure 5: Well temperatures derived from flow models using SGSIM realizations of the permeability field as input parameters.

Numerical flow models based on geostatistical realizations represent a powerful technique to constraint numerical models with geological knowledge and then, through a stochastic process, select the right set of input parameters that mimic some surface measurements of state variables.

This study represents the first step of a research based on uncertainty quantification in geothermal modelling.

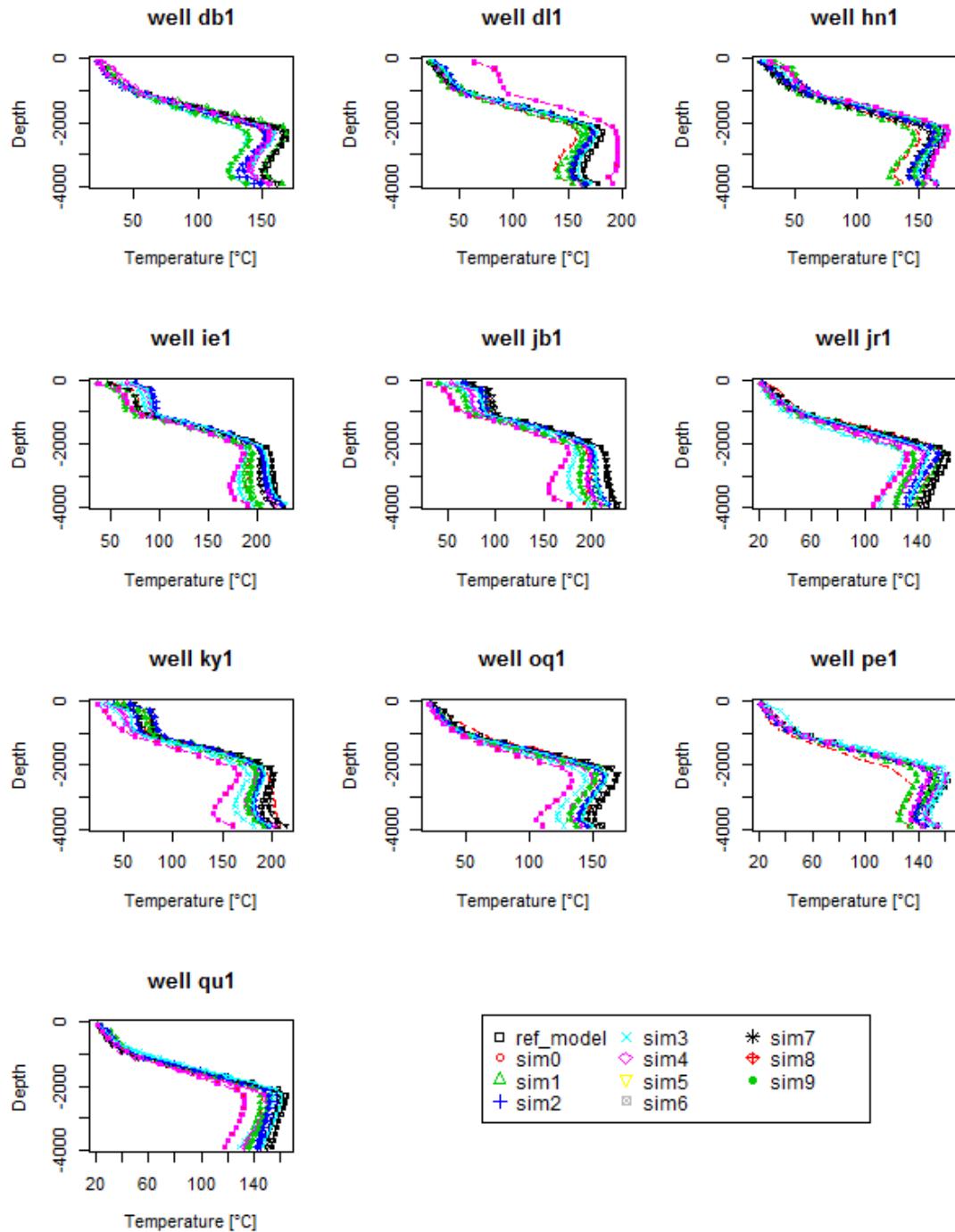


Figure 6: Well temperatures derived from flow models using FILTERSIM realizations of the permeability field as input parameters.

REFERENCES

Boucher, A.: Strategies for modeling with multiple-point simulation algorithms. In: Garner D, Thenin D, Deutsch CV (eds). *Closing the gap: advances in applied geomodeling for hydrocarbon reservoirs*. Canadian Society of Petroleum Geologists, Calgary (2013), pp 67–73.

Chiles, J., P.: *Geostatistics: modeling spatial uncertainty* / Jean-Paul Chiles: Oxford: Wiley-Blackwell (2011).

Deutsch, C.V.: *Geostatistical reservoir modeling*. Oxford University Press (2002), 376 p.

Deutsch, C., V., and Journel, A. G.: *GSLIB: Geostatistical Software Library and User's Guide*, 2nd edn, New York, Oxford University Press (1998).

Finsterle, S., and Kowalsky, M. B.: Joint hydrological-geophysical inversion for soil structure identification. *Vadose Zone Journal*, 7(1), (2008), 287-293.

Goovaerts, P.: *Geostatistics for Natural Resources Characterization*. Oxford University Press, New York (1997).

Guardiano, F., B., and Srivastava R., M.: Multivariate geostatistics: Beyond bivariate moments. In *Geostatistics Tróia '92*, Vol. 1, A. Soares, ed. Kluwer, Dordrecht (1993), pp. 133-144.

Honarkhah, M.: Stochastic simulation of patterns using distance-based pattern modeling, Ph.D. thesis, Stanford University, Stanford, CA (2011).

Hou, Z., Engel, D., Lin, G., Fang, Y., and Fang, Z.: An Uncertainty Quantification Framework for Studying the Effect of Spatial Heterogeneity in Reservoir Permeability on CO₂ Sequestration. *Mathematical Geosciences*, 45(7), (2013), 799-817.

Journal, A., G.: Geostatistics: models and tools for the earth sciences. *Mathematical Geology*, 18(1), (1986), 119–140.

Journal, A., G.: Geostatistics: Roadblocks and Challenges, in A. Soares (ed.), *Geostatistics-Tróia '92*, Vol. 1, Dordrecht, Kluwer Academic Publishers (1993), pp. 213–224.

Journal A., G., and Huijbregts C. J.: *Mining Geostatistics*, Academic Press, London (1978), 600 p.

Journal, A., G., and Rossi, M. E.: When do we need a trend model?. *Mathematical Geology*, 21(7), (1989), 715–739.

Kitanidis, P., K.: On the geostatistical approach to the inverse problem. *Advances in Water Resources*, 19(6), (1996), 333-342.

Krige, D., G.: A statistical approach to some mine valuations and allied problems at the Witwatersrand, M.S. thesis, University of Witwatersrand, South Africa (1951).

Larocque, M., Delay, F., and Banton, O.: A Comparison of Two Stochastic Inverse Methods in a Field-Scale Application. *Ground Water*, 41(1), (2003), 15-23.

Leeder, M.: A quantitative stratigraphic model for alluvium, with special reference to channel deposit density and interconnectedness. *Fluvial Sedimentology*, 5, (1978), 587–596.

Mackey, S., and Bridge, J.: A revised FORTRAN program to simulate alluvial stratigraphy, *Computer & Geoscience*, 18(2), (1992), 119–181.

Mariethoz, G., Renard, P., and Straubhaar, J.: Direct sampling method to perform multiple-point geostatistical simulations, *Water Resources Research*, 46(11), (2010).

Matheron, G.: *Les variables régionalisées et leur estimation. Une application de la théorie des fonctions aléatoires aux Sciences de la Nature*. Masson, Paris (1965).

Remy, N., Boucher, A. and Wu, J.: *Applied Geostatistics with SGeMS - A User's Guide*. Cambridge University Press (2009).

Straubhaar, J., Renard, P., Mariethoz, G., Froidevaux, R., and Besson, O.: An Improved Parallel Multiple-point Algorithm Using a List Approach, *Mathematical Geosciences*, 43(3), (2010), 305-328

Strebelle, S.: Sequential simulation drawing structures from training images, Ph.D. thesis, Stanford University, Stanford, CA (2000).

Strebelle, S.: Conditional simulation of complex geological structures using multiple-point statistics. *Mathematical Geology*, 34(1), (2002), 1–21.

Strebelle, S. and Levy, M.: Using multiple-point statistics to build geologically realistic reservoir models: The MPS/FDM workflow. *Special Publication, Geological Society of London*. 309, (2008), 67-74.

Vogt, C., Mottaghy, D., Wolf, A., Rath, V., Pechnig, R., and Clauser, C.: Reducing temperature uncertainties by stochastic geothermal reservoir modelling. *Geophysical Journal International*, 181(1), (2010), 321-333.

Vogt, C., Kosack, C., and Marquart, G.: Stochastic inversion of the tracer experiment of the enhanced geothermal system demonstration reservoir in Soultz-sous-Forêts — Revealing pathways and estimating permeability distribution. *Geothermics*, 42, (2012), 1-12.

Vogt, C., Iwanowski-Strahser, K., Marquart, G., Arnold, J., Mottaghy, D., Pechnig, R., and Clauser, C.: Modeling contribution to risk assessment of thermal production power for geothermal reservoirs. *Renewable Energy*, 53, (2013), 230-241.

Watanabe, N., Wenqing, W., McDermott, C. I., Taniguchi, T., and Kolditz, O.: Uncertainty analysis of thermo-hydro-mechanical coupled processes in heterogeneous porous media. *Computational Mechanics*, 45(4), (2010), 263-280.

Wellmann, J. F., Finsterle, S. and Croucher, A.: Integrating structural geological data into the inverse modelling framework of iTOUGH2. *Computers & Geosciences*, 65(0), (2014), 95-109.

Wu, J., Boucher, A., and Zhang, T.: A SGeMS code for pattern simulation of continuous and categorical variables: FILTERSIM. *Computers & Geosciences*, 34(12), (2008), 1863-1876.

Yeh, A., Croucher, A. and O'Sullivan, M. J.: Recent developments in the AUTOUGH2 simulator. *Proceedings, TOUGH Symposium 2012*, Lawrence Berkeley National Laboratory, Berkeley, California (2012, September 17-19).

Zinn, B., and Harvey, C.: When good statistical models of aquifer heterogeneity go bad: A comparison of flow, dispersion, and mass transfer in connected and multivariate Gaussian hydraulic conductivity fields. *Water Resources Research*, 39(3), (2003), 1051.