HESS Discussions, 10, 15219–15262, 2013

Challenges in conditioning a stochastic geological model of a heterogeneous glacial aquifer to a comprehensive soft dataset

Correspondence to: J. Koch (juko@geus.dk)

Author comment to review #2: Yoram Rubin.

We would like to thank Yoram Rubin for his review to our manuscript. He brought thoughtful comments forward concerning the statistical theory underlying an ensemble of stochastically generated realizations. We believe that a broader description of the TProGS's principals and algorithms can clarify the reviewer's concerns. We do not find it necessary to include further TProGS theory in our manuscript, because it is already intensively elaborated in the existing literatures. However, we have included more detailed descriptions on the theory in Appendix I to this reply. We appreciate the suggested references, e.g. concept of anchors, and have added them to the revised manuscript in both the introduction and the discussion section. We hope the adjustments made during the revision can help to improve the scientific quality of the manuscript.

> *This paper reports on several challenges addressed by the authors while pursuing this study. The paper is interesting to read, primarily because it raises several interesting issues and challenges.*
> *The main concern I have is that the approach pursued here, although referred to as stochastic, does not present a complete and sound statistical theory. This shows up in different ways:*
>
> *• The approach presented in this paper combines statistical concepts (such as transition probabilities) with multiple decisions that are based on aesthetics and subjective judgment calls. The judgment calls made by the authors are not translated into statistical rules. As a result, the ensemble of realizations that are generated does not constitute a statistically-meaningful ensemble. Rather, it represents a bunch of (presumably and subjectively) reasonably-looking realizations. This creates a confusing mix of statistical models and art. What constitutes "reasonable" and less "reasonable" is not clear. It is a judgment call, and hence it crosses the boundary between science and art. Without a statistically-meaningful ensemble, meaning, without a complete representation of high and low-probability realizations, it becomes impossible to quantifying predictions. Quantifying predictions is the main goal of such studies, and when that is not made possible, what is the point?*

We disagree with some of the comments above; the underlying geostatistical methods in TProGS, namely sequential indicator simulation (SIS) and simulated quenching will ensure that the generated ensemble is statistically meaningful (Please see Appendix I for more details). The two inter-dependent steps will assure that each individual realization honors the conditioning data and the defined model of spatial variability. The model of spatial variability, which is used in the local indicator cokriging estimate in the SIS step, is represented by a continuous 3D Markov model, derived from the computed transition probabilities. A consistent/statistically meaningful TProGS ensemble will reflect the defined model of spatial variability (mean length and proportion) and the conditioning data (hard and soft).

We agree to the comment that "subjective judgment calls are not translated into statistical rules". Subjectivity is part of our study at multiple points: Basing the horizontal and vertical model of spatial variability on different data types, the delineation of the entire geological structure (based on geological interpretations) or the way hardness is assigned to the soft data. Another subjective decision that has to be acknowledged is the definition of the Markov Chain model in TProGS (mean length and proportion). We strongly disagree that these subjective judgment calls are rather art than science; on the contrary they are the strength of this study. Especially the latter, because it allows the incorporation of geological interpretations to the modeling process. The subjective judgment calls do not undermine the meaningfulness of the simulated ensemble. The decision to out-thin the conditioning dataset is driven by the identified problem of overconditioning. We experienced that overconditioning affects both the SIS and the simulated quenching, as spatially correlated data are introduced into the modeling process. Hence, overconditioning hinders the self-consistency of TProGS, because the simulated ensemble does not represent the uncertainties given in the soft dataset which is used for constraining the simulation. Neither it embodies the defined model of spatial variability; e.g. the mean length is undersimulated. Before we can apply the ensemble for subsequent flow modelling we have to assure self-consistency, thus comparing of input and output of TProGS. An ensemble is only fit for quantifying predictive uncertainty if all input data and TProGS parameters are considered accordingly in the simulation. The five validation criteria which are defined in our study test the self-consistency; the criteria go far beyond visual comparisons and subjective judgment calls. We regard this as a systematically approach that address the consistency between input and output and significantly helped us at finding the most appropriate sampling distance and method to reduce the effect of overconditioing.

o Along this line, I should refer to the authors' comment on p. 15221 (line 20) concerning ".. equally likely probable realizations..". First, if the authors believe the realizations are equally-probable, they need to show how these probabilities are computed. Next, I believe that in the absence of a multi-point statistical model, as is the case here, one cannot really use the term "equally-likely". At best these are realizations that were produced following a similar procedure (on p. 15221 line 21, the authors refer to the realizations they produce as "plausible", which I think is more appropriate) . Lastly, the authors should explain why they see merit in producing equally-probable realizations. I believe that in application we are interested in a variety of probabilities, high-and low-probabilities: we know how to prepare for high-probability events. It is the low-probability (less plausible?) events that lead to disasters.

We believe that stochastic modeling of the hydrofacies distribution is an appropriate way of describing geological heterogeneity and geological uncertainty. The stochastic methods can ensure local accuracy through conditioning to observations and "fill the gaps" between observation points, by generating geological features that represent an underlying model of spatial variability. The model of spatial variability is based on a combination of data (transition probabilities) and geological expert knowledge. This in addition with conditioning data, either soft or hard, represents the best combined knowledge of the geological system. Our interests in simulating equally plausible realisations are that they are different and hence will lead to different flow paths when the different geologies subsequently will be used for hydrological flow modelling. Hence we want to to explore how uncertainties in the geological structures lead to uncertainty in groundwater flow paths.

We agree with the reviewer that "equally plausible" is better suited than "equally probable" and the manuscript will be changed accordingly.

• No attempt is reported in this study on computing or employing statistics beyond single- and two-point correlations. This opens up the question of the implications of neglecting higher-order statistics.

We pursued that thought which was also uttered by another reviewer and dedicated it a section in the discussion: “Choice of geostatistical method” (Appendix III; lines 206-238). Here we mostly concentrate on the comparison between two-point and multi-point statistics.

> *• On p. 15220 (line 20), the authors mention five criteria used for validation. This statement and the strategy it represent raise several challenges:*
> *o First, validation is not possible in groundwater applications. This was pointed out in a paper by Naomi Oreskes (http://www.likbez.com/AV/CS/Pre01-oreskes.pdf) and echoed by many scientists ever since. The best the authors could say on this context is that they examined their realizations from five different perspectives.*

The reviewer gives a thoughtful point in his comment above. We are aware of the discussion following the papers by Konokow and Bredehoeft (1992) and Oreskes (1994), and we are aware of the different schools of thoughts on this fundamental issue. We have contributed to that discussion in Refsgaard and Henriksen (2004). We agree that a model cannot be validated if validation is understood in a universal sense as Oreskes does. We have argued in Refsgaard and Henriksen (2004) that the term validation, in line with the scientific philosophy of Popper (1959), can be used in a conditional manner, where the validation in restricted to given site specific locations, given types of applicability and given accuracy. We acknowledge that these details were not included in the manuscript. We have therefore dedicated a few lines on that in the revised introduction (Appendix II; lines 193-196)

> *o The five criteria used for validation represent information that was known a-priori. One could argue that as such, these criteria could have affected the judgment calls made by the authors along the way, and hence they do not represent independent and unbiased evaluation criteria.*

We agree that the five performance criteria are a–priori knowledge and that they are not all truly independent from the simulation. However, we believe that they are unbiased. The sand proportion and mean length define the Markov model of spatial variability and the facies probability distribution and the facies probability – resistivity bias are derived from the soft conditioning dataset. Only the connectivity is not used as an input to TProGS and is therefore the most independent criterion. Nevertheless the reference connectivity is deducted from the categorized SkyTEM dataset. A more correct approach might be to use only borehole data for conditioning and for the definition of the Markov model and validate the ensemble of realizations against SkyTEM data, or vice versa. However, the split sample test in section 5.2 shows that conditioning purely to borehole- or geophysical data gives biased results, because the two data types are different in terms of support scale and density. Therefore only a true integration of the two data types can yield satisfying results. The moving sampling of the conditioning dataset generates multiple independent conditioning datasets. Thus it allows comparing an ensemble conditioning to a decimated dataset to be validated against the entire dataset.

> *o In my opinion, the information represented by the 5 evaluation criteria should be used to construct statistical priors in a Bayesian sense. In this context, there are several papers I should mention, including cf., Woodbury and Rubin, 2000, Hou and Rubin, 2005.*

The study by Hou and Rubin (2005) nicely shows how Bayesian principles can be integrated into inverse modeling of hydraulic parameters in the vadose zone to simulate soil moisture. However, we are not using a Bayesian framework and we do not see how this could help us achieving the objectives of our study

A few more comments in other directions:
1. On p. 15222, line 24, the authors state that "Until now there are no published studies on the incorporation of a comprehensive and continuous soft conditioning datasets..". To my knowledge this is not accurate. As an example, I should mention the concept of anchors, discussed in Rubin et al., (2010) which can be used to condition on so-called "soft" data. Anchors can be used to represent data of all sorts using statistical distributions.

We agree with the reviewer and therefore have updated the introduction and the discussion section accordingly. Collocated cokriging/cosimulation techniques (Babak and Deutsch, 2009) probability aggregation (Mariethoz et al., 2009b) and truncated plurigaussian simulations (Mariethoz et al., 2009a) are briefly presented in the introduction. The MAD (method of anchored distributions) is very promising, because of its flexibility towards inverse modeling of spatial random fields. The associated anchors contain local information in form of distributions of the target variable. The sand probability, derived from the SkyTEM data, could be represented by anchors as "local effects", thus conditioning. Global parameters, like sand proportion and correlation length are also considered in the MAD approach (Rubin et al., 2010). (Appendix II; lines 138-142)

2. On p. 15223 line 4 (and on multiple other locations) the authors refer to "overconditioning" [sic]. The authors do not define what they mean with this term, and my interpretation of it as that it means some sort of challenge related to highly-dense data used for conditioning. This is supported by a statement made on p. 15234 line 13 that "The observed problem of overconditioning is caused by spatially correlated data which are incorporated into the modeling process". The relationship with spatially-correlated data is correct, in my opinion, only that this is an avoidable problem, because it is an outcome of the authors' decision to use kriged data for conditioning. This decision needs to be revisited. Kriging produces point estimates that are optimal in some sense. It does not create, and is not intended to create, fields that are defined the geostatistical models that are used for kriging. Kriging is a smooth interpolator, not a random field generator. Kriging eliminates important variability, and cannot be used for conditioning (see Rubin, 2003, p. 60 and p.71, discussion on estimation vs. simulation). Kriging produces unrealistic and inflated correlation lengths. These correlation lengths do not represent spatial variability of the geophysical variables, because they are obtained from a graphic representation of kriging estimates. This is possibly the reason for the effect referred to by the authors as overconditioning. As an alternative, I would suggest to the authors to generate realizations of the geophysical data for conditioning. I would possibly represent the geophysical data using a series of anchors (each defining a statistical distribution (Rubin et al., 2010).Then, for simulation, I would suggest using a nested structure approach (see Maxwell et al., 2000) which involves (a) generating random skytem field realizations, followed by (b) using each of these realizations as a starting point instead of the kriged estimates. An alternative would be to convert the geophysical data into anchor representation of the facies at selected location, and use these anchors as a starting point for simulation. (Rubin et al., 2010; Murakami, 2010)

We define "overconditioing" as an effect triggered by dense and spatially correlated conditioning data that produces an altered picture of observable uncertainties. We believe that this problem is essentially related to soft conditioning, where the soft data/uncertainty are amplified by the stochastic simulation. A definition is added to the introduction section (Appendix II; lines157-158).

We agree to the reviewer's general concerns about kriging; a smooth interpolator that reduces variability and thus inflates correlation lengths. In the discussion section we advise to only use the direct sounding data instead of the fully distributed 3D kriging map of resistivity. This would reduce the effect of overconditioing and discard any additional uncertainties originating from kriging. The original sounding points are very dense, with over 100,000 sounding points taken at flightlines 50m to 100m apart. These data would be more than sufficient as soft conditioning dataset for a simulation on a 20mx20mx2m grid domain, because all major features are well covered and delineated by the individual sounding data already. The kriging interpolation method forces local accuracy at locations with sounding data and interpolates missing data based on variogram. As the sounding density is rather high, giving multiple points per correlation length, we don't expect a significant inflation of length scales by the kriging. Since the sounding points along the flight lines are 15 m apart, and the distance between the flight lines is around 50 meters, we don't expect that there is too much of spatial smoothing when resampling the data on a 20×20×2 m grid.

We do not see the main cause of overconditioing in the choice of interpolation method, although kriging might boost the overconditioing problem. In fact similar problems are expected to occur when another interpolator is used. The correlation length of SkyTEM data is found to be 500m in the lateral direction. Data on the 20m grid size will therefore always be heavily correlated, independent from the way they are interpolated. The reviewer suggests using SkyTEM realizations, generated by the method of anchored distributions as soft conditioning in TProGS. We agree that this would be a more sophisticated way to use the SkyTEM data than simple kriging. However as a result, the generated fields will still be spatially correlated even if the suggested method was carried out using the same grid size.

> *3. Single point cross-correlations: show examples, explain how done. Explain how the discrepancy between the scale of the borehole measurements on one hand and the scale of the geophysical data as accounted for. In MAD (Rubin et al., 2010) a case is made that anchors could be used to account for that (a scaling model is needed).*

The resistivity data are gridded on a 20m x 20m x 2m domain and get integrated with the borehole data.

The general support scale of a SkyTEM observation increases with depth, as the penetration of the subsurface is shaped as a cone, with 15-20m on the surface to a larger support scale in larger penetration depths (at 30m depth the lateral support size will be in the range of 50m). The glacial sequence which defines the model domain is between 10m and 40m thick. The variable support size makes the analysis difficult. Another difficulty is that SkyTEM inclines to overlook thin sand features, which are especially present in the vertical direction. However, the vertical support scale of the SkyTEM measurement device will not vary as much with penetration depth as the lateral support scale. Therefore, one can assume that the 20m x 20m x 2m is a suitable grid size for near surface resistivity values. However, one has to consider that with increasing depth the support size may grow larger than the defined grid size, where the lateral direction is more affected than the vertical direction.

We agree that the support scale of the borehole data and the geophysical data are different and that this adds to the uncertainty when they are compared in the histogram curve (Figure 3 in the manuscript). The uncertainty on the relationship between the resistivity (geophysical data) and the lithology (borehole data) as reflected by the not very steep curve in Figure 3 is, as discussed in the manuscript, originating from many sources (uncertainty in resistivity data, uncertainty in borehole data, uncertainty in the relationship between resistivity and lithology and the mismatch of support scales). Instead of focusing on each of these uncertainties we have chosen to lump all these uncertainties into one relationship (Figure 3) which we believe is suitable for the purpose of our study. We believe that our methodology has a general applicability, but acknowledge that the specific relationship in Figure 3 does not have generic validity and should not be used at other locations.

A few lines connected to this topic are added to the reviewed discussion section (Appendix III; lines 267-272).

> *Additionally:*
> *• The use of statistical correlations to relate between the geophysical and geological attributes pursued in this study is reported very scantly. It is not clear how good or bad these correlations are, and this needs to be discussed.*

Please find a more detailed explanation of the data integration:
The SkyTEM data, measured in resistivity, needs to be linked to a facies type in order to be used in a stochastic geological simulation. The procedure to connect SkyTEM data to facies information obtained from borehole data used in this study is the histogram probability matching method (HPMM), presented by (He et al., 2013b). The idea behind that method is the assumption that the probability of facies occurrence is positively correlated to the occurrence of resistivity over a certain range. Therefore the continuous SkyTEM data are classified into bins with a defined range. A fixed vertical discretization is defined representing the scale of the assessed heterogeneity in vertical direction, 2m in this case. The geophysical data is then compared with the categorical borehole data at collocated cells and the data pairs are grouped after the chosen bin width (10Ωm). Thus each bin contains a number of data pairs and a facies fraction of the categorized borehole data can be calculated respectively. The fraction can be plotted as bars and polynomial curve fitting allows to translate any resistivity observation into a probability of facies occurrence (Figure 3, section 3.2 in manuscript).

> *• Please discuss and demonstrate the implications of using single- rather than multi-point statistics.*

We agree to this point and dedicated the section “Choice of geostatistical method” (Appendix III; lines 206-238) to discuss the advantages and disadvantages of using two-rather multi-point statistics. This will be placed in the discussion section of the revised manuscript.

> *• There is an extensive body of work on the use of petrophysical models for relating the geophysical and geological attributes (Rubin et al., 1992; Mavko et al., 2009). It would be interesting to know if the statistical correlations provided better results compared to physically-based, statistical models.*

Mavko et al. (2009) present several physically-based empirical models to derive geological parameters, e.g. empirical models to transform seismic velocity into porosity. Opposed, the method used in this study (histogram probability matching method) is purely based on spatial correlations and is not build up on physical relationships. The main limitation is that it is site specific and cannot be applied to other catchments. On the other hand is was never our intention to create a general histogram with universal applicability. It is acknowledged by Jorgensen et al. (2003) that TEM methods provide the same amount of structure detail as the more costly seismic methods. However, to our knowledge no general empirical relationship between resistivity measured by an airborne based TEM method and hydrological attributes has been studied. First attempts with low maturity are given in the PhD thesis by Vest (2003): http://www.hgg.geo.au.dk/rapporter/speciale_phd/AVC_phd.pdf.

> *4. On several occasions in this paper the authors point out that conditioning is producing a trends (e.g., p. 15220 lines 19-20). Stated differently, trends are identified in application. This is a problem because the existence of a deterministic trend indicates that the trend was not removed prior to computing the two-point*

correlations, which violates the requirement for stationarity (Rubin, 2003, p. 58). When a trend exists, it must be accounted for a-priori, and not as an outcome.

Stationarity cannot be attested to the geology of the study site. The geophysical dataset indicates larger sand features in the south of the model domain, which is manifested in an increase of proportion and mean length. This is indicated at several occasions in the manuscript. We agree that non-stationarity trend must be accounted for a-priori, this was tested in our study following the method presented by Seifert and Jensen (1999). Here the TProGS simulation domain is subdivided into smaller sub-domains that fulfill the stationarity requirements and each is simulated individually. After the sub-domains are merged together; hard conditioning along the seamlines should ensure good connectivity between the sub-domains. We subdivided the TProGS domain into three sub-domains; one southern and two northern (north-east and north-west), each was equipped with individual Markov Chain models and hard conditioning was placed along the seamlines. After merging the three sub-domains together one could identify bad connectivity between the sub-domains, which raises questions on the method presented by Seifert and Jensen (1999) or the distinct geological heterogeneity at the study site poses extra challenges to a subdivided simulation approach. However, we claim that comprehensive soft conditioning, where the soft dataset represents the observable trends, can account for any non-stationarity issues. Please find the Master's thesis by Koch (2013) for a more detailed description of the matter (http://nitrat.dk/xpdf/final_juko.pdf, Part 1, pages 20-25).

5. On page 15220, line 13, the authors identify "the incorporation of two distinct datasources [sic] into the stochastic modeling.....sparse borehole data and abundant SkyTEM data" as the "novelty of this study". In making this statement, the authors should recognize the large body of published work that did precisely the same, including: Rubin et al., 1992, Copty and Rubin, 1995, Hubbard and Rubin, 2000, Hubbard et al., 2005, Hou and Rubin, 2005, Kolwaksy et al., (2001, 2004).

We appreciate the reviewer's comment and will correct the given lines respectively. The suggested literature is very relevant and was considered during the revision. We included Hubbard and Rubin (2000) in the revised introduction (Appendix II; lines 166-174).

Appendix I

Extended TProGS-Theory

(Carle, 1996a) composed the second version of the TProGS manual which gives a good insight into the geostatistical software. Further the TProGS workflow design is well presented in (Carle et al., 1998). In indicator geostatistics the indicator variable Ik(x) defines the presence or absence of a category k (e.g. a facies) at a location x.

$$\mathrm{I_k(x)} = \{^{\text{1,if k occurs at x}}_{\text{o,if otherwise}} \qquad \mathrm{k} = 1, \ldots, \mathrm{K} \qquad \text{Eq.1}$$

where K is the number of categories. In a stationary system can the heterogeneity of the category distribution then be modeled by univariate (mean length or proportion) and bivariate spatial statistics (e.g. indicator cross-variogram or transition probability). The most fundamental feature of TProGS is the transition probability tjk (h), which is a measure of spatial variability:

$$\mathrm{t_{jk}(h)} = \mathrm{P\{k\ occurs\ at\ x + h|j\ occurs\ at\ x\}} \qquad \text{Eq.2}$$

where k and j refer to the defined categories, x represents a spatial location vector and h is a separation vector (lag). The definition is possible to be put into words:
'Given that a facies j is present at location x, what is the probability that another (or the same) facies occurs at location x+h.' (Carle, 1996b).
Before computing the transition probabilities and conducting a Markov Chain analysis categories have to be defined representing the relevant facies. Here it is crucial to keep the number of categories minimal, but at the same time, include the required complexity of the model. Markov Chain analysis is regarded as a powerful stochastic model for indicator variables. It assumes that spatial occurrence depends entirely on the nearest data. For the Markov Chain analysis are transition probabilities calculated for strike, dip and vertical direction at specified lag intervals. This denominates a K x K matrix:

$$T(h_\Phi) = \begin{bmatrix} t_{1,1}h_\Phi & \cdots & t_{1,K}h_\Phi \\ \vdots & \ddots & \vdots \\ t_{K,1}h_\Phi & \cdots & t_{K,K}h_\Phi \end{bmatrix} \qquad \text{Eq.3}$$

where the transition probability tjk for all possible transitions (K*K) is denoted at each specific lag (h) in direction Φ. The diagonal entries represent the auto transitions and the off diagonal entries represent the cross transitions. Another important part of the Markov Chain analysis is the transition rate matrix RΦ.

$$R_\Phi = \begin{bmatrix} r_{1,1,\Phi} & \cdots & r_{1,K,\Phi} \\ \vdots & \ddots & \vdots \\ r_{K,1,\Phi} & \cdots & r_{K,K,\Phi} \end{bmatrix} \qquad \text{Eq.4}$$

where the entries rij,Φ describe the rate of change from category i to j in direction Φ, conditional to the presence of i. The transition rate corresponds to the slope of the transition probability as it approaches a lag of zero. When subsurface geology is modelled it is important to parameterize fundamental observable attributes in the model: Volumetric fractions (proportions), mean lengths (thickness and lateral extent) and (asymmetric) juxtapositional tendencies. These attributes can be conveyed by data analysis and geological interpretations and are considered for conceptualization of the facies manifestation, as they control the Markov Chain model. This enables the user to select plausible parameters when defining the model of spatial variability.

The facies proportion (pk) is related to the asymptotic limit of the transition probability by

$$\lim_{h_\Phi \to \infty} t_{j,k}(h_\Phi) = p_k \qquad \text{Eq.5}$$

In a stationary system the proportions are equal for strike dip and vertical direction. The asymptotic limits of each entry in the transition rate matrix will thus correspond to the proportion of the corresponding facies. The mean length of facies j in a given direction Φ can be defined as total length of j along Φ divided by the number of embedded occurrences of j along Φ. The mean length equates to the diagonal transition rate by

$$r_{j,j,\Phi} = -\frac{1}{\overline{\overline{L}}_{j,\Phi}} \qquad \text{Eq.6}$$

The mean length is indicated on a plot of auto-transition probabilities as the intersection of the tangent at the origin with the x-axis. Statistically speaking is the mean length an indicator for the correlation length and for the length scale of the facies.

TProGS computes the realizations of the geology in two uncoupled, but mutual depended steps. An initial configuration of facies distribution is produced by the SIS algorithm (Deutsch and Journel, 1992). Secondly, the initial configuration is reshuffled by the simulation quenching optimization algorithm (Deutsch and Cockerham, 1994). The SIS algorithm incorporates a transition probability based indicator cokriging estimate in order to approximate the local conditional probabilities from data at each simulation cell. This step ensures that the conditioning data is fully honored. The local transition probabilities are incorporated from an interpolated 3D Markov Chain model. A random path is chosen along all unsimulated cells. At each of these cells a local conditional probability distribution is computed by cokriging values of neighboring conditioned data and already simulated cells. The node gets assigned to a category by choosing a random number in respect to the probability distribution. Then the simulation gets updated and repeated until all unsimulated cells are assigned to a category.

The simulated quenching is incorporated to improve the agreement between simulated and modelled transition probabilities. The algorithm utilizes the initial configuration from the SIS and improves it to ensure a better agreement with the defined model of spatial variability, by minimizing the objective function O:

$$\mathrm{O} = \sum_{\mathrm{l}=1}^{\mathrm{M}} \sum_{\mathrm{j}=1}^{\mathrm{K}} \sum_{\mathrm{k}=1}^{\mathrm{K}} (\mathrm{t_{jk}}(\mathrm{h_l})_{\mathrm{SIM}} - \mathrm{t_{jk}}(\mathrm{h_l})_{\mathrm{MOD}})^2 \qquad \text{Eq.8}$$

where hl represent l = 1,…,M specified lag vectors and ’SIM’ and ’MOD’ specify the simulated and modeled transition probabilities, respectively. The algorithm is usually implemented in an iterative manor. In each iteration step, a random path checks at each un- or soft-conditioned cell if a perturbation in category would reduce O; if so, the change is accepted. (Carle, 1997) points put that the quality of the realizations strongly depends on the initial configuration and the number and direction of the quenching lags. Best performance was attested to a simulation with 99 anisotropic lags and an initial configuration computed by cokriging of the nearest 12 data. However the four nearest quenching lags in combination with the three nearest data in the cokriging interpolation gives already good results for a 2D application.

Appendix II

Introduction

Constraints in accurate and realistic solute transport modeling in hydrogeology are caused by the difficulty of characterizing the geological structure. The subsurface heterogeneity heavily influences the distribution of contaminants in the groundwater system. The scale of heterogeneity is often smaller than the data availability (e.g. borehole spacing). In traditional hydrogeological studies, one geological model is built based on the best comprehensive knowledge from often sparse borehole data and subjective interpretations. This can lead to alleged correct results, for instance when addressing the water balance on catchment scale, but will often prove to be inadequate for simulations beyond general flows and heads, e.g. contaminant transport modeling. Therefore, it is proposed by numerous studies that the uncertainty on the geological conceptualization is crucial when assessing uncertainties on flow paths (Neuman, 2003; Bredehoeft, 2005; Hojberg and Refsgaard, 2005; Troldborg et al., 2007; Seifert et al., 2008). One of the strategies often recommended for characterizing geological uncertainty and assessing its impact on hydrological predictive uncertainty is the use of multiple geological models (Renard, 2007; Refsgaard et al., 2012).

In this respect geostatistical tools such as two-point statistics e.g. TProGS (Carle and Fogg, 1996; Carle et al., 1998) and multipoint statistics (MPS) (Strebelle, 2002; Caers and Zhang, 2002; Caers, 2003; Journel, 2004) are powerful tools as they enable the generation of multiple equally plausible realizations of geological facies structure. This study targets the realistic description of heterogeneity in a geological model by introducing diverse data into the stochastic modeling process to generate a set of equally plausible realizations of the subsurface using geostatistics (Strebelle, 2002; Refsgaard et al., 2006).

In geostatistical applications field observations can constrain the simulation as soft or hard conditioning. “Hard conditioning” forces the realizations to honor data completely whereas “soft conditioning” honors the data partly with respect to the uncertainty of the observation (Falivene et al., 2007). This feature is essential because it enables the user to associate uncertainties to the conditioning dataset that can be of either subjective or objective nature. Incorporating a comprehensive and continuous soft conditioning datasets to a stochastic simulation such as TProGS is challenging. Alabert (1987) published an early study on the implications of using sparse soft conditioning data to a stochastic simulation. The analysis shows that soft conditioning significantly increases the quality of the realizations. The same was also observed by McKenna and Poeter (1995) where soft data from geophysical measurements could significantly improve the geostatistical simulation. In the past years, highly sophisticated geophysical methods and advanced computational power allow stochastic simulations that are conditioned to a vast auxiliary dataset. This poses new challenges to the data handling and to the simulation techniques.

Chugunova and Hu (2008) present a study where continuous auxiliary data is introduced directly, without classification to a MPS simulation. MPS requires a site specific training image that represents the geological structure accordingly, which is often the main source of uncertainty in MPS simulations. The above mentioned MPS studies conduct mostly 2D simulations, partly on synthetic data. The training image is the backbone of the MPS method and it has been acknowledged by dell'Arciprete et al. (2012) and He et al. (2013a) that reliable 3D training images are difficult to acquire.

Alternative methods to integrate vast auxiliary information (e.g. geophysics) into the modeling process and at the same time force local accuracy are collocated cokriging or cosimulation techniques (Babak and Deutsch, 2009). Here a linear relationship between the auxiliary variable and the target variable is built in a model of cross covariance. The essentially linear relationship is often too restrictive and does not represent the complex physical processes. Mariethoz et al. (2009b) present a prospective method that extends the collocated simulation method by using a model of spatial variability of the target variable and a joint probability density distribution to depict the conditional distribution of the target variable and the auxiliary variable at any location.

The method of anchored distributions (MAD) (Rubin et al., 2010) is a suitable approach for the inverse modeling of spatial random fields with conditioning to local auxiliary information. Structural parameters such as global trends and geostatistical attributes are considered in a conditional simulation. The conditioning is undertaken by anchored distributions which statistically represent the relationship between any data and the target variable.

The truncated plurigaussian simulation method (Mariethoz et al., 2009a) generates a Gaussian field for the target and the auxiliary variable using variogram statistics. These Gaussian fields are truncated to produce categorical variables that represent the hydrofacies. The truncation is controlled by threshold values that can be defined in a “lithotype rule” that represents the general geological concept. It is a very flexible method, because conceptual understandings are easily incorporated, but non-stationarity and especially directional depended lithotype rules are difficult to incorporate.

TProGS is a well-established stochastic modeling tool for 3D applications and it has been successfully applied to simulate highly heterogeneous subsurface systems by constraining the simulation to borehole data (Carle et al., 1998; Fleckenstein et al., 2006). Weissmann et al. (1999), Weissmann and Fogg (1999) and Ye and Khaleel (2008) use additional spatial information obtained from soil surveys, sequence stratigraphy and soil moisture, respectively for accessing the complex lateral sedimentary variability and thus improving the quality of the model in terms of spatial variability. It has not been tested whether TProGS, is capable of handling abundant soft conditioning data. Moreover, the risk that a cell-by-cell soft constraining may cause an overconditioning of the simulation has not been fully investigated. Overconditioing is defined by the authors as an effect triggered by dense and spatial correlated conditioning data that produces an altered picture of observable uncertainties. Therefore the self-consistency of the stochastic simulation is questioned, because soft constraining should be treated accordingly during the simulation.

Recent studies by Lee et al. (2007) and dell'Arciprete et al. (2012) highlight that TProGS is compatible with other geostatistical methods like, multi-point statistics, sequential Gaussian simulations and variogram statistics (Gringarten and Deutsch, 2001). The distinct strength of TProGS is the simple and direct incorporation of explicit facies manifestations like mean length, proportion and (asymmetric) juxtapositional tendencies of the facies.

Geophysical datasets are valuable information in many hydrogeological investigations. It can considerably improve the conceptual understanding of a facies or hydraulic conductivity distribution and identify non-stationary trends. However, the integration of geophysical data and lithological borehole descriptions is often difficult. A recent study by Emery and Parra (2013) presents an approach to combine borehole data and seismic measurements in a geostatistical simulation to generate multiple realizations of porosity. Hubbard and Rubin (2000) review three methods that allow hydrogeological parameter estimation based on geophysical data. The three methods link seismic, ground penetrating radar (GPR) and tomographic data with sparse borehole data to support the hydrogeological description of the study site. Our study integrates high resolution airborne geophysical data with borehole data to build a probabilistic classification of the subsurface at site. The geophysical data are collected by SkyTEM, an airborne transient electromagnetic method (TEM) that has been used extensively in Denmark for the purpose of groundwater mapping (Christiansen and Christensen, 2003; Jorgensen et al., 2003b; Sorensen and Auken, 2004; Auken et al., 2009). This study utilizes a method that translates SkyTEM observation data into facies probability which enables associating the geophysical data with softness, according to the level of uncertainty. Very few studies have integrated high resolution airborne geophysical data in a stochastic modeling process (Gunnink and Siemon, 2009; He et al., 2013a).

Most stochastic studies only make relatively simple validations of how well the simulations are able to reproduce known geostatistical properties. Carle (1997) and Carle et al. (1998) investigate the goodness of fit between the simulated and the defined spatial variability. The geobody connectivity is used by dell'Arciprete et al. (2012) to compare results originated from two- and multipoint geostatistics. Chugunova and Hu (2008) make a simple visual comparison between the auxiliary variable fracture density and stochastic realizations of the simulated fracture media. A more advanced validation is conducted in Mariethoz et al. (2009b) where simulated variograms and histograms are compared with reference data for the simulation of synthetic examples. In spite of these few studies that have addressed the validation issue, no guidance on which performance criteria to use and how to conduct a systematical validation of a stochastic simulation has been reported so far.

It should be noted that we in line with Refsgaard and Henriksen (2004) do not use the term model validation in a universal manner, but in a site specific context where a model validation is limited to the variables for which it has been tested as well as to the level of accuracy obtained during the validation tests.

The objectives of this study are: (1) to set up TProGS for a study site based on lithological borehole data and high resolution airborne geophysical data and investigate the effect of the two distinct conditioning datasets to the simulation; (2) to assess the problem of overconditioing in a stochastic simulation; (3) to ensure that non-stationary trends are simulated accordingly by TProGS; and (4) to identify and test a set of performance criteria for stochastic simulations that allow the validation against geostatistical properties derived from field data. The results of the present study are intended for application in a hydrological modeling context (Refsgaard et al., 2014).

Appendix III

Discussion

Choice of geostatistical method

The choice of the stochastic method for this study is application driven (Refsgaard et al., 2014). In the Norsminde catchment, it is evident from both borehole and geophysical data that the glacial sequence contains till clay and sand lenses distributed in extremely irregular patterns that are non-stationary. Without dense conditioning data the heterogeneous and non-stationary structures will not be simulated correctly. TProGS among other two-point statistics enables soft conditioning, where the soft information represents the associated level of uncertainty of an observation. The other distinct strength of TProGS is the easy incorporation of observable geological attributes when defining the Markov Chain models. In multi-point statistics (MPS) the definition of a reliable 3D training image is challenging, especially when simulating extremely irregular patterns (Honarkhah and Caers, 2012). Defining a MPS training image for the Norsminde catchment is peculiar, because it could only be based on interpreted SkyTEM data; with inflated length scales in the vertical direction. This makes the model of spatial variability in TProGS more reliable and objective, because it is based on measured transition probabilities and not on an interpreted training image. Further the transition probabilities are based on the data type we trust best: borehole data in the vertical- and SkyTEM data in the horizontal direction. In this study it is of spatial interest to correctly simulate the vertical transition probabilities in order to subsequently simulate the flow paths in the shallow groundwater system most accurately. This requires a detailed description of the spatial variability of the vertical direction, with indication of thin sand lenses, only provided by borehole data.

However, MPS is broadly applied in 2D and 3D applications: The snesim algorithm (Liu, 2006) combines object-based and pixel-based methods in the general MPS framework, to enforce spatial pattern reproduction and local conditioning, respectively. It was successfully applied by He et al. (2013a) in a 3D application. Another promising approach is given by Chugunova and Hu (2008), where MPS is tested on non-stationary 2D structures, by continuous soft conditioning to a secondary variable. Here two training images from the geological structure and from the secondary variable are joint in the simulation.

Many promising geostatistical methods have advanced to incorporate auxiliary information to constrain the simulated target variable: Truncated plurigaussian simulation (Mariethoz et al., 2009a), collocated simulation with probability aggregation (Mariethoz et al., 2009b). Most of them are only tested on 2D applications partly with synthetic data. This present study uses TProGS as the geostatistical tool, because of its reliable model of spatial variability and it is well established in 3D applications with sparse conditioning data. The application of vast soft conditioning data to a TProGS simulation gives valuable information on how such data can influence the stochastic simulation results.

TProGS setup

Direct transformation of geophysical data, such as SkyTEM, into a deterministic subsurface model is risky, because too much reliance on geophysical mapping can lead to seriously wrong hydrogeological models (Andersen et al., 2013). Uncertainties are expected in both, geophysical and lithological data and the shape of the fitted histogram curve reflects those. High uncertainty is associated with the transition zone; around 50% sand probability. Although the cut off value that divides the SkyTEM dataset into sand and clay is calibrated, there is a large quantity of high uncertain cells included which make the measured TPs directly dependent on the cut off value. Therefore the facies proportion and mean length are very sensitive to the selection of the cut-off value. As a result, the MCM in the lateral direction, as part of the TProGS setup, is highly dependent on the way the SkyTEM data is treated. Difficulties in the integration of the two data types are indicated in Figure 2. Small scale heterogeneities indicated by the borehole descriptions are not represented by the coarser SkyTEM dataset. This supports computing the horizontal and vertical TPs individually using SkyTEM and borehole data, respectively.

The SkyTEM dataset used in the present study is a 3D grid of 20m x 20m x 2m which was spatially interpolated from soundings with distances of about 17 m and 50-100 m along and between the flight lines, respectively. To reduce the overconditioning problem it might have been preferable to use the direct sounding data instead of the interpolated dataset. A similar effect is achieved by resampling, but here interpolated data with a higher uncertainty than the direct soundings are used.

Simulating a binary system is a crude simplification of the broad range of sediments in the glacial sequence. However, classifying the SkyTEM data into discrete facies or deriving the soft information on facies membership are peculiar in a multi facies environment. Additionally less abundant facies (e.g. gravel) will show extremely uncertain correlations in the histogram probability matching method. Last the less abundant facies might be represented on a 20m domain, but it will often not be visible on the 100m domain chosen for the subsequent hydrological flow simulations. Dell'Arciprete et al. (2010) present a study where geostatistics are successfully implemented to simulate small scale heterogeneities in a multi facies environment.

Data footprint

Borehole and SkyTEM data are integrated by the histogram probability matching method (He et al., 2013b), where differences in support scale are partly neglected. The support scales of the two data types are expected to vary. The lithological data from the boreholes are aggregated to 2m to be in better vertical agreement with the geophysical dataset. The agreement in the lateral direction is more questionable, because the footprint increases with depth for the geophysical data. The footprint is approximately 15-20m on the surface and in the range of 50m at 30m penetration depth.

Split sample test

Both datasources have advantages and disadvantages: Borehole data have a higher data certainty and a finer spatial resolution in the vertical extent to better represent smaller sand features, but are essentially undersampled in the lateral extend. On the other hand, SkyTEM data have a good spatial coverage and represent the bigger sand features well, but at the same time the data are associated with a higher data uncertainty. At this point, four major sources of uncertainty can be defined: (1) The inversion that transforms the SkyTEM measurement into resistivity, (2) the borehole data, (3) the relationship between lithology and resistivity and (4) the footprint mismatch between small scale borehole data and large scale SkyTEM data. So it is precarious to assume the SkyTEM data as true geology, but it can serve as a reference/benchmark when validating the simulation results. The onlyBH scenario does not capture all of the main sand features, which are revealed by the SkyTEM survey: Only 20% of the high resistivity cells, where the resistivity is greater than 70Ωm are simulated correctly. For the onlySky20 scenario only 44% of the sand descriptions in the boreholes are simulated correctly, which underlines that the SkyTEM data does not measure the finer sand features correctly. The conducted split sample test does not allow to draw firm conclusions on simulation performance, it rather analyses the agreement between the two dataset propagated through the model.

Overconditioning

Correlated data, both temporally and spatially are a common problem in hydrogeological investigations. It has not been previously reported how TProGS is able to handle such a conditioning dataset. TProGS stochastically simulates the subsurface facies system by utilizing the two mutually dependent steps SIS and simulated quenching. It is not assured if the soft information is considered accordingly for the cokriging of the local probability estimate in the SIS step nor if it is accounted for in the objective function used for the simulated quenching in the latest TProGS version. However Deutsch and Wen (2000) successfully integrate exhaustive soft data in simulated quenching. Work around methods have to be developed to overcome the problems associated with overconditioning. The most intuitive approach is to out-thin the original soft dataset by sampling only some of the data and to include a moving sampling strategy to account for the spatial variation in the original dataset. A drawback of this approach is that valuable information might be lost, which again underlines the need for model validation to find a justifiable sampling distance where the original information is best kept. The out-thinning approach works as a very pragmatic solution for a study-specific problem and its generalization might be limited. Thinning the SkyTEM dataset out and only considering data on a 200m spaced moving sampling grid gives the most satisfying results.

Non-stationarity

Non-stationarity can be identified by subdividing the SkyTEM dataset (Figure 2 and 4). It is successfully tested if abundant conditioning data alone is capable of reproducing the observed non-stationary patterns. In a situation of sparse data, e.g. only borehole data for conditioning, these non-stationary trends cannot be reproduced correctly. Seifert and Jensen (1999) present an approach to model non-stationarity, which might be more suitable for sparse conditioning data. They suggested dividing the model domain into several stationary sub-domains, and each subdomain is then characterized using independent MCMs. When subdiving the model domain, care must be taken, that no major features are cut, because it is then difficult to model them correctly. This approach was tested in the present study, but results revealed that this method is not easily applicable in situations of abundant conditioning data, because large coherent sand features are cut by the sub-division and their connectivity could not be simulated adequately.

Performance criteria

We identified and tested five performance criteria for validating the model.

Sand proportion. Artificial conditioning data outside the target area honoring the defined proportion and MCM may help to make the simulation more homogeneous. In that context, exhaustive hard conditioning outside the simulation target can be tested.

Mean length. The simulated and measured TPs are compared by Carle (1997) and Carle et al. (1998). (Carle et al., 1998) simulate a four category system and the simulated quenching yields a perfect match between the modeled TPs and the defined MCM. On the other hand, Carle (1997) underlines that small deviations are to be expected and shows this by various examples where different SIS and simulated quenching parameters are tested.

Geobody connectivity. The connectivity is partly dependent on the proportion. The sand connectivity for the simulation based on the BH-Sky200moving scenario is simulated lower and the sand proportion higher in comparison to the results from the BH-Sky20static scenario. This shows that the geobody connectivity is not fully depending on the proportion in this study. However it is a more feasible performance criterion for proportions far below the percolation threshold.

Facies probability distribution. A good agreement between the simulated facies probability distribution and the original soft dataset doesn’t ensure that the allocation pattern of the simulated probability is correct. This becomes evident when validating the results of the BH-Sky500static scenario.

Facies probability – resistivity bias. The simulated facies probability should be in agreement with a corresponding resistivity observation to ensure that the spatial allocation pattern is simulated correctly. All bins are weighted the same, neglecting the inequality of data in each bin.

We used 25, 10 and 10 realizations to compute the first three performance criteria, respectively. Computing a moving average shows than the mean converges to +/-2% deviation to the final mean after ca. 15 realizations for the first criterion and after ca. 5 realizations for the second and third criteria, which justifies the selected number of realizations. The two latter criteria incorporate the computed probability map based on 25 realizations. Probability maps proved to be a useful tool to investigate the inter variability among realizations (Alabert, 1987; Carle, 2003; Mariethoz et al., 2009b). The results of the onlyBH scenario show the highest inter variability and a moving average tested at 10 random locations in the grid shows that after 20 realizations the mean converges to less than +/-20% from the final mean and to less than +/-10% after 23 realizations. These numbers are supposed to decrease as the conditioning data increase and therefore are 25 realizations in the analysis of the two latter criteria justifiable.

Table 4 compiles the five performance criteria for two different TProGS simulations: The BH-Sky20static- and the BH-Sky200moving scenario. The advantage of using multiple performance criteria is that concentrating on a single criterion may reveal an alleged good result, while another criterion attests a poor performance to the same simulation. Therefore a weighted and balanced analysis of the performance criteria helps to identify the best result. In this study, where abundant data are available, a good performance of the two latter criteria is as important as simulating accurate mean length/proportion. For example, if only considering sand proportion and mean length, it can be argued that the validation favors the BH-Sky20static scenario. However both, the facies probability distribution as well as the facies probability - resistivity bias attest poor performance. On the other hand, if interpreting the probability distribution only, it seems that the validation favors the BH-Sky500static scenario. Collectively, the conclusion is that the BH-Sky200moving scenario generates the overall most balanced results.

Appendix IV

Bibliography

Alabert, F., Stochastic imaging of spatial distributions using hard and soft information, M. S. thesis, 1987.

Andersen, T. R., S. E. Poulsen, S. Christensen and F. Joergensen, A synthetic study of geophysics-based modelling of groundwater flow in catchments with a buried valley, Hydrogeology Journal, 21, 491-503, 2013.

Auken, E., A. V. Christiansen, J. H. Westergaard, C. Kirkegaard, N. Foged and A. Viezzoli, An integrated processing scheme for high-resolution airborne electromagnetic surveys, the SkyTEM system, Exploration Geophysics, 40(2), 184-192, 2009.

Babak, O. and C. V. Deutsch, An intrinsic model of coregionalization that solves variance inflation in collocated cokriging, Computers & geosciences, 35(3), 603-614, 2009.

Buttafuoco, G., T. Caloiero and R. Coscarelli, Spatial uncertainty assessment in modelling reference evapotranspiration at regional scale, Hydrology and earth system sciences, 14(11), 2319-2327, 2010.

Caers, J., History matching under training-image-based geological model constraints, Spe journal, 8(3), 218-226, 2003.

Caers, J. and T. Zhang, Multiple-point geostatistics: a quantitative vehicle for integrating geologic analogs into multiple reservoir models, Stanford University, Stanford Center for Reservoir Forcasting. California, USA, 2002.

Carle, S. F., T-PROGS:Transition Probability Geostatistical Software, University of California, Davis, 1996.

Carle, S. F., Implementation schemes for avoiding artifact discontinuities in simulated annealing, Mathematical Geology, 29(2), 231-244, 1997.

Carle, S. F., Integration of Soft Data into Categorical Geostatistical Simulation. Not published manuscript, Water Resources Research, 2003.

Carle, S. F. and G. E. Fogg, Transition probability-based indicator geostatistics, Mathematical Geology, 28(4), 453-476, 1996.

Carle, S. F. and G. E. Fogg, Modeling spatial variability with one and multidimensional continuous-lag Markov chains, Mathematical Geology, 29(7), 891-918, 1997.

Carle, S. F., G. S. Weissmann and G. E. Fogg, Conditional simulation of hydrofacies architecture: A transition probability approach, SEPM Special Publication, 1(1), 147-170, 1998.

Chugunova, T. L. and L. Y. Hu, Multiple-point simulations constrained by continuous auxiliary data, Mathematical geosciences, 40(2), 133-146, 2008.

dell'Arciprete, D., R. Bersezio, F. Felletti, M. Giudici, A. Comunian and P. Renard, Comparison of three geostatistical methods for hydrofacies simulation: a test on alluvial sediments, Hydrogeology Journal, 20(2), 299-311, 2012.

dell'Arciprete, D., F. Felletti and R. Bersezio, Simulation of Fine-Scale Heterogeneity of Meandering River Aquifer Analogues: Comparing Different Approaches, In P. M. Atkinson and C. D. Lloyd (eds.), geoENV VII - Geostatistics for Environmental Applications, Quantitative Geology and Geostatistics. Springer, 16, 127-137, 2010.

Deutsch, C. V. and P. W. Cockerham, Practical Considerations in the Application of Simulated Annealing to Stochastic Simulation, Mathematical Geology, 26(1), 67-82, 1994.

Deutsch, C. V. and A. Journel, GSLIB: Geostatistical Software Libary and Users Guide, Oxford University Press, New York, 1992.

Deutsch, C. V. and X. H. Wen, Integrating large-scale soft data by simulated annealing and probability constraints, Mathematical Geology, 32(1), 49-67, 2000.

Emery, X. and J. Parra, Integration of crosswell seismic data for simulating porosity in a heterogeneous carbonate aquifer, Journal of Applied Geophysics, 98, 254-264, 2013.

Falivene, O., L. Cabrera, J. A. Munoz, P. Arbues, O. Fernandez and A. Saez, Statistical grid-based facies reconstruction and modelling for sedimentary bodies. Alluvial-palustrine and turbiditic examples, Geologica Acta, 5(3), 199-230, 2007.

Fleckenstein, J. H., R. G. Niswonger and G. E. Fogg, River-aquifer interactions, geologic heterogeneity, and low-flow management, Ground Water, 44(6), 837-852, 2006.

Gringarten, E. and C. V. Deutsch, Teacher's Aide Variogram Interpretationand Modeling, Mathematical Geology, 33(4), 507-534, 2001.

Gunnink, J. and B. Siemon, Combining airborne electromagnetics and drillings to construct a stochastic 3D lithological model, 15th European Meeting of Environmental and Engineerig Geophysics, Dublin, Ireland, 2009.

He, Xiluan., T. O. Sonnenborg, F. Jorgensen, A. S. Hoyer, R. R. Moller and K. H. Jensen, Analyzing the effects of geological and parameter uncertainty on prediction of groundwater head and travel time, Hydrology and earth system sciences, 17(8), 3245-3260, 2013a.

He, Xin., J. Koch, T. O. Sonnenborg, F. Jorgensen, C. Schamper and J. C. Refsgaard, Uncertainties in constructing stochastic geological models using transition probability geostatistics and transient AEM data., Water Resources Research, in revision, 2013b.

Hojberg, A. L. and J. C. Refsgaard, Model uncertainty - parameter uncertainty versus conceptual models, Water Science and Technology, 52(6), 177-186, 2005.

Honarkhah, M. and J. Caers, Direct Pattern-Based Simulation of Non-stationary Geostatistical Models, Mathematical geosciences, 44(6), 651-672, 2012.

Hovadik, J. M. and D. K. Larue, Static characterizations of reservoirs: refining the concepts of connectivity and continuity, Petroleum Geoscience, 13(3), 195-211, 2007.

Hubbard, S. S. and Y. Rubin, Hydrogeological parameter estimation using geophysical data: a review of selected techniques, Journal of contaminant hydrology, 45(1-2), 3-34, 2000.

Jorgensen, F., H. Lykke-Andersen, P. B. E. Sandersen, E. Auken and E. Normark, Geophysical investigations of buried Quaternary valleys in Denmark: an integrated application of transient electromagnetic soundings, reflection seismic surveys and exploratory drillings, Journal of Applied Geophysics, 53(4), 215-228, 2003a.

Jorgensen, F., P. B. E. Sandersen and E. Auken, Imaging buried Quaternary valleys using the transient electromagnetic method, Journal of Applied Geophysics, 53(4), 199-213, 2003b.

Jorgensen, F., P. B. E. Sandersen, E. Auken, H. Lykke-Andersen and K. Sorensen, Contributions to the geological mapping of Mors, Denmark - A study based on a large-scale TEM survey, Bulletin of the Geological Society of Denmark, 52, 53-75, 2005.

Journel, A., Beyond covariance: the advent of multiple-point geostatistics., Geostatistics Banff. Springer, 1, 225-223, 2004.

Krumbein, W. C. and M. F. Dacey, Markov Chains and Embedded Markov Chains in Geology, Mathematical Geology, 1(1), 79-96, 1969.

Lee, S. Y., S. F. Carle and G. E. Fogg, Geologic heterogeneity and a comparison of two geostatistical models: Sequential Gaussian and transition probability-based geostatistical simulation, Advances in Water Resources, 30(9), 1914-1932, 2007.

Liu, Y. H., Using the Snesim program for multiple-point statistical simulation, Computers & geosciences, 32(10), 1544-1563, 2006.

Mariethoz, G., P. Renard, F. Cornaton and O. Jaquet, Truncated Plurigaussian Simulations to Characterize Aquifer Heterogeneity, Ground Water, 47(1), 13-24, 2009a.

Mariethoz, G., P. Renard and R. Froidevaux, Integrating collocated auxiliary parameters in geostatistical simulations using joint probability distributions and probability aggregation, Water Resources Research, 45, W08421, 2009b.

McKenna, S. A. and E. P. Poeter, Field example of data fusion in site characterization, Water Resources Research, 31(12), 3229-3240, 1995.

Neuman, S. P., Maximum likelihood Bayesian averaging of uncertain model predictions, Stochastic Environmental Research and Risk Assessment, 17(5), 291-305, 2003.

Refsgaard, J. C., S. Christensen, T. O. Sonnenborg, D. Seifert, A. L. Hojberg and L. Troldborg, Review of strategies for handling geological uncertainty in groundwater flow and transport modeling, Advances in Water Resources, 36, 36-50, 2012.

Refsgaard, J. C., J. P. van der Sluijs, J. Brown and P. van der Keur, A framework for dealing with uncertainty due to model structure error, Advances in Water Resources, 29(11), 1586-1597, 2006.

Refsgaard, J. C., E. Auken, C. A. Bamberg, B. S. Christensen, T. Clausen, E. Dalgaard, F. Effersoe¸, V. Ernstsen, F. Gertz, A. L. Hansen, X. He, B. H. Jacobsen, K. H. Jensen, F. Joegensen, L. F. Joergensen, J. Koch, B. Nilsson, C. Petersen, G. De Schepper, C. Schamper, K. I. Soerensen, R. Therrien, C. Thirup and A. Viezzoli, Nitrate reduction in geologically heterogeneous catchments - a framework for assessing the scale of predictive capability of hydrological models., Science of the Total Environment, -(- 0), 2014.

Renard, P., Stochastic hydrogeology: What professionals really need?, Ground Water, 45(5), 531-541, 2007.

Renard, P. and D. Allard, Connectivity metrics for subsurface flow and transport, Advances in Water Resources, 51, 168-196, 2013.

Ritzi, R. W., Behavior of indicator variograms and transition probabilities in relation to the variance in lengths of hydrofacies, Water Resources Research, 36(11), 3375-3381, 2000.

Rubin, Y., X. Y. Chen, H. Murakami and M. Hahn, A Bayesian approach for inverse modeling, data assimilation, and conditional simulation of spatial random fields, Water Resources Research, 46, W10523, 2010.

Schamper, C. and E. Auken, SkyTEM Survey Norsminde and Lillebaek. NiCA pproject 2011, HydroGeophysics Group. Aarhus University, 2011-06-16, 2012.

Seifert, D. and J. L. Jensen, Using sequential indicator simulation as a tool in reservoir description: Issues and uncertainties, Mathematical Geology, 31(5), 527-550, 1999.

Seifert, D., T. O. Sonnenborg, P. Scharling and K. Hinsby, Use of alternative conceptual models to assess the impact of a buried valley on groundwater vulnerability, Hydrogeology Journal, 16(4), 659-674, 2008.

Sorensen, K. I. and E. Auken, SkyTEM - a new high-resolution helicopter transient electromagnetic system, Exploration Geophysics, 35(3), 194-202, 2004.

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

Troldborg, L., J. C. Refsgaard, K. H. Jensen and P. Engesgaard, The importance of alternative conceptual models for simulation of concentrations in a multi-aquifer system, Hydrogeology Journal, 15(5), 843-860, 2007.

Weissmann, G. S., S. F. Carle and G. E. Fogg, Three dimensional hydrofacies modeling based on soil surveys and transition probability geostatistics, Water Resources Research, 35(6), 1761-1770, 1999.

Weissmann, G. S. and G. E. Fogg, Multi-scale alluvial fan heterogeneity modeled with transition probability geostatistics in a sequence stratigraphic framework, Journal of Hydrology, 226(1-2), 48-65, 1999.

Ye, M. and R. Khaleel, A Markov chain model for characterizing medium heterogeneity and sediment layering structure, Water Resources Research, 44(9), 2008