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 #1: Anonymous Referee.

We would like to thank the anonymous referee #1 for her/his review which poses thoughtful comments to our manuscript. We appreciate her/his professional assessment in regard to the existing literatures and have tried to associate our study to other existing studies both in the introduction and the discussion section in our manuscript. We are also thankful for the comments on the methodology and have made a rigorous revision of the manuscript following the referee's suggestions. We believe the changes made will improve the scientific quality of the manuscript.

Review of "Challenges in conditioning a stochastic geological model of a heterogeneous glacial aquifer to a comprehensive soft dataset" This manuscript essentially describes a practical approach to integrate auxiliary information in the transition probability approach (Tprogs). Although the approach is quite relevant for this particular application, the existing literature on this problem is largely missed. Integration of exhaustive or dense data in geostatistics has been dealt with from the early days of geostatistics (cosimulaltion, cokriging, kriging/simulation with external drift) until the most recent developments (integration of non-stationarity in multiple-point geostatistics). ***[1]****This study entirely misses this. In the limited context of Tprogs however, it may be true that this problem has not been addressed.* ***[2]****The approach proposed consists of using a decimated version of the auxiliary variable as conditioning data for the primary variable. It is quite a crude way of proceeding since it ignores the differences of support in the different variables, as well as the differences in spatial variability.* ***[3]****I think this is entirely missed in the discussions and is not tested throughout.* ***[4]****The title and abstract have a much too general focus. The paper is essentially only focusing on the transition probability method, and more particularly on the tprogs software.* ***[5]****In my opinion, this manuscript would be more appropriate in a journal focusing specifically on the application of geostatistical methods, such as Mathematical Geosciences. It may also be fit for Hydrogeology Journal because of the emphasis on the case study. In any case I would recommend major revisions before resubmission.*

[1] We appreciate the reviewer's advice and took many of the suggested literatures into consideration. We agree that our choice of literature was too much focused on TProGS and we decided to broaden it in the revised version both in the introduction and the discussion section. Please see Appendix II & III.

[2] Decimating the soft conditioning dataset may seem as an overly simplistic and very crude approach. However we are still convinced that it is adequate for this study. In this study we aim at finding the balance between too few data and too many data. The first case is presented in section 5.2 and 5.4 (BH-Sky500static); one can say that the risk to miss important features is high when conditioning to too few data. Our study mainly deals with the latter case, where too many data lead to an underestimation of the simulation uncertainty. We agree, and have acknowledged in our manuscript, that valuable information might be lost in the process of out-thinning the conditioning dataset, but during the validation process the entire SkyTEM dataset is considered. This ensures that the TProGS results, which are conditioning to a decimated dataset, are validated by the entire geophysical dataset. In our study we tried different sampling techniques; static and moving sampling at various distances. We analyzed the tradeoff between an increase of sampling distance and a decrease in accuracy of a set of realizations, where 200m moving sampling performed best. We believe that a 200m sampling distance is still sufficient to adequately capture all relevant geological features proxied by the entire dataset; this can be argued by the fraction between the observed mean length and the conditioning spacing. The mean length of a sand lens is found to be 500m and can proxy the correlation length. With a horizontal length scale of 500m and sampling at 200m we still condition the simulation with two to three soft data points in each horizontal direction for each mean sized sand feature on average. In the vertical direction we find a very similar conditioning density as in the horizontal direction; 5m mean length of a sand feature and conditioning data is placed in each layer (2m). This means that the fraction between mean length and conditioning spacing is essentially the same in both horizontal and vertical direction, which forms a sound representation of the spatial variability in both directions.

[3] The primary variable "probability of sand occurrence" was derived from the secondary variable "resistivity" by the histogram probability matching method. The reviewer addresses the question of support and spatial variability of the two variables. The resistivity data are given with a 20m x 20m x 2m grid size and integrated with the borehole data which represent local observation with a fine and coarse representation of the subsurface in vertical and horizontal direction, respectively. The borehole data were aggregated to 2m in vertical direction (He et al., 2013b). Thus the support of the two variables can be expected to be similar in vertical direction, but it might be marginally exaggerated in the horizontal direction, because we do not know if the borehole observation supports the 400m^2 of gridded resistivity in the lateral direction. However we can be certain that the 2m in the vertical direction are well represented by the borehole data. The same applies for the spatial variability. This leads to an interesting discussion, but we decided not to take it into consideration in this study. The relevant scale at which we want to capture geological heterogeneity for our subsequent hydrological flow modeling is well represented by a 20m x 20m x 2m resolution, because the flow modelling will be based on a 100m grid size. Therefore deviations in support scale between the two variables can be neglected. Parts of this paragraph are considered in the revised discussion. Appendix III; lines 203-209.

[4] We believe that the title is specific enough, as it narrows the paper down to the most essential aspect of this study: (1) stochastic geological modeling of a (2) heterogeneous glacial aquifer with (3) comprehensive soft conditioning. Further we see the abstract as informative and compact, because it leads the reader stepwise through our study. However we reduced the abstract in length and also made some adjustments of the abstract after revising the paper (please see Appendix I; lines 2-24).

[5] We strongly believe that HESS is the correct choice of journal for our manuscript, because we see HESS as a broad and holistic platform that creates links between Hydrology and other earth sciences towards sustainable management of water resources. Our study was integrated in a hydrological research project dealing with simulation of nitrate transport and reduction in geologically heterogeneous catchments (www.nitrat.dk). This puts our study in an inter-disciplinary context for research aiming at sustainable water resources management, which complements the aims and scope of the HESS journal. Many recent publications in HESS study elaborated mathematical and statistical methods on how to generate input data to a hydrological model, without actually testing the data in a flow simulation. A couple of examples: Buttafuoco et al. (2010) conduct a regional study on reference evapotranspiration and its associated spatial uncertainty. Berne and Uijlenhoet (2007) assess radar rainfall retrieval uncertainties with a Monte Carlo based stochastic simulation.

Specific comments:
p.15222, top: To be a bit more exhaustive, references to truncated Gaussian and Truncated pluriGaussian methods could be added. Object-based models could also be mentioned.

The truncated plurigaussian method presented by (Mariethoz et al., 2009) seems very promising for incorporating auxiliary information. This method would be especially applicable in a situation with more than one auxiliary source. However, limitations lie in the rather subjective “lithotype rule” which has to be defined as the overall geological concept. A brief description is added to the introduction (Appendix II; lines 80-85), because we believe that the reader would benefit from a broader view on alternative methods.

p.15222, l.8: What is meant by compatible?

The two inter-comparison studies mentioned in this context are Lee et al. (2007) and dell'Arciprete et al. (2012). The first study evaluates the representation of geological heterogeneity in two types of models: TProGS and a variogram based sequential Gaussian simulation. The groundwater flow was modeled numerically to simulate a pumping test. Comparing the results with observed data from the pumping test showed that TProGS performed better at simulating the lateral connectivity of high-K-materials. The latter study compares hydrofacies simulation of three geostatistical methods; namely TProGS, multi-point statistics (MPS) and variogram based sequential indicator simulation (SIS). Assessments of the simulated facies proportion and the simulated facies connectivity shows that TProGS and SIS perform well at simulating less abundant facies, whereas MPS is better at simulating the connectivity of the most prominent facies correctly.

The inter-comparison studies underline that TProGS is compatible with other geostatistical methods. Although validation is limited in the above mentioned studies, some of the advantages and disadvantages are highlighted.

p.15222, l.24-25: I don’t agree with this statement: there is a large number of papers that include soft data with geostatistical methods (although maybe not specifically with tprogs). Please have a look at the works on collocated simulation, probability aggregation and tau models. It has been heavily used with indicator geostatistics, multiple-point geostatistics and object-based methods. Soft data is sometimes called "soft probability". A few examples of references: In the context of variogram-based methods:
DEUTSCH, C. V. & WEN, X. H. 2000. Integrating large-scale soft data by simulated annealing and probability constraints. Mathematical Geology, 32, 49-67. MARIETHOZ,

G., RENARD, P. & FROIDEVAUX, R. 2009. Integrating collocated auxiliary parameters in geostatistical simulations using joint probability distributions and probability aggregation. Water Resources Research, 45. In the context of multiple-point simulations:

CAERS, J. 2003. History matching under training-image-based geological model constraints. SPE Journal, 8, 218-226. CHUGUNOVA, T. & HU, L. 2008. Multiple-Point Simulations Constrained by Continuous Auxiliary Data. Mathematical Geosciences, 40, 133-146. These are only a few examples, you will find many other references regarding other geostatistical methods. More generally, the references in the introduction used are often outdated.

We agree with the reviewer's concern that parts of the relevant literature have been missed in the study. Therefore we decided to update both the introduction and the discussion section of our manuscript. The papers that are suggested by the reviewer were really helpful und could substantially contribute to improving the introduction and to set our study in a broader context. Please see Appendix II; lines 60-85 and Appendix III; lines 143-175.

p. 15223,l.12-14: There are numerous studies where exhaustive geophysics is used and validated. Below is a very recent one, however it is a problem that has been heavily studied in the last decade.

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

The revised introduction contains a paragraph on the integration of geophysics (Appendix II; lines 103-111). Emery and Parra (2013) present an interesting approach on how to integrate geophysical data (seismic measurements) and borehole data in a cross-correlated gaussian approach, where the underlying experimental variograms are checked against data. To our understanding this publication does not contribute to a systematical validation scheme for geostatistical simulation. However we could identify similar validation methods in other studies, e.g. Mariethoz et al. (2009b) and Chugunova and Hu (2008) .Please find an updated section in the introduction (Appendix II; lines 120-129)

p. 15223,l.15-16: This statement shows again that an entire side of the literature is missed on geophysical inversion with geostatistics. See for example: HANSEN, T. M., CORDUA, K. S., LOOMS, M. C. & MOSEGAARD, K. 2013. SIPPI: A Matlab toolbox for sampling the solution to inverse problems with complex prior information: Part 2- Application to crosshole GPR tomography. Computers and Geosciences, 52, 481-492.

We don't see the connection between the above mentioned publication and the need for a systematical validation of geostatistical applications as claimed on p. 15226, l.15-16. We don't aim at validating the geophysical inversion itself; instead we want to validate whether the soft data are treated appropriately by TProGS, and whether the generated set of realizations represents all comprehensive understanding of the system (mean length, proportions, connectivity and uncertainty)

p. 15223,l.19-22: These are very vague statements. For geophysical data integration, the usual criterion is a forward problem that calculates the geophysical response given a certain geological model. Valid models are those that reproduce the measures data when such a forward model is applied.

We agree with this point; the five defined validation criteria are used to evaluate the model with respect to the geophysical data, which is not a universal validation, rather a site specific validation. However site specific validation is necessary if such a forward model is applied. In this context this study provides a practical approach on how to validate stochastic realizations in a systematical way. Also, our presented validation method is a forward way of model validation, because we check the simulated facies probability and compare it with the anticipated probability based on the conditioning data. Please see a revised section in Appendix III; lines 130-133.

p. 15228,l.13-15: Such an exhaustive conditioning is never applied. Instead, a general approach is the one of probability aggregation where the probability distribution coming from the spatial model (tprogs or any other method) is combined with the prior probability coming from geophysics. See: ALLARD, D., COMUNIAN, A. & RENARD, P. 2012. Probability Aggregation Methods in Geoscience. Mathematical Geosciences, 44, 545-581. For an example of synthetic application, see LIU, Y. 2006. Using the Snesim program for multiple-point statistical simulation. Computers & Geosciences, 23, 1544-1563.

Pixel-based two point simulation methods (TProGS) are expected to be good at local data conditioning. The realization is simulated one pixel at a time, thus data conditioning is easily achieved. This opens the opportunity to add additional information as hard or soft conditioning at each simulation cell. The 3D continuous SkyTEM dataset in combination with a pixel-based two point simulation allows such an exhaustive local constraining and this study aims to utilize that. We don't aim at a probability aggregation, because the only probability distribution available in this study originates from the SkyTEM data. In our case, the primary variable (facies probability) and the secondary variable (resistivity) are linked by the histogram probability matching method (section 3.2). This method successfully comprises data integration of borehole data and geophysical data.

Section 4.3: The question of support size is not addressed here, however it is critical to know what is the support size of the geophysical data. Is a data point representative of 1m2 or 1km2? This should be a major factor when deciding whether to use subsampling or moving averaging approaches. The description of the moving sampling is not clear enough. Is the same sampling used for each realization? Are the data resampled or interpolated? Why "moving" sampling - is there a moving window defined?

The 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 large 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, but it again underlines the heavy correlation of the conditioning data on a 20m grid size.

The relevant scale at which we want to capture geological heterogeneity for our subsequent hydrological flow modeling is well represented by a 20m x 20m x 2m resolution, because the flow modelling will be based on a 100m grid size.

A few lines connected to this topic are added to the reviewed discussion section (Appendix III)

The moving sampling approach does not use a moving window thus there is no resampling, averaging or interpolation. Instead, it generates different location grids for the samples. The *n* different location grids have the same distance between the samples, but each has an accumulated shift of the origin (+ *sampling distance/n* in X and Y direction). For the TProGS application five location grids are generated, which yields five independent soft conditioning datasets. The first sampling grid have the origin (0,0) the second (40,40), the third (80,80), etc. (Figure 1) A more detailed description will be added to section 4.3. Five realizations are computed for each soft dataset; giving a total of 25 realizations.

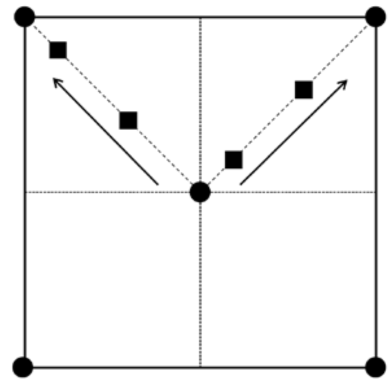

Figure 1. The systematical pattern of the moving sampling of soft data. The circles represent the original grid and the squares the four additional centers of the moved grids. This gives five soft datasets that are equally spaced.

p. 15229, l.5: such an "optimal" combined knowledge should be formulated in a Bayesian framework, which is completely not done here. Therefore the word optimal is not correct in this context.

The formulation was changed to "best combined knowledge" in order not to mislead the reader. It simply represents a merge of all available datasources (borehole and SkyTEM), thus it contains all available information on the system. It will not be further addressed if this is the optimal knowledge.

Section 5.3: Figure 5 is good and justifies the approach. However what is proposed is a "fix" that has no generality. I have nothing against such fixes that work in practice, but they should be acknowledged as such and their limitations should be clearly stated.

The approach we chose in order to overcome the problem of overconditioning, which is indeed a study specific "fix" which might not be applicable for other studies. Throughout the manuscript it is referred to as a "work around method" which to our opinion reflects the pragmatic origin. Please find the relevant section in the revised discussion on the acknowledgment of its limitations in Appendix III.

p. 15234, l.26: The terminology of more/less deterministic is not correct. A model is deterministic or it is not. I would rather speak of higher/lower variability. This is throughout the manuscript.

We agree to that and changes are done accordingly throughout the manuscript.

Section 5.4: In my opinion the approach of comparing the realizations with the probability maps derived from geophysics is flawed. Geophysics does not provide a facies distribution, but a smoothed and coarsened version of it. It is a different variable, indirectly related to the facies. Therefore they are not expected to have the same spatial features and cannot be compared in terms of connectivity or correlation scales. It is clear from figure 8 that the borehole data and the SkyTEM data present different spatial properties.

We partly agree to that point. In most cases it is incorrect to directly derive a deterministic facies distribution from geophysics. SkyTEM has been extensively used for that purpose, but limitations and risks are acknowledged (Andersen et al., 2013). In our point the probabilistic approach for handling geophysical data is much more appropriate, because it can account for uncertainties. We agree that the

3D gridded SkyTEM data are rather smoothed and coarsened, especially in the vertical direction where thin sand features tend to be overlooked. However it significantly improves our understanding of the subsurface lateral heterogeneity, because the lateral direction is essentially undersampled by boreholes. This is manifested in the realizations conditioned to borehole data only (Figure 5).

The probability map derived the geophysics is limited, but the comparison with the realizations is also done for testing self-consistency. Besides the geological validity of the geophysics we want to ensure that the probabilities from the SkyTEM data are treated accordingly during the conditional stochastic modeling with TProGS. Thus we compare input/output of TProGS.

One novelty of this study lies in the integration of two datasources into the modeling study, thus we "cherry pick" the advantages of both datasources. Borehole data for the vertical model of spatial variability, SkyTEM data for the lateral model of spatial variability, both for conditioning in respect to the associated uncertainty. We regard this procedure in combination with stochastic methods as a sound approach to assess geological uncertainties.

Section 6: This section essentially repeats material that was discussed earlier, and therefore can be shortened or removed.

The discussion (section 6) was reduced from 1824 words to 1437 words (21%) to avoid redundant discussions. However new discussions inspired by the reviews are added after shortening the original manuscript.

p. 15241, l.3-5: I'd recommend reading some textbooks on kriging with external drift (in particular the book of Chiles and Delfiner).

We appreciate the reference and it was considered in rewriting parts of the discussion.

Appendix I

Abstract

In traditional hydrogeological investigations, one geological model is often used based on subjective interpretations and sparse data availability. This deterministic approach usually does not account for any uncertainties. Stochastic simulation methods address this problem and can capture the geological structure uncertainty. In this study the geostatistical software TProGS is utilized to simulate an ensemble of realizations for a binary (sand/clay) hydrofacies model of the Norsminde catchment, Denmark. TProGS can incorporate soft data, which represent the associated level of uncertainty. High density (20m x 20m x 2m) airborne geophysical data (SkyTEM) and categorized borehole data are utilized to define the model of spatial variability in horizontal and vertical direction, respectively. Further, both datatypes are used for soft conditioning the TProGS simulations. The category probabilities for the SkyTEM dataset are derived from a histogram probability matching method, where resistivity is paired with the corresponding lithology from the categorized borehole data. This study integrates two distinct datasources into the stochastic modeling process that represent two extremes of the conditioning density spectrum; sparse borehole data and abundant SkyTEM data. Conditioning to vast soft data causes overconditioning; triggered by incorporating spatially correlated data in the modelling process. This is addressed by a work around utilizing a sampling (thinning) of the dataset. In case of abundant conditioning data it is shown that TProGS is capable of reproducing non-stationary trends. The stochastic realizations are validated by five performance criteria: (1) Sand proportion, (2) mean length, (3) geobody connectivity, (4) facies probability distribution and (5) facies probability – resistivity bias. As conclusion, a stochastically generated set of realizations soft conditioned to 200m moving sampling of geophysical data performs most satisfying when balancing the five performance criteria and can be used in subsequent hydrogeological flow modeling to address the predictive uncertainty originated from the geological structure uncertainty.

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.

Berne, A. and R. Uijlenhoet, Path-averaged rainfall estimation using microwave links: Uncertainty due to spatial rainfall variability, Geophysical research letters, 34(7), L07403, 2007.

Bredehoeft, J., The conceptualization model problem--surprise, Hydrogeology Journal, 13(1), 37-46, 2005.

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.