Dear Referee,

Thanks for taking time to review our paper “Automated Monte Carlo-based Quantification and Updating of Geological Uncertainty with Borehole Data (AutoBEL v1.0)”. We highly appreciate your comments and suggestions. Please find blew our responses and manuscript revisions. The provided responses below include 3 part:

- Part 1. Response to your comments (RC1).
- Part 2. Revised manuscript based on your and other referee’s comments.
- Part 3. Responses to comments from other referees (RC2, SC1, SC2).

The responses and manuscript revisions are highlighted in red.

We are looking forward to a positive evaluation.

On behalf of the Authors,
Zhen Yin (David)

Responses to Anonymous Referee #1 on interactive comments (RC1)

Dear authors,

I read with interest your paper entitled “Automated Monte Carlo-based Quantification and Updating of Geological Uncertainty with Borehole Data (AutoBEL v1.0)”. This paper presents a new application of the recently developed BEL framework for the updating of reservoir parameters (both local and global parameters) based on borehole data. The main contributions of the paper are (1) the development of a two-step procedure to sequentially predict lithology-related parameters (thickness of the reservoir and facies) and rock properties (permeability, porosity) and (2) a python toolbox with all the necessary code to automatically apply the methodology. The paper is well-structured, clearly written and deals with important issues related to inverse/prediction problems in Geosciences. I have a series of remarks and suggestions to improve the manuscript. They should be easily handled by slight modifications of the text. Therefore, I suggest publication after minor revision.

We highly appreciate the in-depth review and constructive suggestions from the reviewer. Overall, we agree with the reviewer comments and these have improved the clarity of our paper. Provided below are our point-to-point responses and explains how they are addressed in the revised manuscript.

General comments

1. The paper claims that the main contribution of the paper is the sequential approach (equation 11). However, how this is actually done is not (sufficiently) explicitly described. In the methodology section, it is written that “we will use the posterior realizations of khi and prior realizations of ksi to determine a conditional distribution f(ksi|khi,posterior), then we evaluate this using borehole observations dobs,ksi of ksi.” Later in the manuscript, it is only refer to the use of posterior models as “additional constraints”. My understanding is that the posterior distribution of khi represented by 250 realizations gives a new set of thickness and facies. Since the 250 prior realizations of ksi already exist in the input those can be combined to create a new prior, without having to run a new Monte Carlo sampling (what would require to re-run geostatistical realizations). However, the initial realizations of ksi are initially related to other facies distributions. To me, the only way to apply the methodology is then to have a full spatial distribution of porosity and permeability (i.e. covering the entire model) for each facies, so that those can be combined to any facies distribution. I think this point should be clarified and emphasized as it constitutes the core of the methodology.

Yes, the reviewer’s understanding of the sequential approach is correct and exactly what we proposed in the paper. The prior Monte Carlo model samples provide a full spatial distribution of all the model components, including porosity and permeability for each facies. This allows the calculation of posterior porosity and permeability models to fit any (posterior) facies distributions using direct updating. Therefore, we use the posterior facies model and borehole porosity observations as constraint in AutoBEL to calculate the posterior porosity.

We have added a paragraph to the application section (section 3.4.2) to explicitly explain how the sequential approach is performed.

2. The manuscript lets some ambiguity about the use of synthetic versus field data. I understand that the case is inspired by a field study, but I guess it went through some kind of simplification and that a 251th model of the prior was simulated and considered as the truth? This has obvious implications on the results, as a synthetic case is always easier to handle (the true model is simulated as part of the prior). I would clarify this point and drop some paragraphs or specific sentences that let think this is actually a field study. For example, section 3.1 is named “the field case”, while the first sentence of section indicates that the data set is synthetic. Page 15/Line 12, you also say that “The actual observations of these data (dobs) are measured from the borehole wireline logs”, a clear reference to field data and not simulated data from a synthetic case. However, I assume field wire logs have a higher resolution, so that comparing simulated and field data would require some upscaling? Similarly, section 3.2.1 contains many reference to data generally collected for oil and gas reservoirs (seismic data) and how the thickness model is deduced from it. This is not necessary as seismic data themselves are not included in the prediction process or the falsification process (see recent paper by Alfonzo and Oliver (2019) for example). Hydrogeological applications would probably use other type of geophysical data for the characterization of those elements (ERT or airborne EM) and removing specific explanations would not reduce the clarity while gaining in generality. If the field data from which the case is inspired were actually used but that the model is considered synthetic because of confidentiality issues, this could simply be stated in the manuscript.

Yes, the field data for prior modeling and posterior calculation are from a real case. But the model is considered as synthetic because of simplification for generic application, and because of confidential issues. There is no 251st model of the prior to be considered as the truth. We add a statement for this in the revised manuscript.

The actual well observations are in grid model resolution already. They are upscaled to from higher resolution wireline logs. Unfortunately, we don’t have the original high resolution well logs due to confidential concerns. We have rephrased Page15/Line12 in the revision to avoid the confusions.

In section 3.2.1, we have rephrased and shortened the detailed seismic explanations to gain more generality while keeping the clarity.

3. Automation is very nice as it probably broadens the potential number of users of the framework, but it comes with potential risks: How do you ensure that intermediate steps remain controlled and within a valid range? For example, Hermans et al. (2019) show in a case not extremely complex in terms of spatial distribution, that data-prediction (complex) relationships might lead to the impossibility to derive a linear relationship and to apply the

linear Gaussian analytical solution (what can be identified in the CCA plots). In such a case, the automated method would give an answer, which would be wrong. Similarly, for the model parameters, the use of the sensitivity analysis provides a way to automatically select the sensitive parameters, but the sensitivity (especially for parameters close to the threshold) might be itself dependent on the number of clusters used for DGSA. For the data, do you use a specific threshold for the automatic selection of the dimension? How is this threshold related to noise issues? A short discussion on those issues would help to picture the points where the modeler might need to give some additional input on the inversion process and where further research is needed.

The intermediate steps of automation are ensured by prior/posterior falsifications and sensitivity analysis. Prior falsification guarantees the uncertainty distribution range of prior model is not wrong. Posterior falsification validates the calculated posterior model. The use of sensitivity analysis ensures sensitive model parameters are informed by the data during uncertainty reduction. Besides, the intermediate steps can also be adjusted for users' specific application problems to ensure that the results are meaningful.

We know here for a fact that the forward problem is linear. The borehole data variable (simulated borehole data) are extracted from the model, which is simply a matrix operation. This is shown in Eq.4. Hence the CCA approach will always work here. For more complex non-linear inverse problems as mentioned in Hermans et al. (2019), statistical estimation approaches such as kernel density estimation can be used to replace CCA and Gaussian regression steps, and there are also extensions of CCA to tackle non-linear problems (e.g., Lai and Fyfe, 1999). We have extended the discussion section to address this problem. However, the paper only deals with borehole data

We agree that the cluster number used in DGSA can influence the calculation of sensitive model parameters. The effects of cluster number has been thoroughly studied by Park et al. (2016). We won't repeat this in the manuscript, but we have cited this paper in section 2.2.1. Here we use 3 clusters in DGSA as we focus on the main sensitivity effects.

Regarding the data variable, we select PCs number by preserving 90% variance. This is first because borehole data are in much lower dimension than the spatial models. It does not require much further dimension reduction. Besides, preserving 90% of the variance can be useful to filter out higher order PCs which may relate to local noises. We didn't consider the influence of noise on the threshold. We added more explanations to section 2.2.1 further explain the dimension reduction of model and data variables.

References:

Lai, P., and Fyfe, C.: A neural implementation of canonical correlation analysis, Neural Networks, 12(10), 1391–1397, doi:10.1016/S0893-6080(99)00075-1, 1999.

Lopez-Alvis, J., Hermans, T. and Nguyen, F.: A cross-validation framework to extract data features for reducing structural uncertainty in subsurface heterogeneity, Adv. Water Resour., 133, 103427, doi:10.1016/J.ADVWATRES.2019.103427, 2019.

Park, J., Yang, G., Satija, A., Scheidt, C. and Caers, J.: DGSA: A Matlab toolbox for distance-based generalized sensitivity analysis of geoscientific computer experiments, Comput. Geosci., 97, 15–29, doi:10.1016/J.CAGEO.2016.08.021, 2016.

Specific comments

1. Abstract. I recommend to start the abstract with a general sentence on the need for UQ and new methodologies to deal with it, to give some context to the study.

We have added a general sentence to start the abstract.

2. Page 6. L8-9. It might be worth mentioning here that the linearization might not be optimal. In such a case alternatives can be to linearize around the observed data, or to use a Kernel density approach (Hermans et al., 2019). Such problem is probably mostly encountered when the link data-prediction is less straightforward than here (borehole data directly measures the model parameters) and include some non-linear forward model.

We added a discussion on the limitations of linearization in the discussion section, citing the reference of Hermans et al., 2019.

3. Page 6 L32. Alternatives to PCA to reduce the dimensionality of complex models have been recently proposed such as deep neural network (Laloy et al., 2017, 2018), although they might not be directly applied within the BEL framework.

Thanks for providing the new approaches. The two references are added to the revision to suggest the alternative dimension reduction methods. But here we recommend using PCA for its bijectivity, and simplicity. For more complex geological settings one may indeed need to rely on non-linear techniques.

4. Page 12 L23-29. Maybe I am missing something here. Does the facies also use some secondary variable or trend, such as the thickness or the distance along the Yaxis to be able to represent the belts or is it just the conditioning to borehole data that makes it nicely follow the belt shapes? A simple truncated Gaussian process would rather produce lenses, no?

Yes, a geometrical trend was applied to prior facies modelling using truncated Gaussian. This is to make sure that the orientation and position of modeled facies belts are consistent with geological interpretations in Figure 5(a). The trend in the case study was provided by the field geologists with comprehensive field geology studies, and officially used by the field operator. We therefore quantify the facies model uncertainty under this trend scenario.

5. Page 13 L20. Repetition of the section number.

Section number has been corrected.

6. Page 16. L11-13. Can you shortly comment in the discussion how AutoBel can be adapted if other type of parameters must be used?

Generally, AutoBEL is easy to adapt if the other types of parameters are used for uncertainty quantification. This can be done by simply adding the additional parameters to the model variable $\mathbf{m}$.

We added a short comment on this in the discussion section.

7. Page 16 L15. I guess that higher order components are somehow sensitive to the initial and the number of realizations and potentially the parameters of the sensitivity analysis. Do you use the 250 components for DGSA? How much variance is represented by the different components? Hoffman et al. (2019) in their sensitivity analysis showed that the first 15 PC represented only 23% of the variance, representing the spatial variability with PCA is thus not always straightforward or efficient. Can you comment on that?

Yes, we used all the 250 model components for DGSA. The sensitive PCs of thickness and facies model represents about 19% of total variance. We don't think this is too small or not efficient, it simply a direct result from the fact that for our case, borehole data are in very low dimension with only 7 locations for such a large field. They cannot inform many components of such large dimensional models.

We agree that it is important to know the variance represented by sensitive PCs. This is helpful to understand how informative the boreholes are (depending on their locations). We added Hoffman et al. (2019) to our references.

8. Page 17 L2. It is not clear if you apply PCA to the data and what is you threshold on the total variance to select the dimensions? Is it a user choice or is it fixed in the code?

We select PCs of data by preserving 90% variance (fixed in this code). This is first because borehole data are in much lower dimension than the spatial models. It does not require much further dimension reduction. Besides, preserving 90% of the variance can be useful to filter out higher order PCs which may relate to local noises. We clarified this in the new revision.

9. Figure 12. Not sure to get the color scale for the facies model as the mean is not necessary one of the facies. Maybe show the median, or use a gray scale for one of the facies?

We have replaced the mean by median model

10. Page 20 L7. I guess the updated uncertainty on porosity, permeability and saturation is performed jointly.

Yes, they are updated jointly under the sequential decomposition. We added a paragraph to explain how the joint update is performed.

11. Figures 14 and 17. From the sensitivity analysis, it seems that higher order PCs(41, 22, 26, 131) are sensitive for the permeability and they do not correspond to what is shown in Figure 17 (PC1 and PC4).
In Figure17, we plot two sensitive PCs with highest variances. For permeability, higher order PCs (41, 22, 26, 131) are more sensitive than PC1 and PC4, but PC1 and PC4 contain more information than those higher PCs. Therefore, we choose to plot PC4 and PC1.

We rephrased the figure caption for more clarity.

12. Figure 18. How do you explain that the variance patterns of log-perm and Sw are significantly modified (increase or decrease) after updating in areas where no boreholes are present? Does not an increase in variance indicate a problem with the prediction, i.e. some predicted parameter values are out of the range of the prior? The reason could be that the observed data is at the edge of the linear relationship for some component in the CCA.
Well spotted! Sorry, the prior variance figures of log-perm and Sw are misplaced. In fact, the prior variance figure of log-perm in Figure 18(b) is for Sw, and vice versa. We corrected this mistake in Figure 18. The new comparison between the prior and posterior does not show the significantly increase/decrease of variances. They now are the correct results of uncertainty reduction in log-perm and Sw.

13. Page 25 L7. I guess the 45 minutes do not include the creation of the 250 MC samples forming the input.
Correct. The time used for creating the 250 prior samples are not counted. The 45 minutes are for prior falsification, sensitivity analysis, geological uncertainty reduction with new borehole observations, and posterior falsification. We clarified this in the revision.

14. Page 28 L30. Lopez-Alvis et al. (2019) recently proposed such an automatic approach for falsification of geological scenarios in a Bayesian hierarchical model based on cross-validation.
We added this reference to the revision.

References

Alfonzo, M., & Oliver, D.S. 2019. Evaluating prior predictions of production and seismic data. Computer and Geosciences, 23, 1331-1347.

Hoffmann, R., Dassargues, A., Goderniaux, P., Hermans, T., 2019. Heterogeneity and prior uncertainty investigation using a joint heat and solute tracer experiment in alluvial sediments. Frontiers in Earth Sciences - Hydrosphere, Parameter Estimation and Uncertainty Quantification in Water Resources Modeling 7, 108.

Laloy, E., Hérault, R., Jacques, D., Linde, N., 2018. Training-Image Based Geostatistical Inversion Using a Spatial Generative Adversarial Neural Network. Water Resources Research 54, 381–406. https://doi.org/10.1002/2017WR022148

Laloy, E., Hérault, R., Lee, J., Jacques, D., Linde, N., 2017. Inversion using a new low dimensional representation of complex binary geological media based on a deep neural network. Advances in Water Resources 110, 387–405. https://doi.org/10.1016/j.advwatres.2017.09.029

Lopez-Alvis, J., Hermans, T., Nguyen, F., 2019. A cross-validation framework to extract data features for reducing structural uncertainty in subsurface heterogeneity. Advances in Water Resources 133, 103427.

Thanks for providing the references. They have been added to improve the paper.

Part 2. Revised manuscript:

Automated Monte Carlo-based Quantification and Updating of Geological Uncertainty with Borehole Data (AutoBEL v1.0)

Zhen Yin[1], Sebastien Strebelle[2], Jef Caers[1]

[1]Department of Geological Sciences, Stanford University, Stanford, CA 94305, USA
[2]Chevron, Houston, USA

Correspondence to: Zhen Yin (yinzhen@stanford.edu)

Journal: *Geoscientific Model Development*.

Abstract.

Geological uncertainty quantification is critical to subsurface modeling and prediction, such as groundwater, oil/gas and geothermal, and needs to be continuously updated with new data. We provide an automated method for uncertainty quantification and updating of geological models using borehole data for subsurface developments within a Bayesian framework. Our methodologies are developed with the Bayesian Evidential Learning protocol for uncertainty quantification. Under such framework, newly acquired borehole data directly and jointly update geological models (structure, lithology, petrophysics and fluids), globally and spatially, without time-consuming model re-buildings. To address the above, an ensemble of prior geological models is first constructed by Monte Carlo simulation from prior distribution. Once the prior model is tested by means of falsification process, a sequential direct forecasting is designed to perform the joint uncertainty quantification. The direct forecasting is a statistical learning method that learns from a series of bijective operations to establish "Bayes-linear-Gauss" statistical relationships between model and data variables. Such statistical relationships, once conditioned to actual borehole measurements, allows for fast computation posterior geological models. The proposed framework is completely automated in an opensource project. We demonstrate its application by applying to a generic gas reservoir dataset. The posterior results show significant uncertainty reduction in both spatial geological model and gas volume prediction, and cannot be falsified by new borehole observations. Furthermore, our automated framework completes the entire uncertainty quantification process efficiently for such large models.

1. Introduction

Uncertainty quantification (UQ) is at the heart of decision making. This is particularly true in subsurface applications such as groundwater, geothermal, fossil fuels, CO_2 sequestration, or minerals resources. Uncertainty on the geological structures, rocks and fluids is due to the lack of access to the subsurface geological medium. For most of the subsurface applications, knowledge of the geological settings is mainly gained through the drilling of well boreholes where geophysical or rock physical measurements are made. For example, from several to tens or hundreds of boreholes are drilled in geothermal or groundwater appraisals (e.g. Le Borgne et al., 2006; Klepikova et al., 2011; Vogt et al., 2010), while in mineral resources and shale gas, the number of boreholes can be up to even thousands (e.g. Curtis, 2002; Territory et al., 2013). From borehole data, geological models are constructed for appraisal and uncertainty quantification, such as estimating water volumes stored in groundwater systems or heat storage in a geothermal system. Realistic geological modelling involves complex procedures (Caumon, 2010, 2018; de la Varga et al., 2019). This is due to the hierarchical nature of geological formations: fluids are contained in a porous medium, the porous medium is defined by various lithologies, lithological variation is contained in faults and layers (structure). In addition, boreholes are not drilled all at once, but throughout the lifetime of managing the Earth resource.

Representing the unknown subsurface geological reality by a single deterministic model has been commonly used (Beven, 1993; Royse, 2010), mostly by means of a single realization of the structure (layers/faults), rock and fluid model derived from the borehole data with other supporting geological and geophysical interpretations (e.g., Fischer et al., 2015; Kaufmann and Martin, 2008). However, relying on a single model cannot reflect the inherent geological uncertainty (Neuman, 2003). Recent advances in geostatistics have shown the importance of using multiple model realizations for uncertainty quantification in many geoscience fields, including glaciology (e.g., Cullen et al., 2017), hydrogeology (e.g., Barfod et al., 2018; Zhou et al., 2014), hydrology (e.g. Goovaerts, 2000; Marko et al., 2014), hydrocarbon reservoir modelling (e.g., Caers and Zhang, 2004; Christie et al., 2002; Dutta et al., 2019; Yin et al., 2019), geothermal (e.g. Rühaak et al., 2015; Vogt et al., 2010). Geostatistical approaches can provide multiple geological models that are conditioned/constrained to borehole data). When new boreholes are drilled, uncertainty needs to be updated. While uncertainty updating in forms of data assimilation are commonly applied in various subsurface applications, they are rarely used for updating to newly drilled borehole data, often termed “hard data” in geostatistical literatures (Goovaerts, 1997). Elfeki and Dekking (2007) used coupled Markov chain (CMC) approach to calibrate hydrogeological lithology model by conditioning on boreholes in the central Rhine-Meuse delta from the Netherlands, and then ran Monte Carlo simulation to re-evaluate the hydrogeological uncertainty. Similar approach was also used by Li et al. (2016) to reduce the uncertainty in near-surface geology for the risk assessment of soil slope stability and safety in Western Australia. Jiménez et al. (2016) updated 3D hydrogeological models by adding new geological features identified from borehole tracer tests. Eidsvik and Ellefmo (2013) and Soltani-Mohammadi et al. (2016) investigated the value of information of additional boreholes for uncertainty reduction in mineral resource evaluations.

The problem of geological uncertainty, due to its interpretative nature and the presence of prior information, is often handled in a Bayesian framework (Scheidt et al., 2018). The key part often lies in the joint quantification of the prior uncertainty on all modeling parameters, whether structural, lithological, petrophysical and fluid. A common problem is that the observed data may lie outside the defined prior model, hence is falsified. Another major issue is that most of the state-of-the-art uncertainty updating practices deal with each geological model component separately (a silo treatment of each UQ problem). However, the borehole data informs all components jointly, and hence any separate treatment ignores the likely dependency between the model components, possibly returning unrealistic uncertainty quantification. A final concern, more practically, lies around the automating any uncertainty updating. Geological modeling often requires significant individual/group expertise and manual intervention to make the model adhere to geological rules, hence requiring often months of work when new data is acquired. There is to date, no method that addresses, with borehole data, the falsification, the joint uncertainty quantification and the automation problem.

Recently, a uncertainty quantification protocol termed Bayesian Evidential Learning has been proposed to address decision making under uncertainty, and applied to cases in oil/gas, groundwater contaminant remediation and geothermal energy (Athens and Caers, 2019a; Hermans et al., 2018, 2019; Scheidt et al., 2018). It provides explicit standards that need to be reached at each stage of its UQ design with the purpose of decision making, including, model falsification, global sensitivity analysis, prior elicitation and data-science driven uncertainty reduction under the principle of Bayesianism. Compared to the previous works on BEL, model falsification, statistical learning-based uncertainty reduction approaches and automation are what is of concern in this paper. Also, we will deal with one specific data source: borehole data, through logging or coring, for geological uncertainty quantification. First, we will introduce a scheme to address the model falsification problem involving borehole data by using robust Mahalanobis distance. We will then extend a statistical learning approach termed direct forecasting (Hermans et al., 2016; Satija et al., 2017; Satija and Caers, 2015) to reduce uncertainty of all geological model parameters, jointly, using all (new) borehole data simultaneously. To achieve this, we will present a model formulation that involves updating based on the hierarchy typically found in subsurface formation: structures, then lithology, then property and fluid distribution. Finally, we will show how the proposed framework can be completely automated in an opensource project. With a generalized field case study of uncertainty quantification of gas volume in an offshore reservoir, we will illustrate our approach and emphasize the need for automation, minimizing the need for tuning parameters that require human interpretation.

2. Methodology

2.1 Bayesian Evidential Learning

2.1.1 Overview

We establish the geological uncertainty quantification framework based on Bayesian Evidential Learning (BEL), which is briefly reviewed in this section. BEL is not a method, but a prescriptive & normative data-scientific protocol for designing uncertainty quantification within the context of decision making (Athens and Caers, 2019b; Hermans et al., 2018; Scheidt et al., 2018). It integrates four constituents in UQ – data, model, prediction and decision under scientific methods and philosophy of Bayesianism. In BEL, the data is used as evidence to infer model or/and prediction hypotheses via "learnings" from the prior distribution, whereas decision making is ultimately informed by the model and prediction hypotheses. The BEL protocol consists of six UQ steps: 1) formulating the decision questions and prediction variables; 2) statement of model parametrization and prior uncertainty; 3) Monte Carlo and prior model falsification with data; 4) global sensitivity analysis between data and prediction variables; 5) uncertainty reduction based on statistical learning methods that reflect the principle of Bayesian philosophy; 6) posterior falsification and decision making. Bayesian methods, particularly in the Earth Science rely on the statement of prior uncertainty. However, such statement may be inconsistent with data in the sense that the prior cannot predict the data, hence the important falsification step. We provide next important elements of BEL within the problem of this paper: prior model definition, falsification & inversion by direct forecasting.

2.1.2 Hierarchical model definition

In geological uncertainty quantification any prior uncertainty statement needs to involve all model components jointly. A geological model $\mathbf{m}$ typically consists of four components that are modelled in hierarchical order: structural model $\boldsymbol{\chi}$ (e.g. faults, stratigraphic horizons), rock types $\boldsymbol{\zeta}$ (which are categorical, e.g. sedimentary or architectural facies), petrophysics model $\boldsymbol{\kappa}$ (e.g. density, porosity, permeability), and subsurface fluid distribution $\boldsymbol{\tau}$ (e.g. water saturation, salinity).

$$\mathbf{m} = \{\boldsymbol{\chi}, \boldsymbol{\zeta}, \boldsymbol{\kappa}, \boldsymbol{\tau}\} \tag{1}$$

The uncertainty model then becomes the following sequential decomposition:

$$\mathrm{f}(\mathbf{m}) = \mathrm{f}(\boldsymbol{\chi}, \boldsymbol{\zeta}, \boldsymbol{\kappa}, \boldsymbol{\tau}) = \mathrm{f}(\boldsymbol{\tau}|\boldsymbol{\chi}, \boldsymbol{\zeta}, \boldsymbol{\kappa})\mathrm{f}(\boldsymbol{\kappa}|\boldsymbol{\chi}, \boldsymbol{\zeta})\mathrm{f}(\boldsymbol{\zeta}|\boldsymbol{\chi})\mathrm{f}(\boldsymbol{\chi}) \tag{2}$$

In addition, because of the spatial context of all geological formations, we divide the model variables into global and spatial ones. The global variables, such as proportions, depositional system interpretation, or trend, are scalars and not attached to any specific grid locations, whereas the spatial variables are gridded. Here, we term the global variables as $\mathbf{m}_{\mathrm{gl}}$, and the spatial as $\mathbf{m}_{\mathrm{sp}}$. In this way, the geological model variables are:

$$\mathbf{m} = \{(\boldsymbol{\chi}_{\mathrm{gl}}, \boldsymbol{\chi}_{\mathrm{sp}}), (\boldsymbol{\zeta}_{\mathrm{gl}}, \boldsymbol{\zeta}_{\mathrm{sp}}), (\boldsymbol{\kappa}_{\mathrm{gl}}, \boldsymbol{\kappa}_{\mathrm{sp}}), (\boldsymbol{\tau}_{\mathrm{gl}}, \boldsymbol{\tau}_{\mathrm{sp}})\} \tag{3}$$

The prior uncertainty $\mathrm{f}(\mathbf{m})$ of the global and spatial variables needs to be specified for each model component; this is problem specific and may require substantial amount of work by considering the existing data (e.g. the system is deltaic) and any prior knowledge about the interpreted systems. Using the prior distribution $\mathrm{f}(\mathbf{m})$, we run Monte Carlo to generate a set of L model realizations $\{\mathbf{m}^{(1)}, \mathbf{m}^{(2)}, \ldots, \mathbf{m}^{(L)}\}$. This means instantiating all geological variables $\boldsymbol{\chi}, \boldsymbol{\zeta}, \boldsymbol{\kappa}, \boldsymbol{\tau}$ jointly.

Since borehole data provide information at the locations of drilling, we define the data variables $\mathbf{d}$ through an operator $\mathbf{G}_{\mathrm{d}}$.

$$\mathbf{d} = \mathbf{G}_{\mathrm{d}}\mathbf{m} \tag{4}$$

$\mathbf{G}_\mathrm{d}$ is simply a matrix in which each element is either 0 or 1 identifying the locations of boreholes in the model $\mathbf{m}$. In this sense, borehole data are linear data because of the linear forward operator. By applying $\mathbf{G}_\mathrm{d}$ to prior geological model realizations, we obtained a set of L samples of the borehole data variable.

$$\mathbf{d} = \{\mathbf{d}^{(1)}, \mathbf{d}^{(2)}, \dots, \mathbf{d}^{(\mathrm{L})}\} \qquad (5)$$

Note that we term the actual acquired data as $\mathbf{d}_{\mathrm{obs}}$.

The prediction variable **h,** such as storage volume of a ground water aquifer, or the heat storage of a geothermal reservoir, is defined through another operator (linear or nonlinear):

$$\mathbf{h} = \mathbf{G}_\mathrm{h}(\mathbf{m}) \qquad (6)$$

Applying this function to the prior model realizations we get

$$\mathbf{h} = \{\mathbf{h}^{(1)}, \mathbf{h}^{(2)}, \dots, \mathbf{h}^{(\mathrm{L})}\} \qquad (7)$$

A common problem in practice is that the statement of prior may be too narrow (overconfidence) and hence may not in fact predict the observed data. In falsification, we use hypothetic-deductive reasoning to attempt to reject the prior by means of data, namely we state the null-hypothesis: the prior can predict the observation and attempt to reject it. This step does not involve matching models to data, it is only a statistical test. One way of achieving this is using outlier detection as discussed in the next section.

2.1.3 Falsification using multivariate outlier detection

The goal of falsification of is to test that the prior model is not wrong. The prior model should be able to predict the data. Our reasoning then is that a prior model is falsified if the observed data $\mathbf{d}_{\mathrm{obs}}$ is not within the same population as the samples $\mathbf{d}^{(1)}, \mathbf{d}^{(2)}, \dots, \mathbf{d}^{(L)}$, i.e. $\mathbf{d}_{\mathrm{obs}}$ is an outlier. Evidently, the data variable can be high-dimensional due to a large number of wells with various types of measurements on structure, facies, petrophysics and saturation, which calls for multi-variate outlier detection. We propose in this paper to use a robust statistical procedure based on Mahalanobis distance to perform the outlier detection. The robust Mahalanobis distance (RMD) for each data variable realization $\mathbf{d}^{(\ell)}$ or $\mathbf{d}_{\mathrm{obs}}$ is calculated as:

$$\mathrm{RMD}(\mathbf{d}^{(\ell)}) = \sqrt{(\mathbf{d}^{(\ell)} - \boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\mathbf{d}^{(\ell)} - \boldsymbol{\mu})} \quad , \text{for } \ell = 1, 2, \dots, \mathrm{L} \qquad (8)$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are the robust estimation of mean and covariance of the data (Hubert and Debruyne, 2010; Rousseeuw and Driessen, 1999). Assuming $\mathbf{d}$ distributes as multivariate Gaussian, the distribution of $[\mathrm{RMD}(\mathbf{d}^{(\ell)})]^2$ will be Chi-Squared χ^2_d. We will use choose the 97.5 percentile of $\sqrt{\chi^2_\mathrm{d}}$ as the tolerance for the multivariate dimensional points $\mathbf{d}^{(\ell)}$. If the $\mathrm{RMD}(\mathbf{d}_{\mathrm{obs}})$ falls outside the tolerance ($\mathrm{RMD}(\mathbf{d}_{\mathrm{obs}}) > \sqrt{\chi^2_{\mathbf{d},97.5}}$), the $\mathbf{d}_{\mathrm{obs}}$ will be considered as outliers, which means the prior model has very small probability to predict the actual observations, hence is falsified. It should be noted that the $\mathbf{d}_{\mathrm{obs}}$ dealt in this paper is at model grid resolution. Outlier detection using the Mahalanobis distance has the advantages of providing robust statistical calculations. In addition, diagnostic plots can be used to visualize the result for high-dimensional data. However, it requires the marginal distribution of data to be Gaussian. If the data variables are not Gaussian, other outlier detection approaches such as one-class SVM (Schölkopf et al., 2001) or Isolation Forest (Liu et al., 2008) can be used.

2.2 Direct forecasting

2.2.1 Review

If the prior model cannot be falsified, we will use direct forecasting to reduce geological model uncertainty. Direct forecasting (DF) is a prediction-focused data science approach for inverse modeling (Hermans et al., 2016; Satija et al., 2017; Satija and Caers, 2015). The aim is to estimate/learn the conditional distribution $\mathrm{f}(\mathbf{h}|\mathbf{d})$ between the prediction variable $\mathbf{h}$ and data variable $\mathbf{d}$ from prior Monte Carlo samples. Then, instead of using traditional inverse methods that require re-building models to update prediction, direct forecasting directly calculates the conditional prediction distribution $\mathrm{f}(\mathbf{h}|\mathbf{d}_{\mathrm{obs}})$ through the statistical learning based on data. The learning strategy of direct forecasting is that, by employing bijective operations, the non-Gaussian problem $\mathrm{f}(\mathbf{h}|\mathbf{d})$ can be transformed into a linear-Gauss problem of transformed variables $(\mathbf{h}^*, \mathbf{d}^*)$:

$$\mathbf{h}^* \sim \exp\left(-\frac{1}{2}\left(\mathbf{h}^* - \mathbf{h}^*_{\mathrm{prior}}\right)^{\mathrm{T}} \mathrm{C}^{-1}_{\mathrm{prior}} \left(\mathbf{h}^* - \mathbf{h}^*_{\mathrm{prior}}\right)\right); \mathbf{d}^*_{\mathrm{obs}}; \mathbf{d}^* = \mathbf{G}\mathbf{h}^* \tag{9}$$

where $\mathbf{G}$ is coefficients that linearly map $\mathbf{h}^*$ to $\mathbf{d}^*$. This makes $\mathrm{f}(\mathbf{h}^*|\mathbf{d}^*_{\mathrm{obs}})$ become a “Bayes-linear-Gauss” problem that has an analytical solution:

$$\mathrm{E}[\mathbf{h}^*|\mathbf{d}^*_{\mathrm{obs}}] = \mathbf{h}^*_{\mathrm{posterior}} = \mathbf{h}^*_{\mathrm{prior}} + \mathrm{C}_{\mathrm{prior}}\mathbf{G}^{\mathrm{T}}\left(\mathbf{G}\mathrm{C}_{\mathrm{proir}}\mathrm{G}^{\mathrm{T}}\right)^{-1}\left(\mathbf{d}^*_{\mathrm{obs}} - \mathbf{G}\mathbf{h}^*_{\mathrm{prior}}\right)$$

$$\mathrm{Var}[\mathbf{h}^*|\mathbf{d}^*_{\mathrm{obs}}] = \mathrm{C}_{\mathrm{posterior}} = \mathrm{C}_{\mathrm{prior}} - \mathrm{C}_{\mathrm{prior}}\mathbf{G}^{\mathrm{T}}\left(\mathbf{G}\mathrm{C}_{\mathrm{prior}}\mathrm{G}^{\mathrm{T}}\right)^{-1}\mathrm{G}\mathrm{C}_{\mathrm{prior}} \tag{10}$$

In detail, the specific steps of direct forecasting are:

1. Monte Carlo: generate L samples of prior model, and run forward function to evaluate data and prediction variables.
2. Orthogonality: PCA (Principal Component Analysis) on data variable **d** and prediction variable **h**.
3. Linearization: maximize linear correlation between the orthogonalized data and variables by Normal Score Transform and CCA (Canonical Component Analysis), obtaining transformed $\mathbf{h}^*, \mathbf{d}^*$.
4. Bayes-linear-Gauss: calculate conditional mean and covariance of the transformed prediction variable
5. Sampling: sample from the posterior distribution of transformed prediction variable $\mathbf{h}^*_{\mathrm{posterior}}$
6. Reconstruction: invert all bijective operations, obtaining $\mathbf{h}_{\mathrm{posterior}}$ in the original space.

One key question in direct forecasting is how to determine the Monte Carlo samples size L. Usually, the samples size L lies between 100-1000, according to the studies in water resources (Satija and Caers, 2015), hydrogeophysics (Hermans et al., 2016), hydrocarbon reservoirs (Satija et al., 2017).

The direct forecasting can also be extended to update model variables, by simply replacing the prediction variable **h** by model variable **m** in the above algorithms, to obtain $\mathrm{f}(\mathbf{m}|\mathbf{d}_{\mathrm{obs}})$ without conventional model inversions (Park, 2019). However, the high dimensionality of spatial models (millions of grid cells) imposes challenge to such extension. This is because CCA requires that the sum of input data and model variables dimensions to be smaller than the Monte Carlo samples size L: $L > \dim(\mathbf{d}) + \dim(\mathbf{m})$. Otherwise it will always produce perfect correlations (correlation coefficients be 1) (Pezeshki et al., 2004). Although PCA can significantly reduce the dimensionality of $\mathbf{m}$ from L×P to L×L, where P is the number of model parameters and L≪P, this requirement is still difficult to meet. Global Sensitivity Analysis is therefore applied, to select a subset of the PCA orthogonalized $\mathbf{m}$ that is most informed by the data variables. The subset $\mathbf{m}$ may retain only a few principal components (PCs) (Hoffmann et al., 2019), depending on how informative the boreholes are. For unselected (non-sensitive) model variables, they remain random according to their prior empirical distribution. Both the sensitive and non-sensitive variables will be used for posterior reconstruction at step 6. In this paper, we use a Distance-Based Generalized Sensitivity Analysis (DGSA) method (Fenwick et al., 2014; Park et al., 2016) to perform sensitivity analysis. Compared to the other global sensitivity analysis such as variance-based methods (e.g. Sobol, 2001, 1993), regionalized methods (e.g. Pappenberger et al., 2008; Spear and Hornberger, 1980), or tree-based method (e.g. Wei et al., 2015), DGSA has its specific advantages for high-dimensional problems while requiring no functional form between model responses and model parameters. It can efficiently compute global sensitivity, which makes it preferred for our geological UQ problem where the models are large and computationally intensive. When performing PCA on the data variable $\mathbf{d}$, we select the PCs by preserving 90% variance. Note that borehole data are in much lower dimension than spatial models, hence already low dimension.

2.2.2 Direct forecasting on a sequential model decomposition

We defined our prior uncertainty model (Eq.2) through a sequential decomposition of hierarchical model components. Likewise, the conditioning of such model components to borehole data will be done, using direct forecasting in a sequential fashion:

$$f(\boldsymbol{\chi}, \boldsymbol{\zeta}, \boldsymbol{\kappa}, \boldsymbol{\tau} | \mathbf{d}_{obs}) = f(\boldsymbol{\tau} | \boldsymbol{\chi}_{posterior}, \boldsymbol{\kappa}_{posterior}, \boldsymbol{\zeta}_{posterior}, \mathbf{d}_{obs,\tau}) f(\boldsymbol{\kappa} | \boldsymbol{\chi}_{posterior}, \boldsymbol{\zeta}_{posterior}, \mathbf{d}_{obs,\kappa}) f(\boldsymbol{\zeta} | \boldsymbol{\chi}_{posterior}, \mathbf{d}_{obs,\zeta}) f(\boldsymbol{\chi} | \mathbf{d}_{obs,\chi}) \quad (11)$$

Following this equation, the joint uncertainty quantification is equivalent to a sequential uncertainty quantification, where uncertainty quantification of one model component conditions to borehole data and posterior models of the previous components. Direct forecasting has not been applied within this framework of Eq (11), hence this is one of the new contributions in this paper. In applying direct forecasting we will use the posterior realizations of $\boldsymbol{\chi}$ and prior realizations of $\boldsymbol{\zeta}$ to determine a conditional distribution $f(\boldsymbol{\zeta} | \boldsymbol{\chi}_{posterior})$, then we evaluate this using borehole observations $\mathbf{d}_{obs,\zeta}$ of $\boldsymbol{\zeta}$.

To apply this framework to discrete variables such as lithology, we need a different method for dimension reduction than using PCA. PCA relies on a reconstruction by linear combination of principal component vectors, which becomes challenging when the target variable is discrete. Figure 1 shows this problem that discrete lithology model cannot be recovered from inverse PCA. To avoid this, a level set method of signed distance function (Osher and Fedkiw, 2003; Deutsch and Wilde, 2013) is employed to transform rock type models into a continuous scalar field of signed distances before applying PCA. Here, considering S discrete rock types in model $\boldsymbol{\zeta}$, for each s-th ($s = 1, 2, \dots, S$) rock type, the signed distance $\psi_s(\mathbf{x})$ from location $\mathbf{x}$ to its closest boundary $\mathbf{x}_\beta$ can be computed as:

$$\psi_s(\mathbf{x}) = \begin{cases} +\|\mathbf{x}-\mathbf{x}_\beta\|, & \text{if } \boldsymbol{\zeta}(\mathbf{x}) = s \\ -\|\mathbf{x}-\mathbf{x}_\beta\|, & \text{otherwise} \end{cases} \qquad s = 1, 2, \dots, S \tag{12}$$

Figure 2 illustrates the concept of using a signed distance function to first transform a sedimentary lithology model to continuous signed distances for PCA. We observe that, with the signed distance as an intermediate transformation, the inverse PCA recovers the lithology model. In the case of multiple categories, we will have multiple signed distance functions.

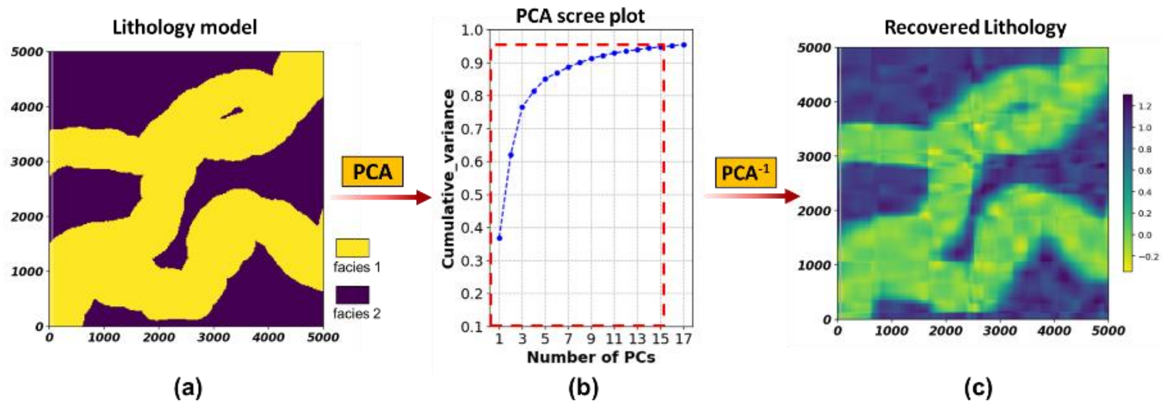

Figure 1. PCA on discrete lithology model: (a) the original lithology model (b) Scree plot of PCA on the lithology model. (c) The reconstructed model from inverse PCA using the preserved PCs (marked by the red dash line on the scree plot).

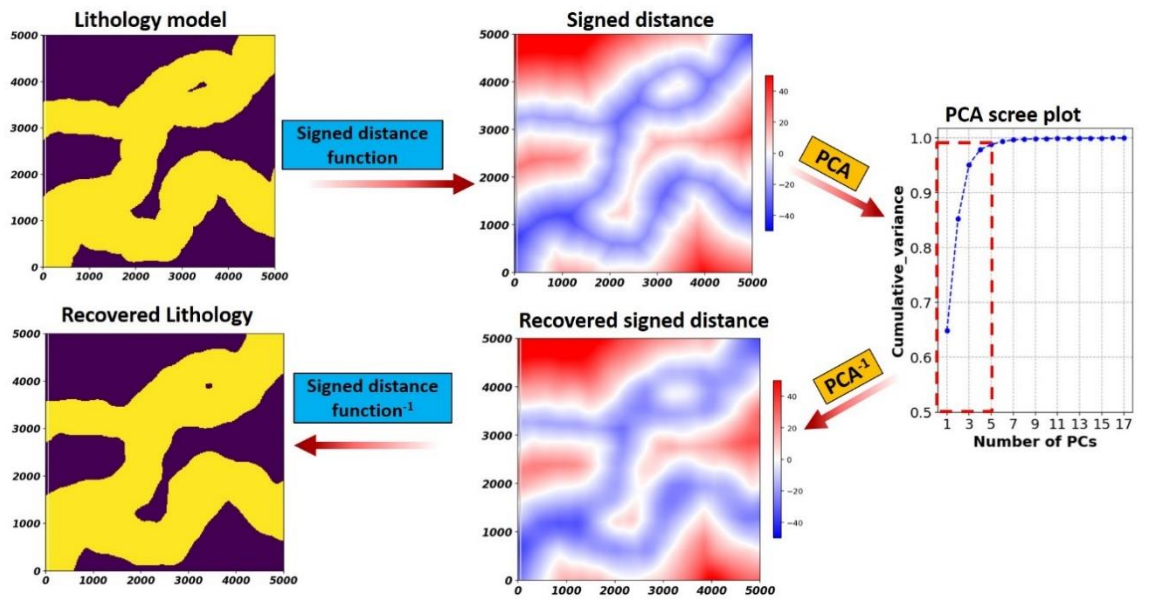

Figure 2. Example of transforming categorical lithology model to continuous signed distances for performing PCA.

2.3 Automation and Code

Our objective of automation is to allow for seamless uncertainty quantification once the prior uncertainty models have been established. Therefore, following the above described geological UQ strategies, we design a workflow in Figure 3 to automate the implementation. The workflow starts with the prior model Monte Carlo (MC) samples and borehole observations as input. All following steps including extraction of borehole data variables, prior falsification, sequential direct forecasting, posterior prediction and falsification (if required) are completely automated. With this workflow, we develop an open source Python implementation to execute the automation (named "Auto-BEL"). This opensource project can be accessed from Github (repository: https://github.com/sdyinzhen/AutoBEL-v1.0, DOI: 10.5281/zenodo.3479997). Figure 4 briefly explains the structure of the Python implementation. This automation implementation allows that, once new borehole observation and prior model is provided from "*Input*" directory, the uncertainty quantitation and updating can be performed automatically by running the Jupyter Notebook "*Control panel*". The results from the automated uncertainty quantification are stored in the "*Output*", classified as "*Model*", "*Data*", and "*Prediction*".

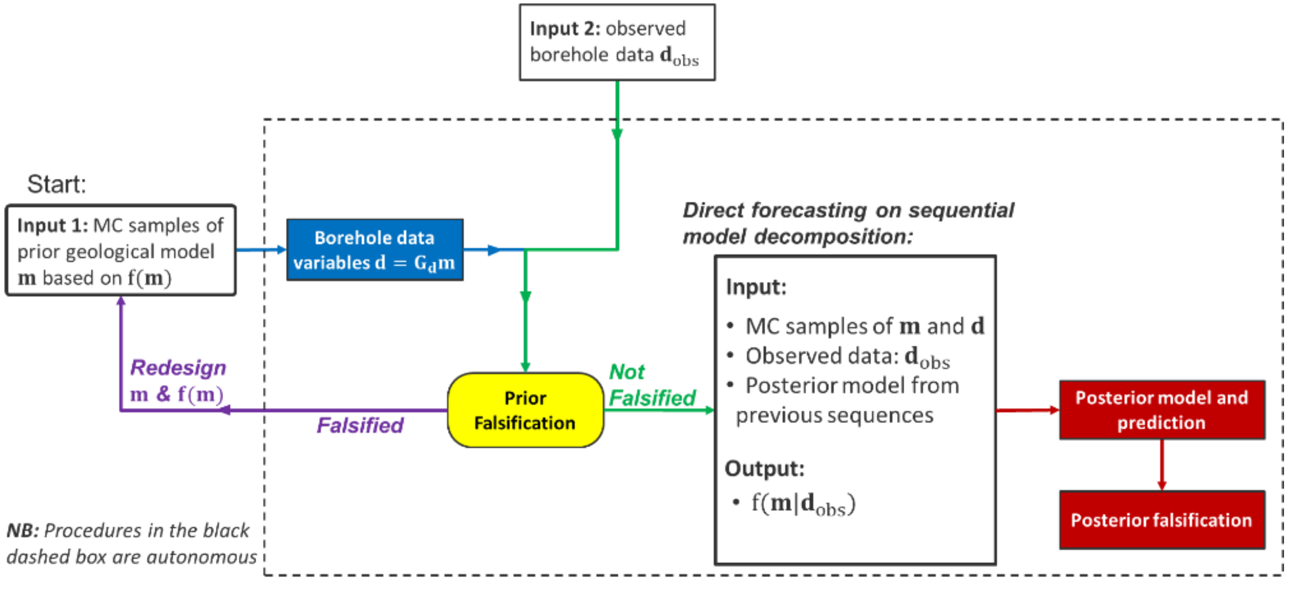

Figure 3. Proposed workflow to automate the geological uncertainty quantification.

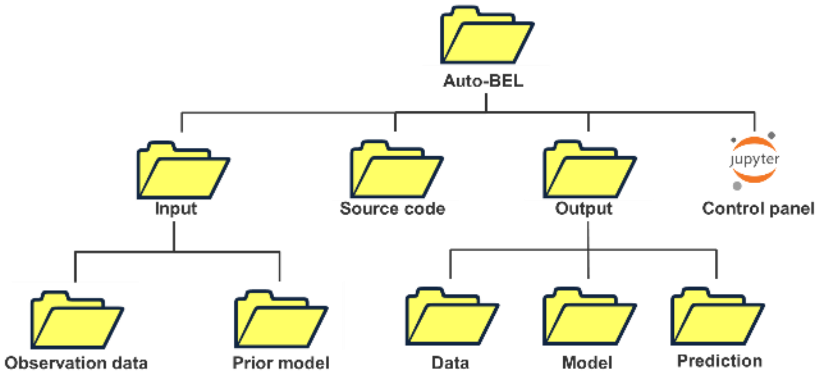

Figure 4. The structure of the Auto-BEL python implementation project.

3. Application example

3.1 The field case

We demonstrate the application of the automated UQ framework using synthetic dataset inspired by a gas reservoir located offshore Australia. This case study is considered as synthetic due to simplification for generic application and because of confidentiality issues. Its spatial size is around 50 km (EW) ×25km (NS) with thickness ranging from 75 meter to 5 meters. The reservoir rocks deposited at shallow marine environment, with four lithological facies belts corresponding to four different types of porous rocks (Figure 5a). The rock porous system contains natural gas and formation water. The major challenges lie in quantifying spatial geological uncertainty, appraising gas initially in place (GIIP), and then fast updating the uncertainty quantification when new boreholes are drilled. This will directly impact the economic decision making for reservoir development.

Initially, the reservoir geological variation is represented on a 3D model (Figure 5b) with a total of 1.5million grid cells with dimension of 200 ×100 ×75 (layers). Companies often drill exploration and appraisal wells before going ahead with producing the reservoir. They would like to decrease uncertainty by such drilling to a point where the risk is considered tolerable to start actual production. To mimic such setting, we consider that initially 4 well-bores (w1, w2, w3, w4, marked in Figure 5b) have been acquired and that models have been built using the data from these wells. Then 9 new wells (w5 to w13 in Figure 5b) are drilled, and uncertainty needs to be updated. The idea is to use the 9 new wells to automatically update the reservoir uncertainty using the above developed procedures. In order to validate our results, we will use observations from w7 to w13 to reduce the uncertainty, whereas observations from w5 and w6 to analyze the obtained uncertainty quantification.

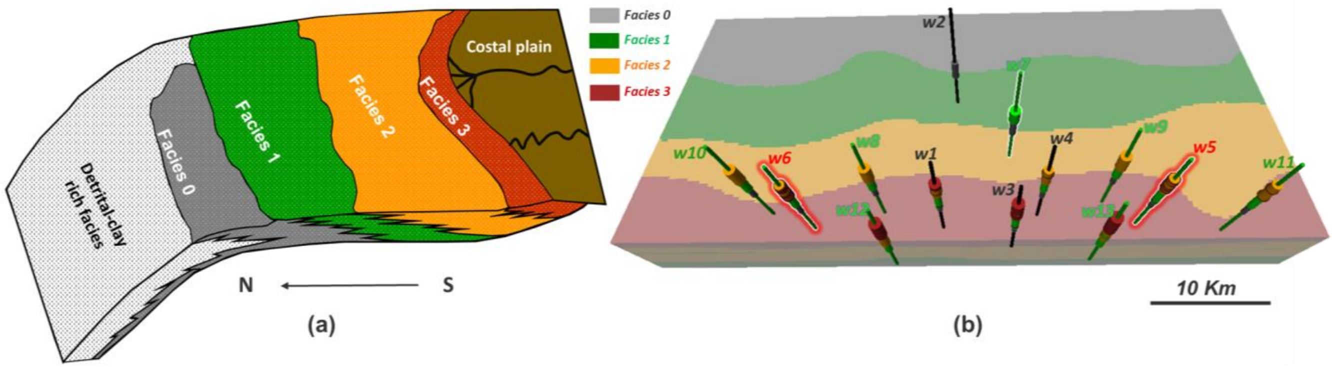

Figure 5. (a) The field geology conceptual model with the four facies belts. (b) The initial 3D geological model of facies with locations of existing boreholes and newly drilled boreholes.

3.2 Prior model parameterization and uncertainty

3.2.1 Approaches

The reservoir geological properties responsible for reserve appraisals are spatial variations of (1) reservoir thickness, spatial distributions of (2) lithological facies belts, (3) 3D porosity, (4) 3D formation water (saturation); while the spatial heterogeneity of (5) 3D permeability is critical to future production of gas, but not used in volume appraisal. Constructing a prior uncertainty model for these properties requires a balance between considering aspects of the data and overall interpretation based on such data. The strategy in the BEL framework is not to state too narrow uncertainty initially, rather to explore a wide range of possibilities. Based on interpretation from data, Table 1 containing all uncertainties and their prior distribution was constructed. We will clarify how these uncertainties were obtained.

Table 1. The global model parameter m_{gl} and its prior uncertainty distribution $f(m_{gl})$. The initial prior distributions of the parameters are mostly assumed to be uniform (formulated as U[min, max]) due to limited available data.

Model	Global parameters: m_{gl}	Prior uncertainty: $f(m_{gl})$	Source for prior uncertainty statement
Reservoir Thickness	Thickness expectation – Z_{mean}	U[36, 51] meter	Geophysical seismic interpretations, initial borehole measurements.
	Variogram range of trend – T_{range}	U[10000, 40000] meter	
	Variogram sill of trend – T_{sill}	U[350, 650]	
	Variogram range of residual – R_{range}	U[1000, 5000] meter	
	Variogram sill of residual – R_{sill}	U[4, 100]	
Lithological Facies	Proportion of facies 1 – fac1	U[0.22, 0.36]	Boreholes gamma ray logs, seismic amplitude maps,
	Proportion of facies 2 – fac2	U[0.07, 0.27]	
	Proportion of facies 3 – fac3	U[0.13, 0.19]	
Porosity & Permeability	Porosity mean in facies 1 – $\phi1$	U[0.175, 0.225]	Borehole neutron porosity logs, laboratory measurements on core samples
	Porosity mean in facies 2 – $\phi2$	U[0.275, 0.325]	
	Porosity mean in facies 3 – $\phi3$	U[0.225, 0.275]	
	Porosity mean in facies 0 – $\phi0$	U[0.125, 0.175]	
	Variogram range of porosity – ϕ_{range}	U[4000,10000] meter	
	Variogram sill of porosity – ϕ_{sill}	U[0.0015 0.003]	
	Correlation coeff. between Porosity and log-perm – $r_{\phi k}$	Normal(0.80, 0.0025)	
	log-perm mean in facies 1 – k1	U[0.3, 1.3] log(mD)	
	log-perm mean in facies 2 – k2	U[1.6, 2.6] log(mD)	
	log-perm mean in facies 3 – k3	U[1, 2] log(mD)	
	log-perm mean in facies 0 – k0	U[-1.6, -0.6] log(mD)	
	Variogram range of permeability – k_{range}	U[4000,10000] meter	
	Variogram sill of permeability – k_{sill}	U[0.9, 1.4]	
Saturation (Sw)	Coeff. a of Eq.14 (capillary pressure model) – a	U[0.041, 0.049]	Laboratory capillary pressure experiments on rock core and fluid samples
	Coeff. b of Eq. 14 – b	U[0.155, 0.217]	
	Coeff. c of Eq. 14 – c	U[0.051, 0.203]	

Thickness

First, the thickness uncertainty is mainly due to limited resolution of the geophysical seismic data and uncertainty in velocity modeling (not shown in this paper). Seismic interpretations show no faults in the geological system, but the thickness variations follow a structural trend. To model thickness uncertainty, we decompose thickness $\mathrm{Z}(\mathbf{x})$ into an uncertain trend $\mathrm{T}(\mathbf{x})$ and uncertain residual $\mathrm{R}(\mathbf{x})$

$$\mathrm{Z}(\mathbf{x}) = \mathrm{T}(\mathbf{x}) + \mathrm{R}(\mathbf{x}) \tag{13}$$

Note that most common geostatistical approaches do not consider uncertainty in trend. Uncertainty in $\mathrm{T}(\mathbf{x})$ can be estimated using geophysical data such as seismic, electrical resistivity tomography or airborne electromagnetic. This case study uses seismic data. We describe uncertainty on trend using a 2D Gaussian process (Goovaerts, 1997) with uncertain expectation and spatial covariance. The expectation is interpreted from seismic data with vertical resolution of 15 meters, while the uncertain spatial covariance is modeled using a geostatistical variogram on seismic data with uncertain range (spatial correlation length) and sill (variance). The residual $\mathrm{R}(\mathbf{x})$ is modeled using a zero-mean 2D Gaussian process with unknown spatial covariance. This term is highly uncertain, in particular the covariance, because the residual term is observed only at 4 initial borehole locations. However, the variogram range is assumed to be much smaller than the trend variogram, as residuals aim to represent more local features. Once the Gaussian process is defined, it can be constrained (conditioned) to the actual thickness observation at the vertical boreholes through the generation of conditional realizations. Note that these conditional realizations contain the uncertainties of trend and residual terms (Figure 6).

Facies

The lithological facies are considered to have rather simple spatial variability and described as “belts” (see Figure 5a). These are common in the stratigraphic progression, typical of shallow marine environments. To describe such variation, we use a 3D Gaussian process that is truncated (Beucher et al., 1993), thereby generating discrete variables. This truncated Gaussian process has specific advantage in reproducing simple organizations of ordered lithologies, thus making a useful model in our case. Because 4 facies exist, three truncations need to be made on the single Gaussian field. The truncation bounds are determined based on facies proportions. The uncertain facies proportions are obtained from lithological interpretations on borehole gamma ray logs and geophysical seismic interpretation.

Porosity and permeability

For each facies belt, rock porosity and permeability (logarithmic scale, termed as log-perm) are modelled, using two correlated 3D Gaussian processes. The cross-covariances of these processes are determined via Markov-models (Journel, 1999) that only require the specification of a correlation coefficient. Laboratory measurements on the borehole rock core samples show that permeability is linearly correlated to porosity with a coefficient 0.80, and a small experimental error (around 6% random error according to the lab scientists by repeating the experiments). The marginal distributions of porosity and log-perm are assumed to be normal but with uncertain mean and variances. The mean of porosity and log-perm is based on borehole neutron porosity logs and core sample measurements. Similar to the thickness residual modelling, the spatial covariances are modeled via a variogram respectively for porosity and permeability, with uncertain range and sill. Limited wellbore observations make variogram range and sill highly uncertain, and therefore large uncertainty bounds are assigned.

Saturation

Rocks contain gas and water; hence the uncertain saturation of water (Sw) will affect the uncertain gas volume calculations. The modelling of Sw is based on a classical empirical capillary pressure model from Leverett J-function (Leverett et al., 1942), formulated as:

$$\mathrm{Sw} = 10^{-\mathrm{a}*[\log(\mathrm{j})]^2-\mathrm{b}*\log(\mathrm{j})-\mathrm{c}} \tag{14}$$

where $\mathrm{j} = 0.0055 * \mathrm{h}\sqrt{\emptyset/\mathbf{k}}$, and h is height above the reservoir free water level. The uncertainty parameters in this fluid modelling are the coefficients a, b, c. Their prior distributions are provided by capillary pressure experiments using rock core plugs and reservoir fluids as shown in Table 1.

3.2.2 Monte Carlo

By running Monte Carlo from the given prior distribution in Table 1

3.2.2 Monte Carlo

, a set of 250 geological model realizations are generated. Figure 6 displays Monte Carlo realizations of the geological model: thickness trend and corresponding thickness model, facies, porosity, permeability (log-perm) and Sw. With prior samples of geological model, prior prediction of GIIP are calculated, using the following linear equation:

$$\mathrm{GIIP} = \mathrm{study\ area} * \mathrm{thickness} * \mathrm{porosity} * (1-\mathrm{Sw})/\mathrm{Bg} \tag{15}$$

where the Bg is the gas formation volume factor provided from laboratory measurements. The calculated GIIP prediction is plotted in Figure 7. The plot shows that the initial prediction of reservoir gas storage volume has wide range, which means significant risk can exist during decision makings for field development.

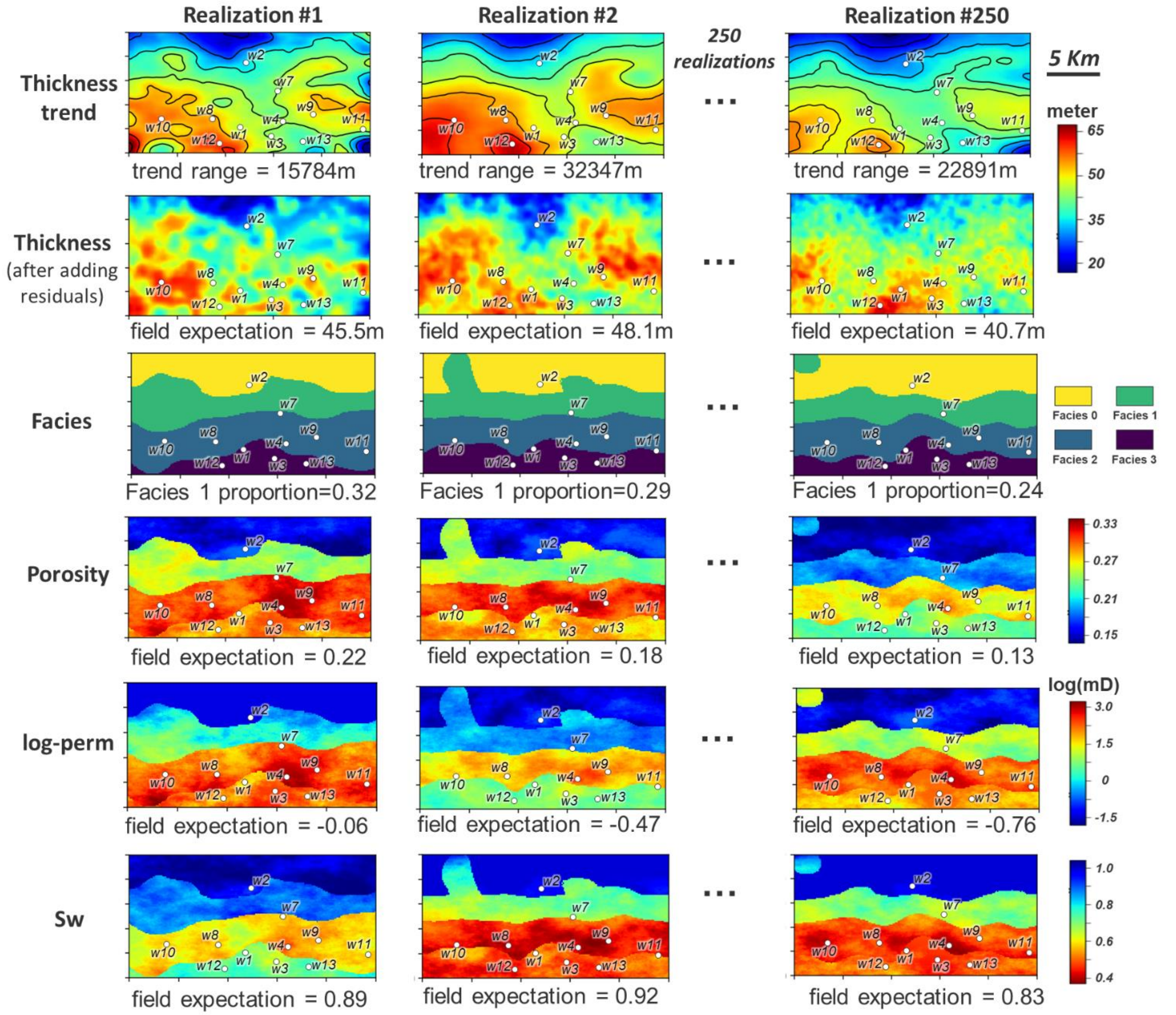

Figure 6. Layer view of prior Monte Carlo model samples of thickness trend and corresponding thickness, facies, porosity, permeabilty (logarithmic, termed as log-perm), and Sw.

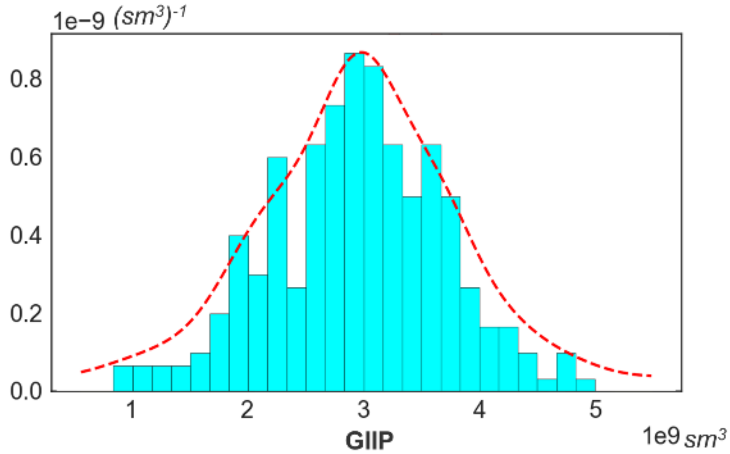

Figure 7. Uncertainty quantifiation of GIIP based on prior uncertainty and 4 boreholes.

3.3 Prior falsification with newly acquired borehole data

Table 1 is a subjective statement of prior uncertainty. When new data is acquired, this statement can be tested, using a statistical test (section 2.1.3) that may lead to a falsified prior. To perform falsification, borehole data variables at the seven new well locations (from w7 to w13) are extracted by applying the data forward operator $\mathbf{G}_{\mathrm{d}}$ to the 250 prior model realizations. It simply means extracting all thickness, facies, petrophysics and saturation at the borehole locations in the prior model. For the 2D thickness model, the new boreholes provide seven data extraction locations. For the 3D model of facies, porosity, permeability and Sw, each vertical borehole drills through 75 grid layers, thus the seven boreholes provide 2100 extracted data measurements (75 data measurements/well × 7 wells × 4 model components = 2100 data measurements). The dimensionality of data variable $\mathbf{d}$ in this case therefore equals to 2107. The actual observations of these data ($\mathbf{d}_{\mathrm{obs}}$) are measured from the borehole wireline logs and upscaled to the model resolution vertically. As described in section 2.1.3, prior falsification is then conducted by applying the Robust Mahalanobis Distance outlier detection to $\mathbf{d}$ and $\mathbf{d}_{\mathrm{obs}}$. Figure 8 shows the calculated RMD for $\mathbf{d}_{\mathrm{obs}}$ and the 250 samples of $\mathbf{d}$, where the distribution of the calculated $\mathrm{RMD}(\mathbf{d})$ falls to a Chi-Squared distribution, with the $\mathrm{RMD}(\mathbf{d}_{\mathrm{obs}})$ falls below the 97.5 percentile threshold. This shows with (97.5) confidence that the prior model is not wrong.

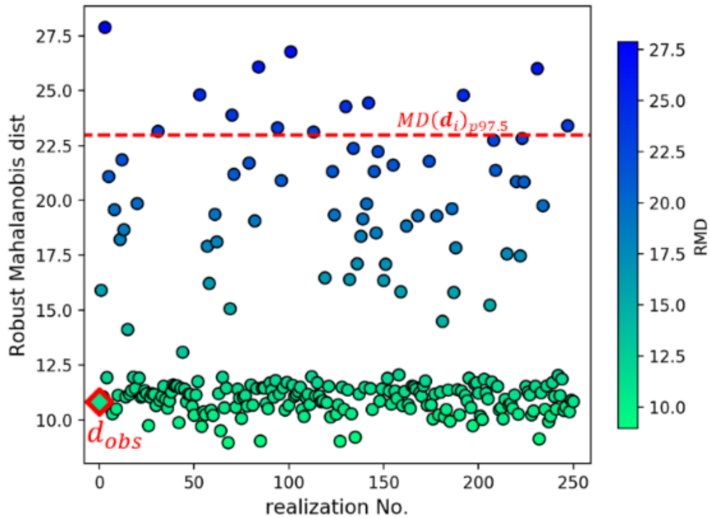

Figure 8. Prior falsification using robust Mahalanobis Distance (RMD). Circle dots represent the caculated RMD for data variable samples. The red-squared dot is the RMD for borehole observations. The red dash line is the 97.5 percentile of the Chi-Squared distributed RMD.

3.4 Automatic updating of uncertainty with new boreholes

After attempting to falsify the prior uncertainty model, we use the automated framework to jointly update model uncertainty with the new boreholes. The joint model uncertainty reduction is performed sequentially as explained in section 2.2.2. Under the AutoBEL GitHub repository instruction (https://github.com/sdyinzhen/AutoBEL-v1.0/blob/master/README.md), we also provide a supplement YouTube video to demonstrate how this automated update is performed.

3.4.1. Thickness and facies

Uncertainty in facies and thickness models can be updated jointly, as they are two independent components for this case. The Auto-BEL first transforms the categorical facies to continuous model using signed distance function. The transformed signed distances are then combined with thickness model to perform orthogonalization using mixed PCA (Abdi et al., 2013). As shown in Figure 9, the first eigen-image (first principal components (PC1)) of thickness reflects the global variations of reservoir thickness, while higher order eigen-images (e.g. eigen-image of PC40) represent more local variation features. To evaluate what model variables impact thickness variation at the boreholes, DGSA (Fenwick et al., 2014) is then performed to analyze the sensitivity of model variables to data. Figure 10(a) plots the main effects in a Pareto plot. As shown in the plot, DGSA identifies sensitive (measure of sensitivity >1) and non-sensitive (measure of sensitivity <1) model variables. Thickness global parameters of both trend (Z_{mean}, T_{range}, T_{sill}) and residuals (R_{range}) show sensitivity to the borehole data. In terms of facies, proportions of the facies 1 (fac1) and 2 (fac2) are sensitive. There are totally 26 sensitive principal components from the spatial model. These sensitive global variables and principal component scores are now selected for uncertainty quantification.

Following the steps of direct forecasting (see section 2.2.1), uncertainty reduction proceeds by mapping all sensitive model variables into a lower dimensional space such that the "Bayes-linear-Gauss" model can be applied. This requires the application of CCA to the selected model variables and data variables, then normal score transformation. Figure 10b shows two examples of cross-plot between model and data variables of the first and tenth canonical components, where we observe linear correlation coefficient of 0.84 even for the tenth canonical components. Once the Bayesian model is specified, one can sample from the posterior distribution and back-transform from lower-dimensional scores into actual facies and thickness models. Figure 10c shows the distribution of the posterior model realizations in comparison to the corresponding prior, showing the reduction of the model uncertainty. Figure 10d shows the comparison between the prior and posterior distributions of the scores for the first 4 sensitive PCs, where the reduction of uncertainty is observed (while noting that uncertainty quantification involves all the sensitive PC score variables).

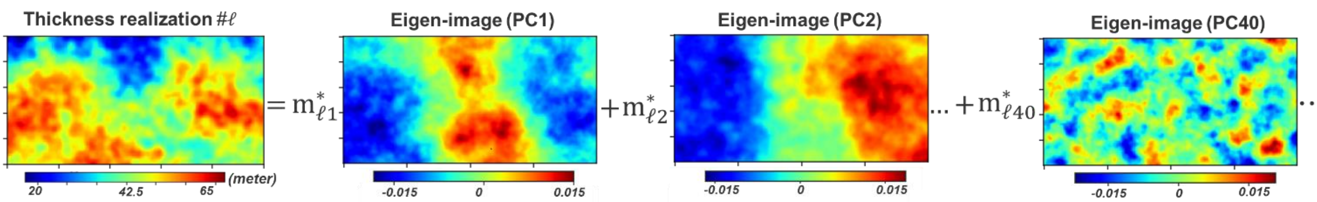

Figure 9. Example of applying PCA on thickness model. One model realization ℓ ($\ell = 1, 2, \ldots, L$) can be represented by the linear combination of eigen-images scaled by the PC scores $\mathbf{m}^*_{\ell}$.

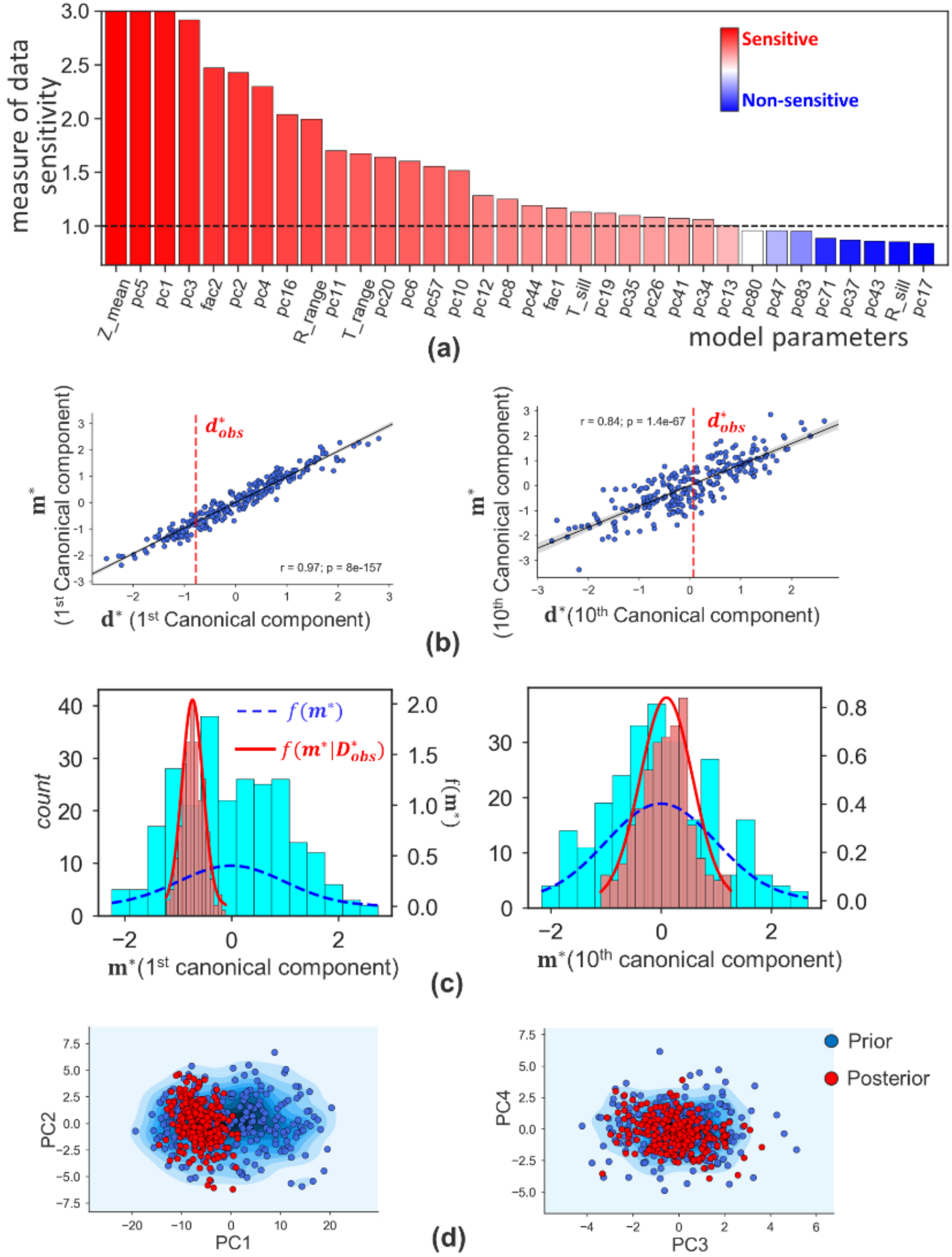

Figure 10. Uncertainty reduction of thickness and facies: (a) global sensitivity of model parameters to borehole data. (b) First and tenth canonical covaraites of data and model variables. The dash redline is the observation data. (c) Posterior and prior distributions of model variables (first and tenth canonical components, corresponding to b). (d) Prior and posterior PC score distributions of first 4 sensitive PCs.

Figure 11 plots the reconstructed posterior global parameters in comparison to the prior. Uncertainty reduction of sensitive global parameters is observed, while the distribution of non-sensitive global parameters (R_{sill} and fac3) is unchanged. To assess the reconstructed posterior spatial model realizations, we calculate the mean for thickness (namely “ensemble mean”), and the median realization of facies. Variance is also calculated for thickness and facies respectively (“ensemble variance”). Figure 12 shows show the ensemble mean and median of the thickness and facies realizations, while the ensemble variances is shown in Figure 13. The results in Figure 12 imply that the posterior model thickness is thicker on average than the prior. This change mainly occurs in areas where the new boreholes are drilled. Referring to the actual borehole observations plotted on Figure 12, we also find that the posterior thickness adjusts to the borehole observations at both training (w7-w13) and validating (w5, w6) locations. This improvement is significant compared to the prior model. Furthermore, the ensemble variances (Figure 13) are reduced in the posterior model, mostly in vicinity of the new boreholes. This implies reduction of the spatial uncertainty. One should note that our method does (not yet) result in an exact match of the thickness at borehole data. This is an issue we will comment on in the discussion and conclusion section. For the facies model, the magnitudes of the uncertainty reduction are not as remarkable, because prior uncertainty at borehole locations was small to start with.

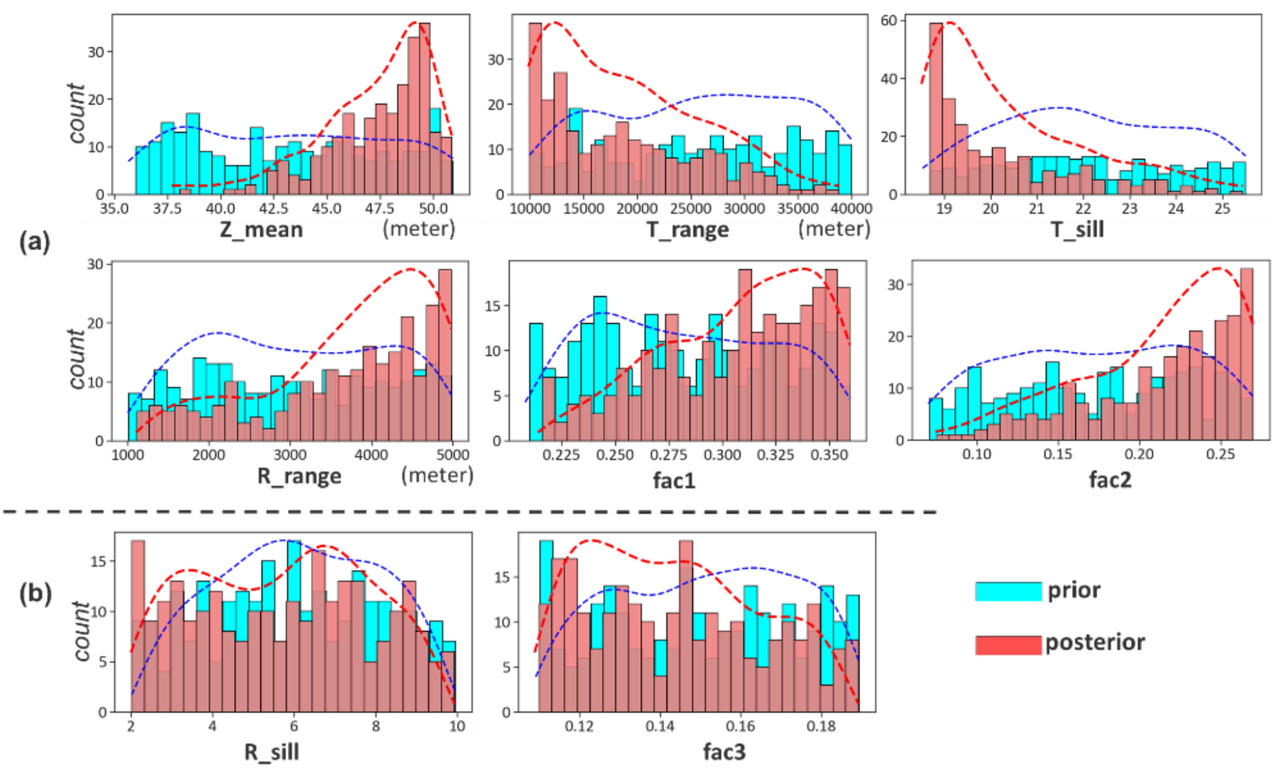

Figure 11. Uncertainty updating of (a) sensitive, and (b) non-sensitive global model parameters at the first sequence. The dashed lines are estimated kernel density with Gaussian kernels.

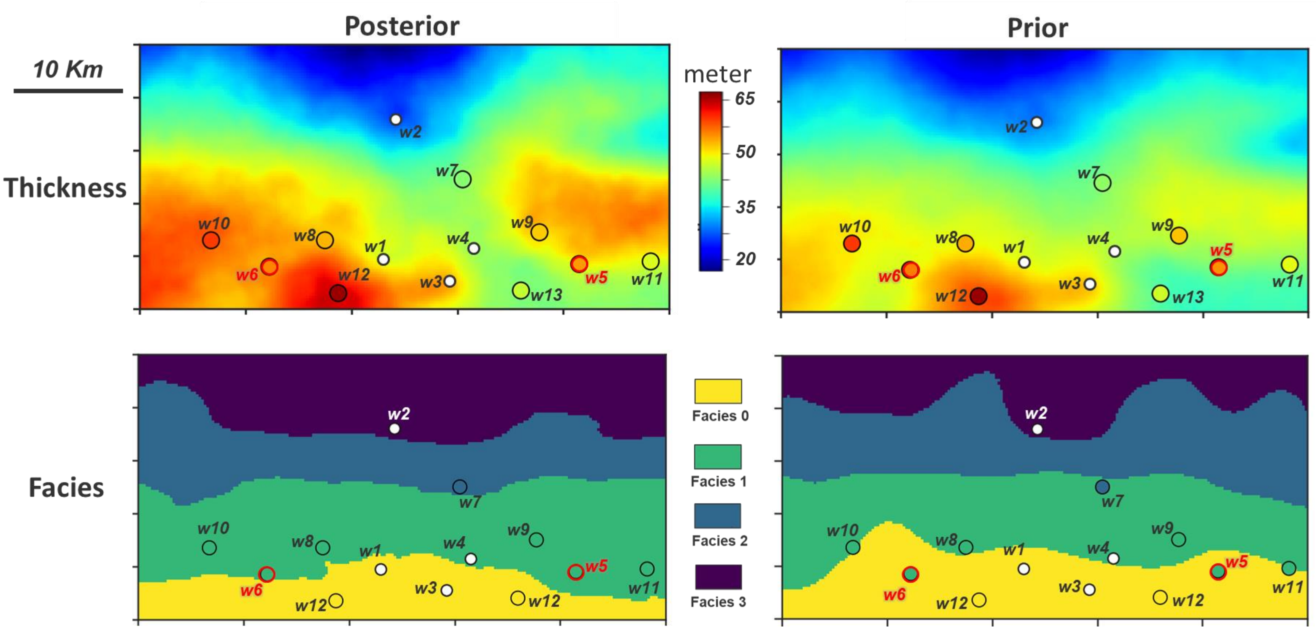

Figure 12. First row: ensemble mean of posterior and prior thickness. Second row: the median realization of posterior and prior facies. The dots are borehole locations and their color represent the actual borehole observation values. The boreholes and models share the same color legend.

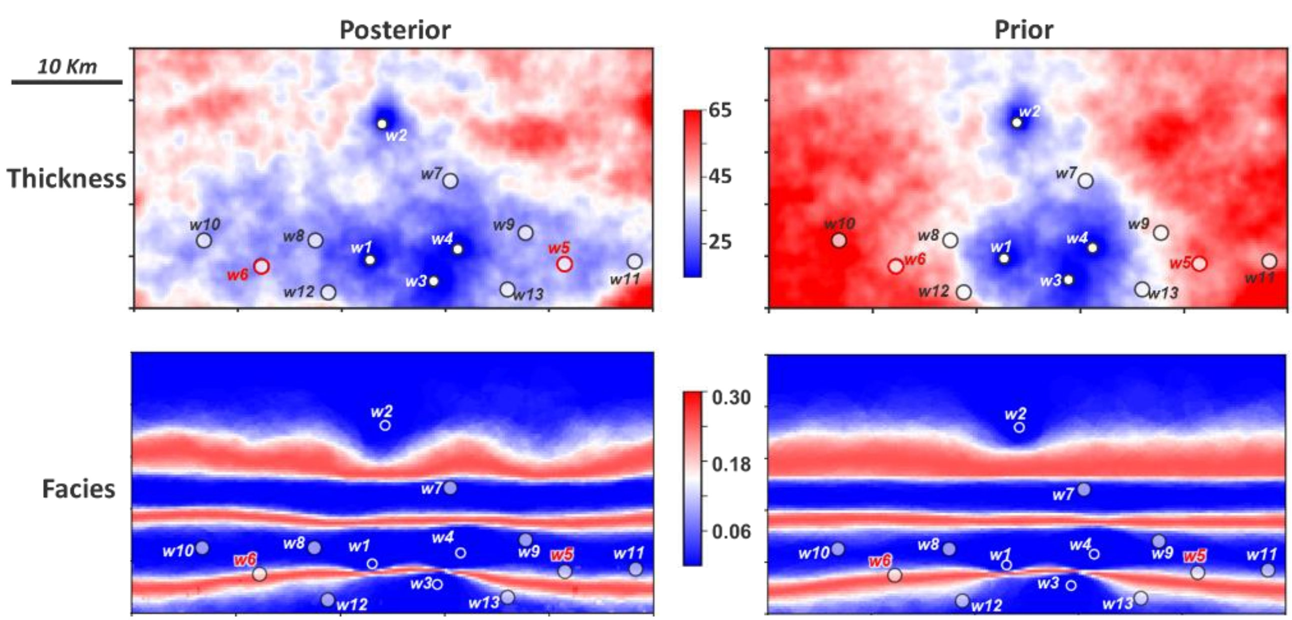

Figure 13. Ensemble variance of the posterior and prior thickness and facies models from the first sequence.

3.4.2 Porosity, permeability and saturation

Auto-BEL is now applied to update uncertainty of porosity, permeability, and saturation under the sequentially decomposition.

The prior Monte Carlo samples have provided a full distribution of porosity for each facies. This allows for the calculation of posterior porosity to fit the obtained posterior facies models. Therefore, we condition to posterior facies model and borehole porosity observations in Auto-BEL to calculate the posterior porosity. Similarly, for permeability and saturation model, the Auto-BEL is applied by additionally conditioning to posterior models from previous model components.

Figure 14, Figure 15 and Figure 16 show the results. In Figure 14, we see sensitive global and spatial model variables that are selected for uncertainty reduction. Figure 15 shows the constructed the linear correlation between data and sensitive model variables by means of CCA. Figure 16 plots the posterior model realizations (250 realizations) computed from the "Bayes-linear-Gauss" model, where reduced uncertainty is observed when comparing to the prior. The posterior spatial model PC scores are also plotted in Figure 17.

Finally, by back-transformation, we can reconstruct all original model variables. Figure 18 compares ensemble means and variances of the reconstructed posterior porosity, log-perm and Sw, to their corresponding prior models, with actual borehole observations plotted on the top. Taking w7 for example, the actual borehole observations show low values of porosity, permeability and Sw, while the prior model initially expects those values to be large at this location. This is adjusted in the posterior. From the ensemble variance maps, we notice that spatial uncertainty is significantly reduced from prior to posterior in areas near w7. The updates of model expectations and reduction of spatial uncertainty are also observed from the other wells. It implies that the posterior models have been constrained by the borehole observations.

Figure 19 shows one example realization of the spatial models. It shows that, same as the hierarchical order in the prior (Figure 19a), the spatial distributions of posterior porosity and log-perm follow the spatial patterns of their corresponding facies belts (Figure 19b). However, if the joint model uncertainty reduction is performed without the sequential decomposition (not conditioning to the posterior models from previous sequences), the model hierarchy from facies to porosity and permeability is lost (marked by the purple boxes on Figure 19c). This is because they are treated as independent model variables, which violates the imposed geological order of variables. The linear correlation between porosity and log-perm is also preserved due to the sequential decomposition. We observe similar correlation coefficients from prior (Figure 20a) to posterior (Figure 20b). But without sequential decomposition, this important feature cannot be maintained as the results shown from Figure 20c: 1) the four clouds pattern (representing the four facies) of the covariate distribution between porosity and log-perm is lost; 2) the correlation coefficient has changed significantly for facies 0, 2 and 3.

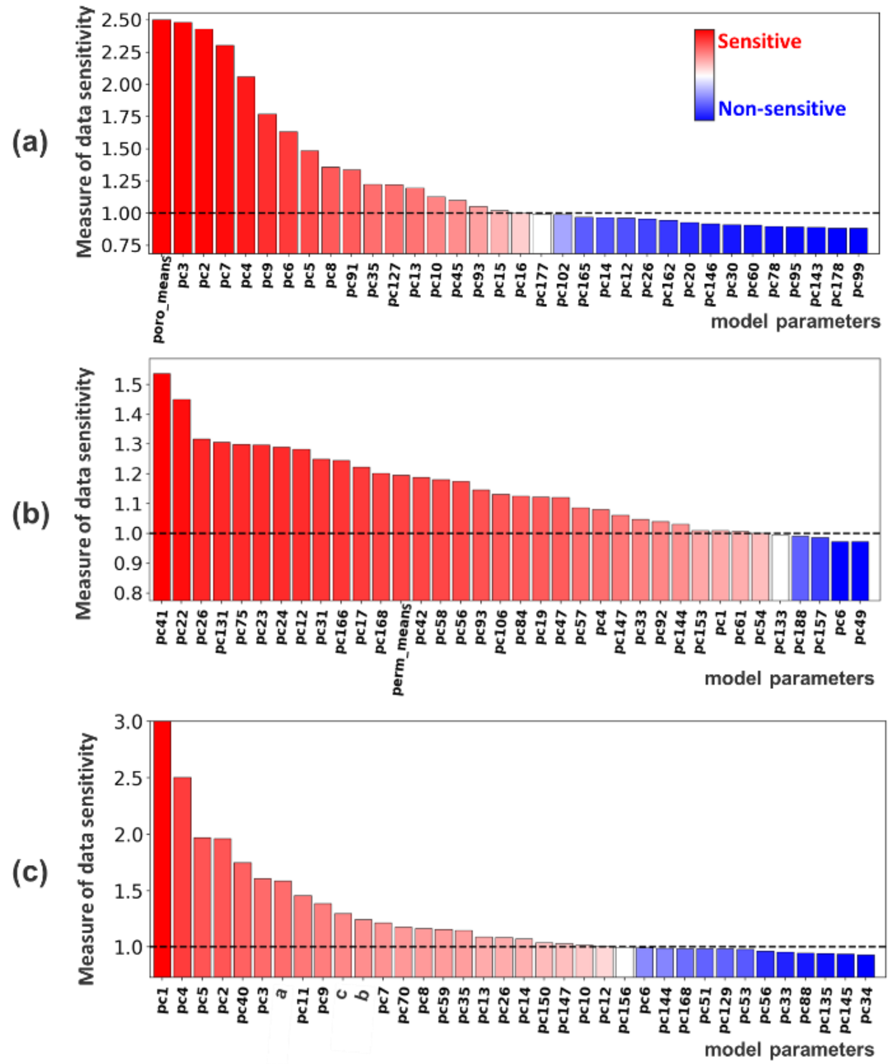

Figure 14. Results from global sensitivity analysis using DGSA at (a) porosity (b) log-perm and (c) Sw.

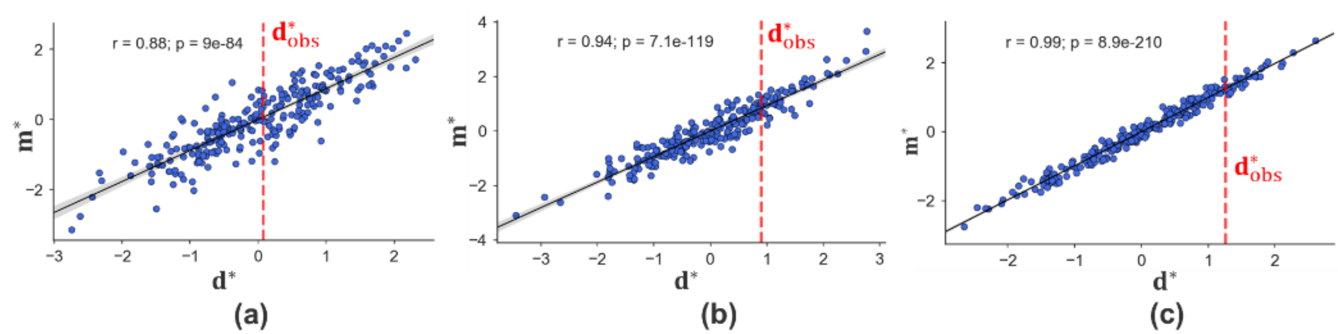

Figure 15 First canonical covariates of data and model variables from (a) porosity (b) log-perm and (c) Sw.

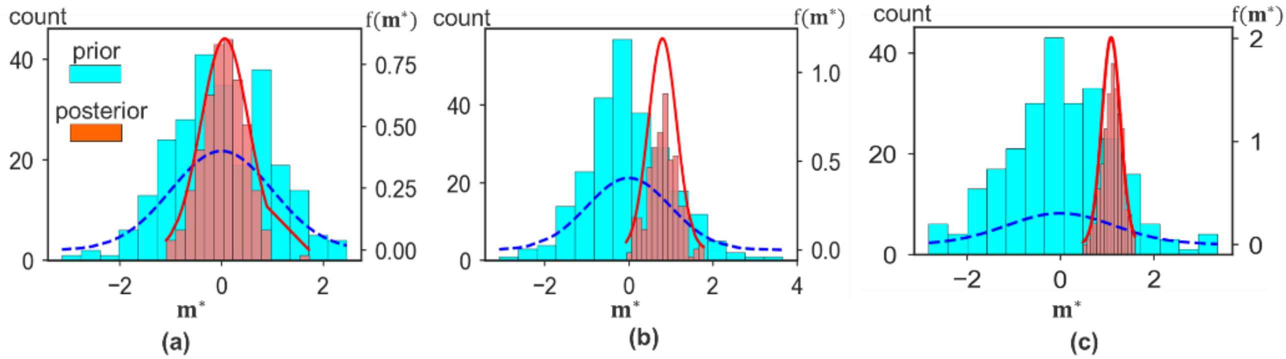

Figure 16. Reduction of uncertainty of the first model canonical component: (a) porosity (b) log-perm and (c) Sw

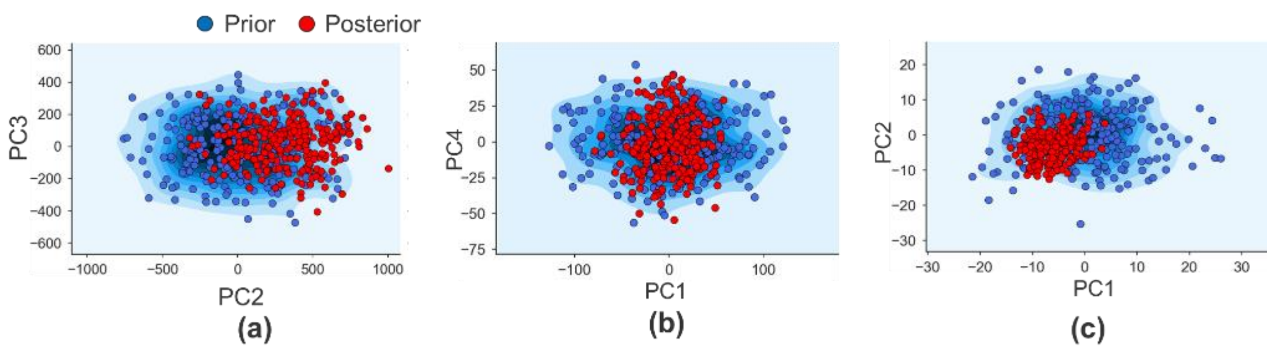

Figure 17. Prior and posterior distribution of the scores of the two sensitive PCs with highest variances: (a) porosity (b) log-perm and (c) Sw.

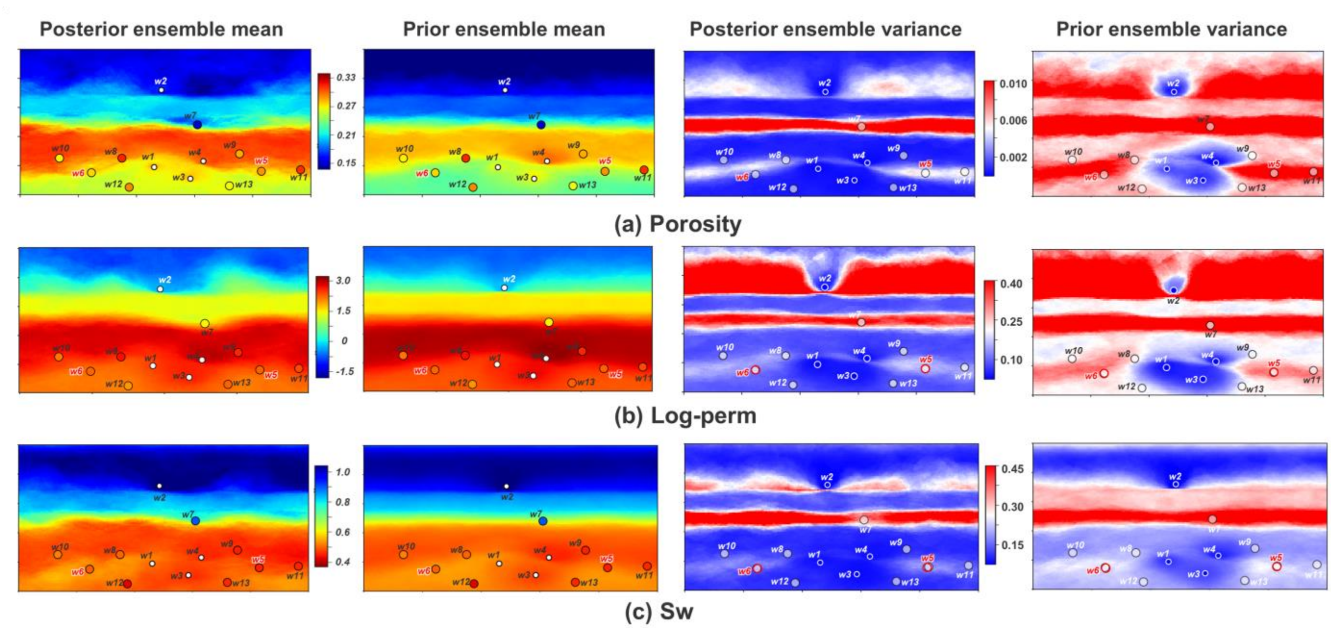

Figure 18. Ensemble mean and variance of posterior and prior geological models: (a) porosity; (b) log-perm; (c) Water saturation. The dots represents locations of the boreholes, where color of the dots represents observation values.

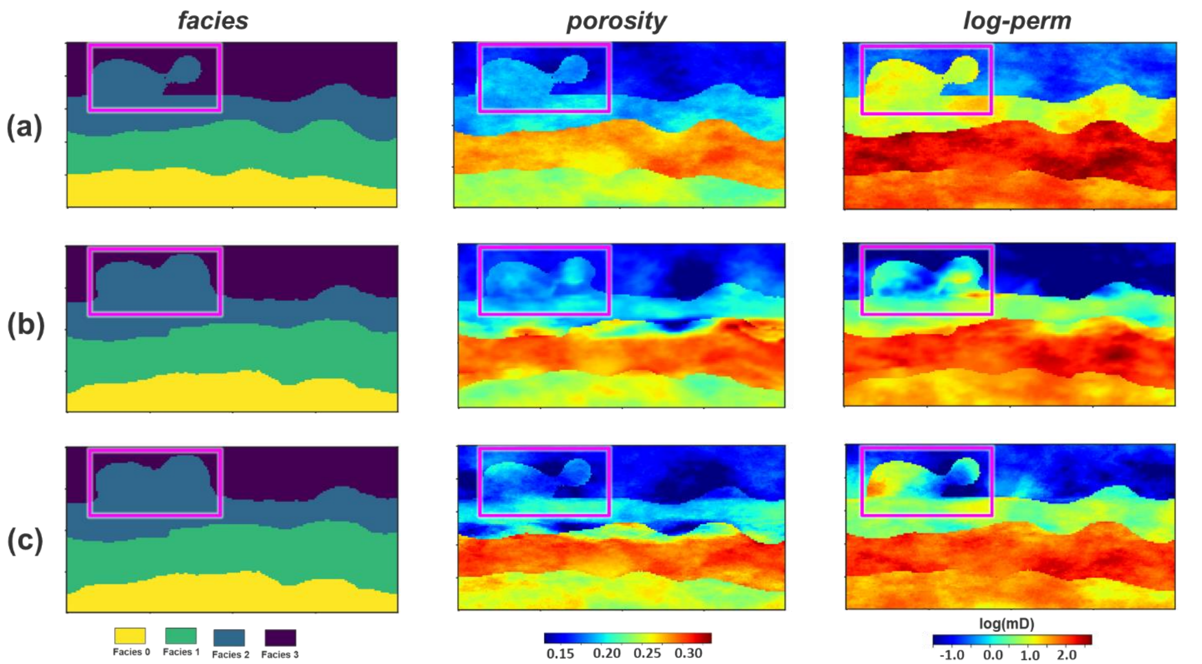

Figure 19. Prior and posterior facies, porosity and log-perm of realization #1 (a) prior model; (b) posterior model from the sequential decomposition; (c) posterior from joint uncertainty reduction without sequential decomposition.

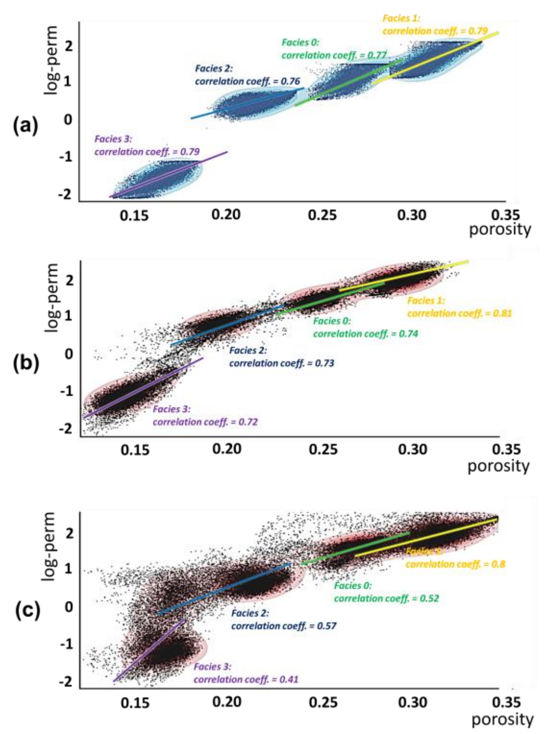

Figure 20.Bivarite distribution between porosity and log-perm model of realization #1 (a) prior; (b) posterior from the sequential decomposition; (c) posterior without performing sequential decomposition. The correlation coefficeint is examined for each facies.

3.4.3 Posterior prediction and falsification

Gas storage volume is calculated using the posterior geological models and plotted in Figure 21. The result highlights a steep uncertainty reduction in comparison to the initial prior prediction. The posterior predicted GIIP leads to a major shift of the expected gas volumes to a more positive direction (higher than initially expected). More importantly, the forecast range is significantly narrowed. This provides critical guidance to the financial decisions on the field development. It also in return confirms the value of the information of the newly drilled wells. In total, the whole application of "Auto-BEL" to this test case took about 45 minutes (not including the time on prior modeling) when running on a laptop with an Intel Core i7-7820HQ processor and 64 GB of Ram.

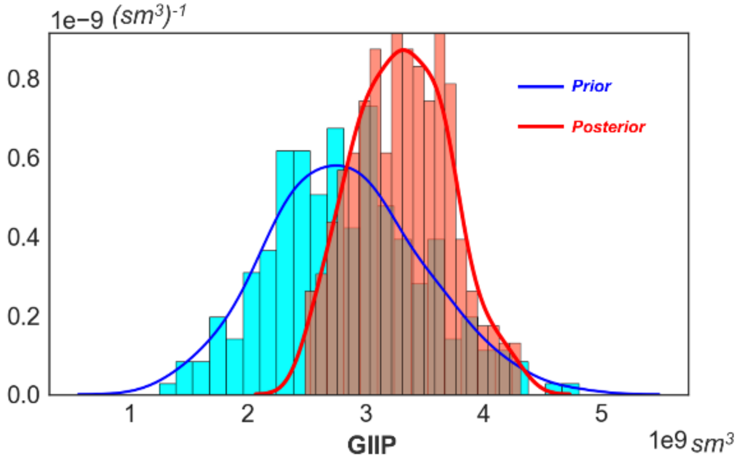

Figure 21.The prior and posterior prediction of GIIP

To test the posterior, we perform posterior falsification using data from validating boreholes (w5 and w6). Figure 22 plots the result from applying Robust Mahalanobis Distance outlier detection to the posterior data of the two wells. The statistical test shows that the test borehole observation falls within the main population of data variables, significantly below the 97.5 threshold percentile. We also want to further examine if the posterior models can predict the validating boreholes (regarded as future drilling wells) with reduced uncertainty. To do so, we compare the prior and posterior predicted thickness at the two borehole locations, together with their actual measurements (Figure 23). For 3D models of facies, porosity, log-perm and Sw, this comparison is performed on vertical average values across the 75 layers. We notice that these future borehole observations are predicted by posterior models with significantly reduced uncertainty.

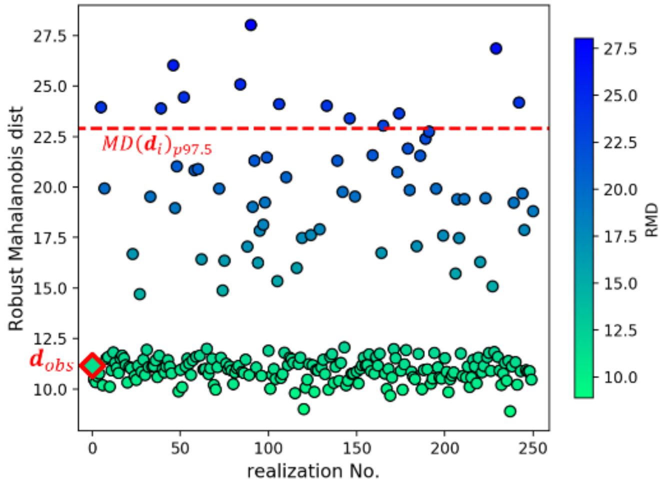

Figure 22. Posterior falsification using the Robust Mahalanobis Distance outlier detection method using the data from (w5 and w6).

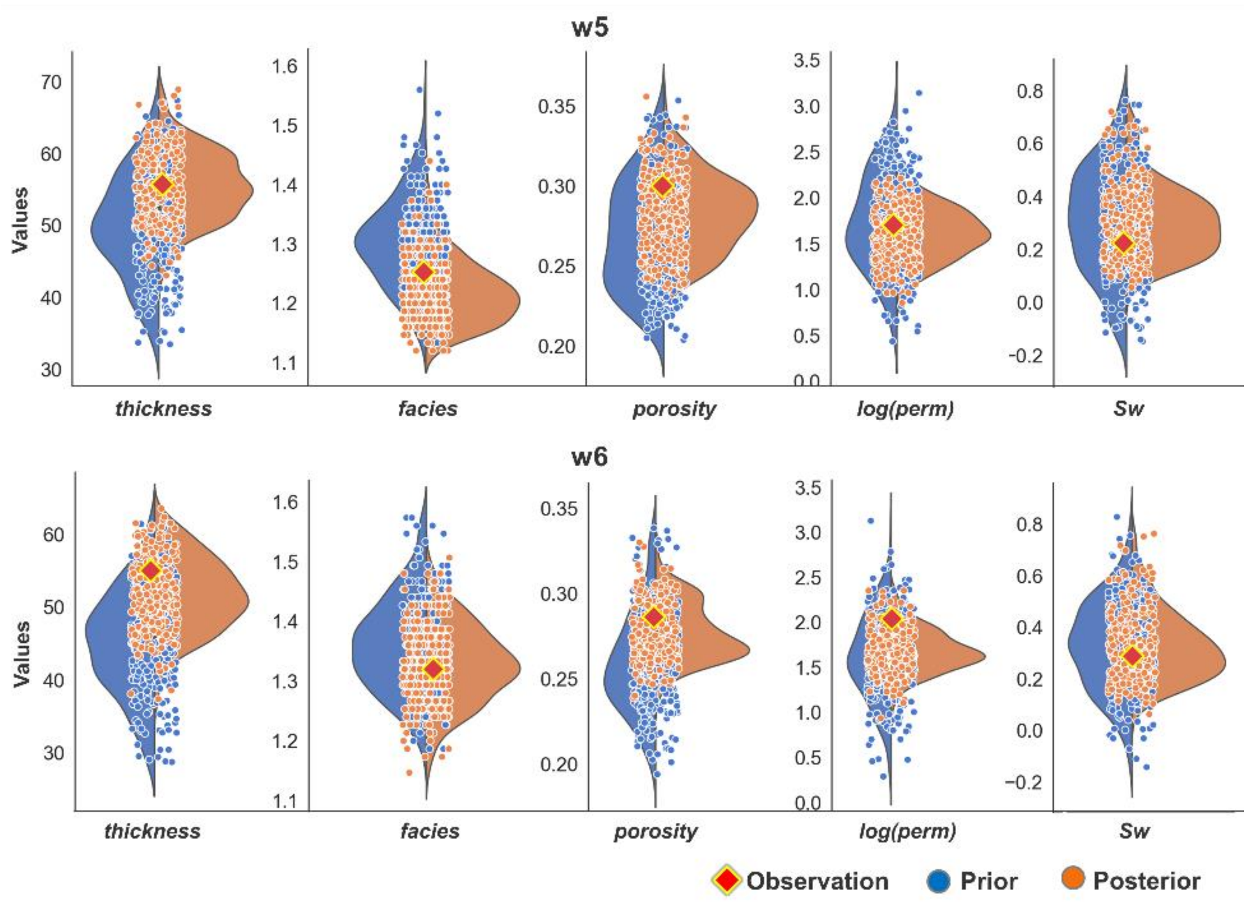

Figure 23. Prior and posterior predicted thickness, facies, porosity, log-perm and Sw at validating boreholes. The bule and brown colored dots respectively represents the prior and posterior prediction, while red squared dots are the actural observations.

4. Discussion

One main purpose of this paper is to introduce automation to geological uncertainty quantification when new borehole data is acquired. We tackle this challenge by following the protocol of Bayesian Evidential Learning to build an automated UQ framework. BEL formulates a protocol involving falsification, global sensitivity analysis, and statistical learning uncertainty reduction. When establishing such framework for geological UQ, three important questions have to be addressed. The first is on how to preserve the hierarchical relationships and correlations that commonly exist in geological models. We propose a sequential decomposition by following the chain rule under Bayes' theorem. This allows to assess the joint distribution of multiple model components while honoring the geological rules. The second one is on how to falsify the geological model hypotheses, especially when data becomes highly dimensional. We employ multivariate outlier detection methods. They provide quantitative and robust statistical calculations when attempting to falsify the model using high dimensional data. The last but most practical one, is to deploy data-science driving uncertainty reduction. Uncertainty reduction of geological model is usually time-consuming, because conventional inverse methods require iterative model rebuilding. When it comes to real cases, the daunting time-consumption and computational efforts of conventional methods can hamper practical implementations of automation. Direct forecasting helps to avoid this, as it mitigates the uncertainty reduction to a linear problem in much lower dimension. There are many dimension reduction methods for complex models such as deep neural network (Laloy et al., 2017, 2018), but here we use PCA because it is simple and bijective, and the structure models are not complex (e.g. channels). However, direct forecasting of geological model is faced with two new challenges. One is to accommodate direct forecasting algorithm to the sequential model decomposition. This is achieved by additionally conditioning to the posterior from previous sequences. The other challenge is that DF cannot be directly applied to categorical models such as lithological facies. We therefore introduce signed distance function to convert categorical models to continuous properties before performing the DF. Field application has shown benefits of using the proposed framework. Since the posterior in the case study cannot be falsified, its uncertainty can be further reduced by repeating the automated procedures with the validating borehole observations. This suggests that the proposed framework has potentials for life-of-field uncertainty quantification for applications where new boreholes are regularly drilled.

The main challenge addressed in this paper is to apply such uncertainty quantification within a Bayesian framework. Most method applied in this context simply rebuild the models by repeating the same geostatistical methods that were used to construct the prior model. In such approach, all global variables and their uncertainty need to be re-assessed. The problem with such approach is twofold. First, it does not address the issue of falsification: the original models may not be able to predict the data. Hence, using the same approach to update models with a prior that may have been falsified may lead again to falsification, thereby leading to invalid and ineffective uncertainty quantification. As a result, the uncertainty quantification of some desirable property, such as volume exhibits a yoyo effect (low variance in each UQ but shifting mean). Second, there is no consistent updating of global model variables. Often such uncertainties are assessed independently of previous uncertainties. The challenge addressed in this paper is to jointly update global and spatial variables and do this jointly for all properties.

The proposed method offers a Bayesian consistency to uncertainty quantification in the geological modeling setting. However, unlike geostatistical methods, the posterior models do not fully match to local borehole observations. The current method is only designed to globally adjust the model, not locally at the borehole observation. This can be an important issue if using the model for subsurface flow simulations. To tackle this problem, one possible path we like to explore in the future is to combine geostatistical conditional simulation as posterior step to the current methodology. A second limitation is that the method does (not yet) treat discrete global variables, such as a geological interpretation. In the cased study, only one interpretation of the lithology was used. The way such variables would be treated is by assigning prior probabilities to each interpretation (e.g. of depositional system), then updating them into posterior probabilities. This has been done by treating the interpretation independent of other model variables in some studies (e.g. Aydin and Caers, 2017; Grose et al., 2018; Wellmann et al., 2010). For example, one could first update the probabilities of geological scenarios, then update the other variables (Lopez-Alvis et al., 2019). Regarding the automation of BEL, its intermediate steps can also be adjusted depending on users' specific applications. Taking the direct forecasting step for example, here we adapt it for uncertainty quantification using borehole data, which is a linear problem. But for more complex non-linear inverse problems, it may be difficult to use CCA to derive a "Bayes-linear-Gauss" relationship in DF. Statistical estimation approaches such as kernel density estimation (Lopez-Alvis et al., 2019) can be used for such cases, and there are also extensions of CCA to tackle non-linear problems(e.g., Lai and Fyfe, 1999). AutoBEL can also be adapt if other types of parameters (other than spatial model parameters) are used for uncertainty quantification. This can be done by simply adding the additional parameters to the model variable **m**. A final, and perhaps more fundamental concern not limited to our approach is on what should be done when the prior model is falsified with new data. According to the Bayesian philosophy this would mean that any of the following could have happened: uncertainty ranges are too small; model is too simple or some combination of both. The main problem is that it is difficult to assess what the problem is exactly. Our future work will focus on this issue.

5. Conclusions

In conclusion, we generalized a Monte Carlo-based framework for geological uncertainty quantification and updating. This framework, based on Bayesian Evidential Learning, was demonstrated under the context of geological model updating using borehole data. Within the framework, a sequential model decomposition was proposed, to address the geological rules when assessing joint uncertainty distribution of multiple model components. For each component, we divided model parameters into global and spatial ones, thus facilitating the uncertainty quantification of complex spatial heterogeneity. When new borehole observations are measured, instead of directly reducing model uncertainty, we first strengthen the model hypothesis by attempting to falsify it via statistical tests. Our second contribution was to show how direct forecasting can jointly reduce model uncertainty under the sequential decomposition. This requires posterior model from previous sequences as additional inputs to constrain the current prior. Such sequential direct forecasting showed to maintain important geological model features of hierarchy and correlation, whilst avoiding the time-consuming conventional model rebuilding. In terms of discrete model such as lithology, signed distance function was employed, before applying directly forecasting to reduce uncertainty. The third contribution, but maybe more important, is that the proposed framework allows automation of geological UQ. We developed an opensource Python project for this implementation. Its application to a large reservoir model showed that the automated framework ensures the model is objectively informed by data at each step of uncertainty quantitation. It jointly quantified and updated uncertainty of all model components, including structural thickness, facies, porosity, permeability and water saturation. The posterior model showed to be constrained by new borehole observations globally and locally, with dependencies and correlations between the model components preserved from the prior. It predicted validating observations (future drilling boreholes) with reduced uncertainty. Since posterior cannot be falsified, the uncertainty reduced GIIP prediction can be used for decision makings. The whole process takes less than 1 hour on a laptop workstation for this large field case, thus demonstrating efficiency of the automation

Code availability.

AutoBEL is a free, open-source Python library. It is available at https://doi.org/10.5281/zenodo.3479997, and the source code is maintained on GitHub https://github.com/sdyinzhen/AutoBEL-v1.0 under MIT license.

Author contributions

Zhen Yin: contributed the concept and methodology development; wrote and maintained the code; conducted the technical application and drafted this paper.

Sebastien Strebelle: prepared data for the methodology application and provided critical insights during the research developments.

Jef Caers: provided overall supervision and funding to this project; contributed major and critical ideas to the research development and revised the manuscript.

Competing interests

The authors declare that they have no conflict of interests.

Acknowledgements

We thank Chevron for sponsoring this research project. The authors would like to specially thank Maisha Amaru, Jairam Kamath, Lewis Li, Sarah Vitel, Lijing Wang and Celine Scheidt for the technical discussions and supports.

References

Abdi, H., Williams, L. J. and Valentin, D.: Multiple factor analysis: principal component analysis for multitable and multiblock data sets, Wiley Interdiscip. Rev. Comput. Stat., 5(2), 149–179, doi:10.1002/wics.1246, 2013.

Athens, N. D. and Caers, J. K.: A Monte Carlo-based framework for assessing the value of information and development risk in geothermal exploration, Appl. Energy, 256, 113932, doi:10.1016/J.APENERGY.2019.113932, 2019a.

Athens, N. D. and Caers, J. K.: A Monte Carlo Based Framework for Geothermal Risk Assessments, Appl. Energy, Paper Acce, 2019b.

Aydin, O. and Caers, J. K.: Quantifying structural uncertainty on fault networks using a marked point process within a Bayesian framework, Tectonophysics, doi:10.1016/j.tecto.2017.04.027, 2017.

Barfod, A. A. S., Møller, I., Christiansen, A. V, Høyer, A.-S., Hoffimann, J., Straubhaar, J. and Caers, J.: Hydrostratigraphic modeling using multiple-point statistics and airborne transient electromagnetic methods, Hydrol. Earth Syst. Sci., 22(6), 3351–3373, doi:10.5194/hess-22-3351-2018, 2018.

Beucher, H., Galli, A., Le Loc'h, G. and Ravenne, C.: Including a Regional Trend in Reservoir Modelling Using the Truncated Gaussian Method, in Geostatistics Tróia '92: Volume 1, edited by A. Soares, pp. 555–566, Springer Netherlands, Dordrecht., 1993.

Beven, K.: Prophecy, reality and uncertainty in distributed hydrological modelling, Adv. Water Resour., 16(1), 41–51, doi:10.1016/0309-1708(93)90028-E, 1993.

Le Borgne, T., Bour, O., Paillet, F. L. and Caudal, J.-P.: Assessment of preferential flow path connectivity and hydraulic properties at single-borehole and cross-borehole scales in a fractured aquifer, J. Hydrol., 328(1–2), 347–359, doi:10.1016/J.JHYDROL.2005.12.029, 2006.

Caers, J. and Zhang, T.: Multiple-point Geostatistics: A Quantitative Vehicle for Integrating Geologic Analogs into Multiple Reservoir Models, edited by G. M. Grammer, P. M. "Mitch" Harris, and G. P. Eberli, Integr. Outcrop Mod. Analog. Reserv. Model., 0, doi:10.1306/M80924C18, 2004.

Caumon, G.: Towards Stochastic Time-Varying Geological Modeling, Math. Geosci., 42(5), 555–569, doi:10.1007/s11004-010-9280-y, 2010.

Caumon, G.: Geological Objects and Physical Parameter Fields in the Subsurface: A Review, in Handbook of Mathematical Geosciences: Fifty Years of IAMG, edited by B. S. Daya Sagar, Q. Cheng, and F. Agterberg, pp. 567–588, Springer International Publishing, Cham., 2018.

Christie, M., MacBeth, C. and Subbey, S.: Multiple history-matched models for Teal South, Lead. Edge, 21(3), 286–289, doi:10.1190/1.1463779, 2002.

Cullen, N. J., Anderson, B., Sirguey, P., Stumm, D., Mackintosh, A., Conway, J. P., Horgan, H. J., Dadic, R., Fitzsimons, S. J. and Lorrey, A.: An 11-year record of mass balance of Brewster Glacier, New Zealand, determined using a geostatistical approach, J. Glaciol., 63(238), 199–217, doi:10.1017/jog.2016.128, 2017.

Curtis, J. B.: Fractured Shale-Gas Systems, Am. Assoc. Pet. Geol. Bull., 86(11), 1921–1938, doi:10.1306/61EEDDBE-173E-11D7-8645000102C1865D, 2002.

Deutsch, C. V. and Wilde, B. J.: Modeling multiple coal seams using signed distance functions and global kriging, Int. J. Coal Geol., 112, 87–93, doi:10.1016/J.COAL.2012.11.013, 2013.

Dutta, G., Mukerji, T. and Eidsvik, J.: Value of information analysis for subsurface energy resources applications, Appl. Energy, 252, 113436, doi:10.1016/J.APENERGY.2019.113436, 2019.

Eidsvik, J. and Ellefmo, S. L.: The Value of Information in Mineral Exploration Within a Multi-Gaussian Framework, Math. Geosci., 45(7), 777–798, doi:10.1007/s11004-013-9457-2, 2013.

Elfeki, A. M. M. and Dekking, F. M.: Reducing geological uncertainty by conditioning on boreholes: the coupled Markov chain approach, Hydrogeol. J., 15(8), 1439–1455, doi:10.1007/s10040-007-0193-x, 2007.

Fenwick, D., Scheidt, C. and Caers, J.: Quantifying Asymmetric Parameter Interactions in Sensitivity Analysis: Application to Reservoir Modeling, Math. Geosci., 46(4), 493–511, doi:10.1007/s11004-014-9530-5, 2014.

Fischer, T., Naumov, D., Sattler, S., Kolditz, O. and Walther, M.: GO2OGS 1.0: a versatile workflow to integrate complex geological information with fault data into numerical simulation models, Geosci. Model Dev., 8(11), 3681–3694, doi:10.5194/gmd-8-3681-2015, 2015.

Goovaerts, P.: Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall, J. Hydrol., 228(1–2), 113–129, doi:10.1016/S0022-1694(00)00144-X, 2000.

Goovaerts, P. and Goovaerts, D. C. E. E. P.: Geostatistics for Natural Resources Evaluation, Oxford University Press. [online] Available from: https://books.google.com/books?id=CW-7tHAaVR0C, 1997.

Grose, L., Laurent, G., Aillères, L., Armit, R., Jessell, M. and Cousin-Dechenaud, T.: Inversion of Structural Geology Data for Fold Geometry, J. Geophys. Res. Solid Earth, 123(8), 6318–6333, doi:10.1029/2017JB015177, 2018.

Hermans, T., Oware, E. and Caers, J.: Direct prediction of spatially and temporally varying physical properties from time-lapse electrical resistance data, Water Resour. Res., 52(9), 7262–7283, doi:10.1002/2016WR019126, 2016.

Hermans, T., Nguyen, F., Klepikova, M., Dassargues, A. and Caers, J.: Uncertainty Quantification of Medium-Term Heat Storage From Short-Term Geophysical Experiments Using Bayesian Evidential Learning, Water Resour. Res., 54(4), 2931–2948, doi:10.1002/2017WR022135, 2018.

Hermans, T., Lesparre, N., De Schepper, G. and Robert, T.: Bayesian evidential learning: a field validation using push-pull tests, Hydrogeol. J., 27(5), 1661–1672, doi:10.1007/s10040-019-01962-9, 2019.

Hoffmann, R., Dassargues, A., Goderniaux, P. and Hermans, T.: Heterogeneity and prior uncertainty investigation using a joint heat and solute tracer experiment in alluvial sediments, Front. EARTH Sci., 7, 15 [online] Available from: http://dx.doi.org/10.3389/feart.2019.00108, 2019.

Hubert, M. and Debruyne, M.: Minimum covariance determinant, Wiley Interdiscip. Rev. Comput. Stat., 2(1), 36–43, doi:10.1002/wics.61, 2010.

Jiménez, S., Mariethoz, G., Brauchler, R. and Bayer, P.: Smart pilot points using reversible-jump Markov-chain Monte Carlo, Water Resour. Res., 52(5), 3966–3983, doi:10.1002/2015WR017922, 2016.

Journel, A. G.: Markov Models for Cross-Covariances, Math. Geol., 31(8), 955–964, doi:10.1023/A:1007553013388, 1999.

Kaufmann, O. and Martin, T.: 3D geological modelling from boreholes, cross-sections and geological maps, application over former natural gas storages in coal mines, Comput. Geosci., 34(3), 278–290, doi:10.1016/J.CAGEO.2007.09.005, 2008.

Klepikova, M. V., Le Borgne, T., Bour, O. and Davy, P.: A methodology for using borehole temperature-depth profiles under ambient, single and cross-borehole pumping conditions to estimate fracture hydraulic properties, J. Hydrol., 407(1–4), 145–152, doi:10.1016/J.JHYDROL.2011.07.018, 2011.

de la Varga, M., Schaaf, A. and Wellmann, F.: GemPy 1.0: open-source stochastic geological modeling and inversion, Geosci. Model Dev., 12(1), 1–32, doi:10.5194/gmd-12-1-2019, 2019.

Lai, P. . and Fyfe, C.: A neural implementation of canonical correlation analysis, Neural Networks, 12(10), 1391–1397, doi:10.1016/S0893-6080(99)00075-1, 1999.

Laloy, E., Hérault, R., Lee, J., Jacques, D. and Linde, N.: Inversion using a new low-dimensional representation of complex binary geological media based on a deep neural network, Adv. Water Resour., 110, 387–405, doi:10.1016/J.ADVWATRES.2017.09.029, 2017.

Laloy, E., Hérault, R., Jacques, D. and Linde, N.: Training-Image Based Geostatistical Inversion Using a Spatial Generative Adversarial Neural Network, Water Resour. Res., 54(1), 381–406, doi:10.1002/2017WR022148, 2018.

Leverett, M. C., Lewis, W. B. and True, M. E.: Dimensional-model Studies of Oil-field Behavior, Trans. AIME, 146(01), 175–193, doi:10.2118/942175-G, 1942.

Li, D.-Q., Qi, X.-H., Cao, Z.-J., Tang, X.-S., Phoon, K.-K. and Zhou, C.-B.: Evaluating slope stability uncertainty using coupled Markov chain, Comput. Geotech., 73, 72–82, doi:10.1016/J.COMPGEO.2015.11.021, 2016.

Liu, F. T., Ting, K. M. and Zhou, Z.: Isolation Forest, in In ICDM '08: Proceedings of the 2008 Eighth IEEE International Conference on Data Mining. IEEE Computer Society, pp. 413–422., 2008.

Lopez-Alvis, J., Hermans, T. and Nguyen, F.: A cross-validation framework to extract data features for reducing structural uncertainty in subsurface heterogeneity, Adv. Water Resour., 133, 103427, doi:10.1016/J.ADVWATRES.2019.103427, 2019.

Marko, K., Al-Amri, N. S. and Elfeki, A. M. M.: Geostatistical analysis using GIS for mapping groundwater quality: case study in the recharge area of Wadi Usfan, western Saudi Arabia, Arab. J. Geosci., 7(12), 5239–5252, doi:10.1007/s12517-013-1156-2, 2014.

Neuman, S. P.: Maximum likelihood Bayesian averaging of uncertain model predictions, Stoch. Environ. Res. Risk Assess., 17(5), 291–305, doi:10.1007/s00477-003-0151-7, 2003.

Osher, S. and Fedkiw, R.: Signed Distance Functions, in Level Set Methods and Dynamic Implicit Surfaces, pp. 17–22, Springer New York, New York, NY., 2003.

Pappenberger, F., Beven, K. J., Ratto, M. and Matgen, P.: Multi-method global sensitivity analysis of flood inundation models, Adv. Water Resour., doi:10.1016/j.advwatres.2007.04.009, 2008.

Park, J.: Uncertainty Quantification and Sensitivity Analysis of Geoscientific Predictions with Data-Driven Approaches, Stanford University., 2019.

Park, J., Yang, G., Satija, A., Scheidt, C. and Caers, J.: DGSA: A Matlab toolbox for distance-based generalized sensitivity analysis of geoscientific computer experiments, Comput. Geosci., 97, 15–29, doi:10.1016/J.CAGEO.2016.08.021, 2016.

Pezeshki, A., Scharf, L. L., Azimi-Sadjadi, M. R. and Lundberg, M.: Empirical canonical correlation analysis in subspaces, in Conference Record of the Thirty-Eighth Asilomar Conference on Signals, Systems and Computers, 2004., vol. 1, pp. 994-997 Vol.1., 2004.

Rousseeuw, P. J. and Driessen, K. Van: A Fast Algorithm for the Minimum Covariance Determinant Estimator, Technometrics, 41(3), 212–223, doi:10.1080/00401706.1999.10485670, 1999.

Royse, K. R.: Combining numerical and cognitive 3D modelling approaches in order to determine the structure of the Chalk in the London Basin, Comput. Geosci., 36(4), 500–511, doi:10.1016/J.CAGEO.2009.10.001, 2010.

Rühaak, W., Guadagnini, A., Geiger, S., Bär, K., Gu, Y., Aretz, A., Homuth, S. and Sass, I.: Upscaling thermal conductivities of sedimentary formations for geothermal exploration, Geothermics, 58, 49–61, doi:10.1016/J.GEOTHERMICS.2015.08.004, 2015.

Satija, A. and Caers, J.: Direct forecasting of subsurface flow response from non-linear dynamic data by linear least-squares in canonical functional principal component space, Adv. Water Resour., 77, 69–81, doi:10.1016/J.ADVWATRES.2015.01.002, 2015.

Satija, A., Scheidt, C., Li, L. and Caers, J.: Direct forecasting of reservoir performance using production data without history matching, Comput. Geosci., 21(2), 315–333, doi:10.1007/s10596-017-9614-7, 2017.

Scheidt, C. Ã., Li, L. and Caers, J.: Quantifying Uncertainty in Subsurface Systems, Wiley. [online] Available from: https://books.google.com/books?id=xvRYDwAAQBAJ, 2018.

Schölkopf, B., Platt, J. C., Shawe-Taylor, J. C., Smola, A. J. and Williamson, R. C.: Estimating the Support of a High-Dimensional Distribution, Neural Comput., 13(7), 1443–1471, doi:10.1162/089976601750264965, 2001.

Sobol, I. .: Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, Math. Comput. Simul., 55(1–3), 271–280, doi:10.1016/S0378-4754(00)00270-6, 2001.

Sobol, I. M.: Sensitivity estimates for nonlinear mathematical models., Math. Model. Comput. Exp., 1(4), 407–414, 1993.

Soltani-Mohammadi, S., Safa, M. and Mokhtari, H.: Comparison of particle swarm optimization and simulated annealing for locating additional boreholes considering combined variance minimization, Comput. Geosci., 95, 146–155, doi:10.1016/J.CAGEO.2016.07.020, 2016.

Spear, R. C. and Hornberger, G. M.: Eutrophication in peel inlet—II. Identification of critical uncertainties via generalized sensitivity analysis, Water Res., 14(1), 43–49, doi:10.1016/0043-1354(80)90040-8, 1980.

Territory, N., Abbott, J. and Geol, B. A.: Technical Report Mineral Resource Estimation for the Wonarah Phosphate Project, , (March), 2013.

Vogt, C., Mottaghy, D., Wolf, A., Rath, V., Pechnig, R. and Clauser, C.: Reducing temperature uncertainties by stochastic geothermal reservoir modelling, Geophys. J. Int., 181(1), 321–333, doi:10.1111/j.1365-246X.2009.04498.x, 2010.

Wei, P., Lu, Z. and Song, J.: Variable importance analysis: A comprehensive review, Reliab. Eng. Syst. Saf., 142, 399–432, doi:10.1016/J.RESS.2015.05.018, 2015.

Wellmann, J. F., Horowitz, F. G., Schill, E. and Regenauer-Lieb, K.: Towards incorporating uncertainty of structural data in 3D geological inversion, Tectonophysics, 490(3–4), 141–151, doi:10.1016/J.TECTO.2010.04.022, 2010.

Yin, Z., Feng, T. and MacBeth, C.: Fast assimilation of frequently acquired 4D seismic data for reservoir history matching, Comput. Geosci., 128, 30–40, doi:10.1016/J.CAGEO.2019.04.001, 2019.

Zhou, H., Gómez-Hernández, J. J. and Li, L.: Inverse methods in hydrogeology: Evolution and recent trends, Adv. Water Resour., 63, 22–37, doi:10.1016/J.ADVWATRES.2013.10.014, 2014.

Part 3.

Responses to Guillaume Caumon (Referee) on interactive comments (RC2)

Guillaume Caumon (Referee)
guillaume.caumon@ensg.univ-lorraine.fr
Received and published: 27 December 2019

General comments

This paper proposes an application of the “Bayesian Evidential Learning" approach to the problem of uncertainty assessment in integrated reservoir modeling. Although the general approach was previously described in Scheidt et al (2018), this paper contains significant new elements, applications and discussions, which are very interesting for the community.

In terms of form, the paper is very well written and clearly presented, apart from minor issues. It includes a link to a Jupyter notebook implementing the methodology and applying it to the reservoir thickness. The implementation works fine, after some twiddling to install scikit-fmm. The overall structure of the code seems to allow for managing facies (some unused functions for facies modeling are present in the code), but the demo notebook assumes porosity to be 1 and water saturation to be zero. Even so, the reproducibility is much better than in most similar papers on this topic.

Overall, I congratulate the authors for the very interesting approach which represents a paradigm shift as compared to the current practice. I have, nonetheless, several comments, questions, and suggestions, which I hope will help to improve the paper. My recommendation is to proceed with minor/moderate revision.
We are very thankful to the referee’s thorough and in-depth review. The comments almost cover all the aspects of our paper. They are insightful and very helpful to improve our work. We overall agree with them. Provided below shows how the referee’s comments are addressed point-to-point and incorporated to the revised manuscript

Specific comments

• Overall, the introduction makes a good job introducing the general problem, but more precise explanations about the exact contributions of the paper would be welcome at this stage (in particular with regard to the other recent contributions of SCREF).
We rephrased the last paragraph of “Introduction” to clarify the exact contributions of our paper with regard to the previous papers on this topic (of BEL).
The main contributions are

- To propose a model falsification scheme using robust Mahalanobis distance.

- Extension direct forecasting based on sequential model decomposition to honor the hierarchical rules in geological modeling.
- A complete automation of geological uncertainty quantification using borehole data.

• The borehole data are generally at much higher resolution than the reservoir grid data. However, as in most reservoir modeling approaches, this paper assumes that the borehole data has been upscaled to grid resolution, a process which is source of inaccuracies in reservoir models. One of the points of the proposed approach is that the falsification step (Section 2.1.3) could in principle integrate the scale change. Comments on this would be welcome. Also, some additional precisions about the management of categorical variables during falsification would be welcome (in addition to the last sentence on page 5). From Eq. (8), it seems that the robust Mahalanobis distance accounts for spatial redundancy; please confirm.

In this paper we deal with borehole observations that are already upscaled to model grid resolution. Upscaling errors is not within the scope of this paper's research, but is for us a very active area of our current research. The way we approaching this problem is to add such error to the data covariance matrix in Eq10, using some Monte Carlo analysis, but as said, see our next paper. As a result the falsification then indeed tests only a necessary condition (upscaled data) but not a sufficient test (actual data)

One rather ad-hoc trick around that is to choose a lower threshold value as the tolerance of Chi-squared distribution of $\mathrm{RMD}(\mathbf{d})$ (making the test more powerful) We have added a few sentences in Section 2.1.3 to further clarify this problem.

Yes, spatial redundancy is accounted by performing dimension reduction (PCA) on the well data. Eq (8) uses the PC scores of $\mathbf{d}$.

• Overall, I am not fully clear about the falsification step. As this is a key aspect of the proposed methodology, it would be good if the authors could insist on the consequences of this approach as compared to the conventional method which creates models sampling exactly the borehole data. Errors between model and data may be acceptable for some applications such as hydrocarbon in place, as they will average out, but would they yield reliable forecasts of porous flow and transport if borehole data are not exactly honored by some model realizations?

The goal of falsification is not to check errors between model and data, but to check whether the model can predict the data, even if we would account for some tolerance (or error). If that is not the case, then one would be averaging out over a wrong model. Therefore, we are not looking here at variance of error, but bias in the prior

• Although this is not the main point of the paper, some parameters for generating the prior models could be described more precisely and discussed. For example, I am a bit puzzled by uniform distributions taken for the facies 1,2,3, which suggests that facies 0 will adjust so that the total is equal to 1, which may be a source of bias (see Haas Formery, Math. Geol. 2002, or compositional data analysis literature). What are the variogram ranges for facies modeling? In the figure, there seems to be a facies trend, but what are the parameters of this trend? Are the variogram models isotropic? Is the variogram of porosity the same for all facies?

Yes, the prior modeling is not the focus of our paper. We therefore prefer not to stress much on the prior modeling, but to focus more on our main contributions on falsification, direct forecasting on sequential decomposition and automation. Additionally, Referee #1 suggests to remove specific explanations on the prior modeling to gain more generality (see General comment #2 of Referee #1). We agree with this comment.

Regarding the prior facies modeling, a deterministic trend was applied. This is to make sure that the orientation and position of modeled facies belts are consistent with geological interpretations in Figure 5(a). The trend in the case study was provided by the field geologists, and officially used by the field operator. The variogram models are isotropic for facies, anisotropic for porosity. The variogram of porosity is the same in all facies. We quantify the prior model uncertainty under this scenario.

• Overall, I get the overall idea for facies, but I am not fully clear about the consequence of the facies processing. The signed distance is mentioned, and then the Truncated Gaussian. I first understood that the TG was used in the generation of prior models, and the signed distance for the workflow steps (which would mean in general 3 distance fields for 4 facies). But I was then puzzled by Fig 12 which suggests that maybe one single scalar field is used. So at the end, I am not sure about what was done exactly. Clarification of this would be needed in the paper.

The referee is correct on understanding the overall idea for facies uncertainty quantification. We agree that the main problem is from Fig 12 which is confusing. We replotted Fig 12 to avoid such confusions. The figure is replaced by the median posterior facies model to show the final results from Auto-BEL, which is also suggested by Referee #1.

• I looked at the code to try to understand the facies management, but it is not fully integrated in the high-level functions. Adding facies management in the code would improve reproducibility. If not possible, then please explicitly mention in Section 2.3 that the provided python code illustrates the workflow for thickness only.

We appreciate the referee's test of the Auto-BEL code. Regarding the problem of facies management, we couldn't provide the facies model data because it is classified as confidential by the company. However, all the code functions for facies management are provided in the repository, including signed distances calculations and back transform ("signed_distance_functions.py"), mixed PCA ("dmat_4mixpca.py"). We are working on create a synthetic facies data set so that the user can. We explicitly explained this problem on the provided github AutoBEL repo. we prefer not to write this on the paper because the code will continue to be updated.

• I have some doubts about the back transformation process when not enough PCs are retained. Some artifacts are visible in the realizations on Fig. 19 and on the poro/perm plots of Fig. 20 (breaking the consistency between petrophysical features within each facies). Could this also break the match to borehole data if not enough PCs were selected?

Here, we retain all the PCs of $\mathbf{m}$ in back transformation. Uncertainty reduction is performed only on the sensitive PCs, while for the non-sensitive PCs of $\mathbf{m}$, they remain random according to their prior empirical distribution. But both are used in back transformation. Therefore "not enough PCs" is not a problem in this paper. We further clarified this operation in the methodology section 2.2.2.

• As the aim of the paper is to "minimize the need for tuning parameters" it would be good to summarize the updated parameters as in Table 1 to help the reader assess to what extent the global model parameters have been updated by the process. I cannot help but wondering about how the updating of global parameters such as facies proportions or variogram range would compare to a classical process where statistical inference would be repeated by domain experts as new data become available.

We agree with the referee's suggestion. The updated global parameters in Table 1 are plotted in Figure 11, including the global mean, variogram ranges and facies proportions. In the Figure 11, they are also compared to their prior uncertainty to show the uncertainty reduction from the method. We prefer not to summarize them again in a table as the figure is more effective for the comparison than text table. Note that our method is Bayesian and therefore in accordance with the rules of probability (which domain experts may violate).

• Performance: I assume the 45 minutes do not include the prior model generation? Please clarify.

Yes, the 45 minutes do not include the prior model generation. We have clarified this in the revision.

• Discussion: the discussion in its current form highlights the main points of the method and some challenges, but does not really discuss some aspects which are often considered critical in subsurface models, such as the match of individual realizations to borehole data, or the preservation "geological consistency" such as the petrophysical distributions for various rock types. I suspect some moderate violations do not really matter for the accumulation problem considered in this paper, but I have more doubts about what would happen for modeling objectives involving highly non-linear physics, such as flow simulation. SOme balanced discussion on these aspects would probably be useful. Another question is whether there are any guidelines about the various sensitivity / confidence levels involved in the method, as these parameters likely impact the results.

We further extended the discussion section to stress more on the critical aspects mentioned by the referee. The extension includes matching individual realizations to local borehole observations and using of BEL for non-linear problems. The discussion section now stands out as a single section. The application of BEL or some BEL steps (e.g. direct forecasting) to non-linear physical problems has been investigated by Satija and Caers (2015), Scheidt et al (2018) on subsurface flows of oil/gas and groundwater reservoirs, and by Hermans et al (2018), Athens and Caers (2019) on geothermal heat prediction. These works are referred at the introduction to distinguish the unique contributions of our paper.

The guideline about sensitivity study has been thoroughly studied by Fenwick et al (2014) and Park et al (2016). We won't repeat this in the manuscript, but we refereed to these papers in revised section 2.2.1.

References:

Fenwick, D., Scheidt, C. and Caers, J.: Quantifying Asymmetric Parameter Interactions in Sensitivity Analysis: Application to Reservoir Modeling, Math. Geosci., 46(4), 493–511, doi:10.1007/s11004-014-9530-5, 2014.

Park, J., Yang, G., Satija, A., Scheidt, C. and Caers, J.: DGSA: A Matlab toolbox for distance-based generalized sensitivity analysis of geoscientific computer experiments, Comput. Geosci., 97, 15–29, doi:10.1016/J.CAGEO.2016.08.021, 2016.

Satija, A. and Caers, J.: Direct forecasting of subsurface flow response from non-linear dynamic data by linear least-squares in canonical functional principal component space, Adv. Water Resour., 77, 69–81, doi:10.1016/J.ADVWATRES.2015.01.002, 2015.

Scheidt, C. Ã., Li, L. and Caers, J.: Quantifying Uncertainty in Subsurface Systems, Wiley. [online] Available from: https://books.google.com/books?id=xvRYDwAAQBAJ, 2018.

Athens, N. D. and Caers, J. K.: A Monte Carlo-based framework for assessing the value of information and development risk in geothermal exploration, Appl. Energy, 256, 113932, doi:10.1016/J.APENERGY.2019.113932, 2019a.

Technical corrections

In several places: the term "data-scientific" looks hype but I don't get the exact meaning. lease define or use another term.

We replaced the term "data-scientific" by "statistical learning" for more precise description of our approach.

Also, the term "constraint" used in several places could probably be replaced with more accurate terminology.

We rephrased the sentences to remove the term "constraint" to for more accuracy.

p.1, l.19: "A generalized synthetic data set motivated by a gas reservoir": please rephrase. Seems to me this is a gas reservoir study which has been simplified.

The statement is rephrased to "a generic gas reservoir dataset.".

Yes, this is a reservoir simplified from real gas reservoir for general research implementations.

p.4, Eq. (3): Unless I am missing something, the notation could be simpler using $\boldsymbol{\chi}_{\mathrm{gl}}, \boldsymbol{\chi}_{\mathrm{sp}}$ etc.

Notation corrected accordingly

p.6, l.10: Please define G (and make it bold?). Fix typos in Eq. (10): "proir" should read prior.

G are bolded and defined in the revised Eq (10). Typos corrected accordingly.

p.6, l.24-25: Please rephrase the sentence for more clarity. This is more an explanation about why it works in practice than an actual "truth".

We remove this sentence because it introduces confusion, and previous explanations has already explained the problem.

p.7, l.1: Instead of “model grid cells”, I’d suggest for generality: model parameters

Revised accordingly

p.7, l.3-6: I’d recommend to factually summarize the DGSA approach rather than summarizing its advantages (as compared to what?).

We rephrased the sentences to compare DGSA to the other global sensitivity analysis methods such as variance-based methods (e.g. Sobol, 2001, 1993), regionalized methods (e.g. Pappenberger et al., 2008; Spear and Hornberger, 1980), or tree-based method (e.g. Wei et al., 2015).

References:

Sobol, I. .: Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, Math. Comput. Simul., 55(1–3), 271–280, doi:10.1016/S0378-4754(00)00270-6, 2001.

Sobol, I. M.: Sensitivity estimates for nonlinear mathematical models., Math. Model. Comput. Exp., 1(4), 407–414, 1993.

Spear, R. C. and Hornberger, G. M.: Eutrophication in peel inlet—II. Identification of critical uncertainties via generalized sensitivity analysis, Water Res., 14(1), 43–49, doi:10.1016/0043-1354(80)90040-8, 1980.

Wei, P., Lu, Z. and Song, J.: Variable importance analysis: A comprehensive review, Reliab. Eng. Syst. Saf., 142, 399–432, doi:10.1016/J.RESS.2015.05.018, 2015.

p.7, l.11-20: The notations of Eq 11 are unclear to me. I get the point of the sequential updating, but I am not sure it is correct to write “the posterior model of χ becomes prior model for ”, as both variables are different. Maybe using the subscript such as χ_posterior in Eq (11) would help make the point clearer.

We rephrased the sentences and revised the equation for more clarity.

P.7, l. 21-25: PCA on an image can be achieved in a variety of manners. I suspect that in Figs. 1 and 2, the PCA factors are linear combinations of image columns, but please explain so that the reader does not have to guess.

Fig 1 and 2 are to show the challenges when performing PCA on categorical models. We added further explanation for this problem.

P.7, l. 27: I’d suggest to use ψ_s, as it relates to the distance to facies s. I also think that $\mathbf{x}_\beta$ should be the closest location equal to (and not different from)s in the second term (otherwise, the definition enlarges the facies by 1 voxel). Or maybe just the distance to the boundary of facies *s*?

We changed the notation. $\mathbf{x}_\beta$ is the closest boundary of facies s. We further clarified this.

p.8, l. 7: Typo: the prior uncertainty models *have*

corrected accordingly

p.9, l.1-3: Please check convoluted sentence.

Sentence rephased

p.9, l.13: Did you mean “the reservoir rocks deposited at shallow marine environments”?

Yes, we added “rocks” to the sentence.

p.10, l.2: Confusion between in-place resources (GIIP) and recoverable reserves, which also depends on flow behavior.

We removed the word “reserve” to only keep GIIP to avoid the confusion.

p.10, l.12-13: Please check syntax.

Syntax corrected

P.10, Fig.5: a scale would be needed so that variogram ranges provided later in the paper can be related to model size and well spacing.

Scale bar is added to the figure.

p.11, table 1: Typo in “gammy ray”

Typo corrected

p.12, l.2: In addition to resolution, velocity is a significant source of thickness uncertainty.

We added velocity as another source of uncertainty

p.13, l.10-15: h is the height above the free water level, not the reservoir depth.

Corrected according.

p.13, l.19-20: Please fix sentence.

Sentence fixed

p.16, l.11: The independence between thickness and facies is stated as a fact. In my view, it is a working hypothesis (probably a reasonable one), but not a general truth.

Yes, the independency between thickness and facies is a fact for the case, not a general truth. We rephrased the sentence to gain clarity.

p. 18, Fig.10: Would dobs correspond to a line? I guess it should have some thickness due to data noise and to PCA projection.

We agree with the referee on this. For this synthetic case, we didn’t consider the data noise, so it is one value represented by a line.

p.19, l.1-2: “the uncertainty... their prior”: Please chech grammar.

Grammar checked and corrected for this line.

p.19, Fig.11: Please tell what the dash lines correspond to (kernel density?)

The dash lines are the estimated probability density using Gaussian kernels. We add this explanation to the figure caption.

p.20, Fig. 12: The visualization for posterior facies distributions gives a qualitative hint about what happens, but I am not sure about what we are exactly looking at. It is the blending of discrete color maps or the average of the underlying Gaussian field (but then, facies threshold change depending on facies proportions)?

We agree with the referee’s comments on the Figure 12. This figure has been replaced by the median prior and posterior model. This is also suggested by Referee #1.

Fig. 15: Typo in legend

Typo corrected

p.25: Please remove mention to reserves, as no recovery factor is involved.

"reserves" removed in the revision

p.25, l.17: The cross-validation tests on wells 5 and 6 seems a bit optimistic, as vertical averaging on the 75 layers essentially amounts to making a two-dimensional model. Again, erors average out, so the reduction of uncertainty in such a case is no proof of the actual forecasting ability of the model in three dimensions.

The aim here is to test the resulting posterior models are not wrong and can predict future new borehole observation with reduced uncertainty. We rephased these lines near l.17 by removing the statements on forecasting ability to avoid the ambiguity and stress on our main point.

p. 24, Fig. 20: The correlation coefficient is not really meaningful on such multi-modal distributions, even more so as the facies proportions likely change from one distribution to another. It would make more sense to examine the bivariate statistics per facies.

We have replotted the Fig 20 to examine the bivariate statistics facies by facies.

P.26, Fig. 23: please explain what we see: the curves are densities, but what does the point horizontal spread mean? And again, what is the facies value?

The horizontal spread of the scatters is for better visualization. Otherwise, the points will overlap each other, making it difficult to observe their distributions. The facies values on the plot are averaged across the 75 layers.

p.27 l.12, l.15, l.24 : "results an extreme fast computation"; "be able predict", "do not full match": Please fix English

These sentences are fixed.

p.28, L.11-21: This paragraph nicely explains the problem in simple terms, so I think it would be better integrated in the introduction than in the discussion.

We put this paragraph to the newly created discussion chapter because it fits better to the context below and above.

p.28, L.16: this sentence mixes the falsification of parameters and falsification of a methodology, which I think are very different. I suggest rephrasing.

The sentence is rephrased to focus on the problem from falsified prior model.

p.28, l.30: Not sure I understand the references in this context.

We rephrased the sentence.

Responses to Tao Feng (taof76@hotmail.com) on interactive comments (SC1)

Dear authors,

This paper presents an automated workflow to build geo-model using some hard data. The work is very interesting, and the paper is well written. I strongly suggest publication after some revision. I have a few questions and comments. I hope authors can clarify it.

We thank the reviewer for the interest and comments on our paper. Please see below our responses.

1. In BEL, you mentioned prediction. There is nothing related prediction from the field case, am I right? It seems to me this work is mainly related model building using data. The prediction variable h is the model parameters, right?

In the field case, the prediction variable **h** is reservoir storage volume, which is directly calculated from the models. Our main work was to develop an automated approach (Auto-BEL) to jointly update spatial models and storage volume uncertainties using data. The steps of Auto-BEL include build/update models using data.

2. Using observation d_obs, you can detect some outlier realizations. Do you just remove these realizations from your prior in practice?

No, we didn't remove these realizations. We keep all the prior realizations if they are not falsified. Removing the outlier realizations can change the prior distribution.

3. h*, d* are some subspace of h and d, right? If we talking permeability field with millions of cells, could you give me roughly number of h* compared to h? Is Formula (9) standard way to formula linear-Gauss problem?

Yes, h*, d* are subset of orthogonalized h and d (after PCA, and CCA). The dimension of h* is the total number of model realizations, although h can be up to millions of cells. Yes, for linear-Gauss problem, Formula (9) is standard.

4. Your Python tool can be used to build geo-model (grid, etc..)?

Yes, it can directly update/rebuild geo-models, once new borehole data are provided, without using any external modelling tools. But it requires Monte Carlo of prior geo-models for "training". These prior models were built with a geomodeling tool (Petrel)

5. After CCA, elements in h* is more independent (less correlated), right?

We don't think so. CCA didn't change the independency of h* elements. h* elements are independent already before CCA, because they are orthogonalized via PCA.

6. I am very interested in "sequential update model". It will be nice if the authors can describe this in more details.

Thanks for the suggestion. We have added a paragraph to the application section (section 3.4.2) to explicitly explain how the sequential approach is performed.

7. Every time you update a parameter, do you use the posterior as a prior for next parameter update?

Exactly, the posterior from previous sequence is used as prior for the next update.

Responses to Dexiang Li (dexiangli_pe@163.com) on interactive comments (SC2)

Received and published: 16 December 2019

Geological models can be more accurate and actual with coupling more borehole data in models. Meanwhile, the data sizes of geological models increase with the developments of field projects and participation of new borehole data. During dynamic process of subsurface applications such as groundwater, geothermal, oil, gas, and CO2 geostorage, uncertainty quantification is the key for decision making. As the authors mentioned, uncertainty reduction is a time-consuming work which requires iterative model rebuilding using conventional inverse methods. In order to make the model adhere to geological rules, geological modeling often requires significant individual/group ex-pertise and manual intervention which will need often months of work after new data is achieved. In this paper, the authors generalized a Monte Carlo-based framework for geological uncertainty quantification and updating. Their methodologies were developed with the BEL protocol for uncertainty quantification. The extension of directly forecasting results an extreme fast computation of posterior geological model, by avoiding conventional model rebuilding. The proposed framework also allows automation of geological UQ. This paper is interesting and in an area worthy of investigation. Overall, this paper is well-organized and well-written. This paper can be accepted by addressing the following minor comments.

We appreciate the independent reviewer's in-depth understanding of our paper. The comments are helpful to improve our paper. Please see below our responses and explanations on the revision.

1. The advantages and disadvantages using your method for UQ and updating should be further illustrated by comparing with typical conventional method. At the same time, its applicable scenarios are suggested to be provided which can give guidelines for field application.

We extended the discussions on limitations in the last paragraph of "Discussion and conclusions". The advantages of this method have been discussed in second and third paragraph (page 27, 28). The application scenarios are for uncertainty quantification using borehole data in geological modeling and prediction, which are common in oil&gas, geothermal, CO2 sequestration applications. We have added a statement at the beginning of abstract for more context of this.

2. As you mentioned, current method is only designed to globally adjust the model, not locally at the borehole observation. Could you provide your idea on further solution in more details?

As we explained at the discussion section, one possible solution we like to explore is to combine geostatistical conditional simulation as posterior step to our current methodology. For example, once the posterior global parameters are calculated from AutoBEL, they can be used as the input to geostatistical simulation conditioned to the local well observations. This will enable posterior models locally matched to the borehole observations with reduced global uncertainty.

3. Could provide the specific performance parameters of CPU which can show the improvement on calculation efficiency more accurately?

The CPU is Intel Core i7-7820HQ. We added this specification to the revision.

4. The authors are suggested to unify the multiplication sign through the whole manuscript?

We unified the multiplication symbols in the revision.

5. Please add a "." between "Figure 19" and the "Prior and posterior..." to keep in accordance with other figures. Please check similar problems accordingly.

"." has been added to the figure captions.

6. The usage of abbreviation such as DF should be noticed. Abbreviations should be defined when they are first mentioned in the text and should always be used afterwards.

The abbreviation DF is defined at Page6/Line4

7. Discussion and conclusions are suggested to be separated into two parts. Please provide conclusions point by point which can help reader to understand the main con-tributions of the paper. Meanwhile, future researches should be clarified according to the limitations of proposed method.

Thanks for this suggestion. Discussion section now stands as a single section for more in-depth discussions. The conclusions are point by point already. In paragraph 1 of this section, we have clearly itemized the contributions by words such as "generalized", "second contribution", "third contribution" … We extended the discussion on future researches and provided relevant references in the new revision.