Dear Guillaume Caumon,

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 which are highlighted in red.

On behalf of the Authors,
Zhen Yin (David)

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.