Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2019-232-RC1, 2019

Geoscientific Model Development Discussions
Open Access
EGU

Interactive comment on “Automated Monte Carlo-based Quantification and Updating of Geological Uncertainty with Borehole Data (AutoBEL v1.0)” *by* Zhen Yin et al.

Anonymous Referee #1

Received and published: 19 November 2019

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,

Printer-friendly version

Discussion paper

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.

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.

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.

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.

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.

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.

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.

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 Y-axis 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?

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

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 ?

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 ?

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 ?

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 ?

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

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).

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.

13. Page 25 L7. I guess the 45 minutes do not include the creation of the 250 MC samples forming the input.

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.

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.

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2019-232, 2019.