Solid Earth Discuss.,
https://doi.org/10.5194/se-2019-78-RC1, 2019

Solid Earth
Discussions

Interactive comment on “Topological Analysis in Monte Carlo Simulation for Uncertainty Estimation” *by* Evren Pakyuz-Charrier et al.

Guillaume Caumon (Referee)

guillaume.caumon@ensg.univ-lorraine.fr

Received and published: 11 June 2019

1 General comments

This paper proposes and evaluates a method to define clusters in a population of structural models obtained by spatial data perturbation. An original feature of the approach is that the clustering uses only topological distances. I would like to congratulate the authors for addressing this difficult problem, which is very relevant uncertainty management in structural studies.

The method builds mainly on previous papers by Thiele et al (2016) and on the DB-SCAN algorithm. The evaluation is made on two populations of models obtained with

varying degrees of data uncertainty. The results are very interesting, as the proposed method does identify relevant population subsets, but only very few clusters manage to be identified even when the method's parameters are tuned. This suggests the structure of the problem is very continuous, and tends to generate models which are mutually very close one to another when considering the adjacencies of geological units.

I mainly have form comments, which I hope will will help to improve the paper.

2 Acronyms and wording

Overall, the paper is well written and easy to read, but heavily uses three acronyms: MCUE, UIM, PGM, which I have not seen in other author's work working on similar topics. So, I wonder if we really need these acronyms.

In particular, the term MCUE (Monte Carlo Simulation for Uncertainty Estimation) is very general and not specific to the proposed method, so I'd recommend to change the name to better explain that the data are perturbed / sampled to generate a set of probabilistic geological structural models.

PGM (Probabilitic geological models) is quite clear, but I am not 100

UIM (Uncertainty Index Models) is expanded in the introduction, but not explained before page 3 (mention to the work of Wellmann and Regenauer-Lieb, 2012), so could be difficult to understand upfront. Why not just mention a map of local uncertainty?

Another point of vocabulary is the term "geologically incompatible models" and "geologically [in]consistent". To me, geologically incompatible would mean transforming a reverse fault into a normal fault, or changing the geological history. I understand that the proposed data perturbation may change a normal fault, but this is not what the clustering detects. In this paper, consistency / compatibility essentially means topological similarity, so I'd rather use that term. Of course, topological differences may have implication in the geological history (e.g., the juxtaposition of one formation against another may create paths for subsurface to migrate), but this is not mandatory. Geometrical variability between models of the same topology could have a similar effect, see for example Abrahamsen et al (2015). I would recommend to stay with descriptive terms (topologically similar / dissimilar) in the bulk of the manuscript and discuss the geological implications in the case studies and in the discussion.

Overall, I think the paper could be improved by more thoughly citing and refering works on structural uncertainty done by other teams. In Section 4 of Wellmann and Caumon (2018), we tried to review the various approaches to structural uncertainty assessment, so I hope this could be a useful entry point.

3 Motivation

The introduction in its present form presents the problem in a very general way, which is nice... but maybe too general. I think the intro could do a better job to motivate the need for model clustering, which is the key aspect of this paper. This goes along the lines of clarifying (or replacing) the term "incompatibility", and possibly also of explaining a bit more in what sense a categorization of models could help to reduce uncertainties (as mentioned at the beginning of section 3). This would certainly call for some additional references to inverse problem theory, which can look around a particular scenario in model parameter space, but has more difficulties to work with problems of varying numbers of parameters. Carter et al (2006); Suzuki et al (2008), Cherpeau et al (2012), Scheidt et al (2018) could be useful references to discuss this.

4 Heteroscedasticity and error correlation

In Section 1.2 and other places, heteroscedasticity in the data set is invoked to imply dependency within the data. I agree, but this is not the only reason. Multiple examples of spatial correlation have been documented in the literature, especially when the data are interpreted from seismic images (where location errors stem from velocity errors). Although rare, this could also occur in principle with field geological data (e.g., poorly calibrated instrument leading to systematic measurement bias in some areas). At the bottom of page 4, the authors seem to suggest that heteroscedasticity always implies spatial correlation. I am not sure whether this is correct and would argue that heteroscedasticity and spatial correlation are two different (and important) aspects of data uncertainty.

5 What does topology exactly means?

If I understood correctly, the type of topology used in this paper is primarily “Lithological topology” (sensu Thiele et al., 2016), ie, the nodes of the topological graph represent faulted, folded and possibly eroded geological formations while the edges of the graph represent the adjacency between these formations. This is clear from the title of section 3.1, but it was not clear to me when I first read the paper, which trigerred many interrogations. Having now understood that lithological topology is considered, I still have two comments:

- Considering 1's on the diagonal seems like a choice that a formation is considered adjacent to itself (although one could argue that this is not really adjacency). Did I miss something here? Actually, I do not think the daigonal is so important in the characterization, so would it make sense to just ignore the diagonal in further steps?

- The “lithological topology” considers only geological formations and not the connected components of these formations (termed “cellular topology” by Thiele et al.), whose number may change from one stochastic structural model to the next. So, the existence of the same number of lithologies in all structural models seem like a prerequisite to apply the proposed method. This should be more explicit. The variability in lithologies only summarizes much of the variability that would be observed considering the adjacencies of connected components. I suspect that this contributes to the reason why the clustering algorithm has difficulties to segragate realizations.

6 Cluster Entropy

I am not completely sure I understand the cluster entropy concept, because I suspect there is a typo in the equation: I am not sure about the k in the log, and it seems to me that (if k is indeed a mistake), the result will always be zero (sum of 0 log(0) and 1 log (1)). I suspect there should be some average connectivity involved (probably $\frac{A(k)_i^j}{c} log \left(\frac{A(k)_i^j}{c}\right)$ and not just $A(k)_i^j log \left(A(k)_i^j\right)$). Maybe I am just missing something, but in any case some references would be welcome.

7 Minor remarks

- page 3: “flattened to triangulated surfaces or shrink to triple lines”: Unclear to me.
- page 5: the mention to adjacency, overlap and separation are already made in Thiele et al. (2016), and only adjacency is used in this paper, so maybe there is

no need for discussing the combinatorial aspects here.

- It took me som guesswork to unsrstand Table 4. Please explain that the 1-8 codes correspond to lithologies; having the table of lithologies would help analyzing the results and following the discussion. The lower right matrix is the difference of the matrices in the first raw, right? Please add "See text for detail" in the caption, not all elements are described in the caption.
- Considering the most significant topological classes (page 8) is acceptable but it is arguable for uncertainty quantification is high dimensional spaces, as it may artificially reduce uncertainties. I think this should be mentioned.
- Some of the discussion on distances could possibly benefit from references to the recent book of Scheidt et al (2018).
- The Appendix provides interesting details about spherical orientation

Please see also annotated manuscript.

8 References

Only references not already cited in the paper are included:

Abrahamsen, P., Dahle, P., Hauge, V.L., Almendral-Vazquez, A., Vigsnes, M., 2015. Surface Prediction using Rejection Sampling to Handle Non-linear Constraints. Bulletin of Canadian Petroleum Geology 63, 304–317.

Carter, J.N., Ballester, P.J., Tavassoli, Z., King, P.R., 2006. Our calibrated model has poor predictive value: An example from the petroleum industry. Reliability Engineering System Safety 91, 1373–1381. https://doi.org/10.1016/j.ress.2005.11.033

Cherpeau, N., Caumon, G., Caers, J., Lévy, B., 2012. Method for Stochastic Inverse Modeling of Fault Geometry and Connectivity Using Flow Data. Mathematical Geosciences 44, 147–168. https://doi.org/10.1007/s11004-012-9389-2

Scheidt, C., Li, L., Caers, J., 2018. Quantifying uncertainty in subsurface systems. John Wiley Sons.

Suzuki, S., Caumon, G., Caers, J., 2008. Dynamic data integration for structural modeling: model screening approach using a distance-based model parameterization. Computational Geosciences 12, 105–119.

Please also note the supplement to this comment:
https://www.solid-earth-discuss.net/se-2019-78/se-2019-78-RC1-supplement.pdf

Interactive comment on Solid Earth Discuss., https://doi.org/10.5194/se-2019-78, 2019.