Solid Earth
Discussions

Open Access

SED

Interactive
comment

Interactive comment on "Bayesian geological and geophysical data fusion for the construction and uncertainty quantification of 3D geological models" *by* Hugo K. H. Olierook et al.

Hugo K. H. Olierook et al.

hugo.olierook@curtin.edu.au

Received and published: 3 May 2019

Reviewers' comments:

Reviewer #1: 1. General comments Olierook et al. present a study where they integrate information from geological maps and observations, petrophysical measurements and geophysical data. They focus on a small area in Western Australia and use interpretation from a nearby seismic survey to constrain their modelling. Their aim is to present an example of holistic inverse modelling where gravity and magnetic data are constrained using all the available information.

We thank the reviewer for their rapid turnaround time and detailed comments. Please find attached a zip file containing comments to all reviewers, a manuscript with track changes and a clean manuscript with all changes incorporated.

There is some novelty in their approach and the study they show can be published after the authors address a series of comments about specific issues in the text and more general issues which, at the moment,is problematic. To date, Monte Carlo approaches for geological uncertainty have focussed on regional scale studies, while the work is clearly at a smaller scale. I think that this should be highlighted as one of the novelties of this paper.

AGREE. We have now highlighted this in the abstract, introduction and discussion.

I think that the formulation of the geophysical inversion problem should be described in more detail and that giving essential equations about geophysics and uncertainty assessment would improve the manuscript greatly.

AGREE. We have added many of essential equations that were previously only present in Scalzo et al., in review. We also clarify the covariance function for boundaries and the prism approximation used in calculating potential fields (citing Li & Oldenburg).

The only equation shown in the paper does not suffice to provide a good understanding of the basic mechanics of the methodology. Besides, after a quick manual derivation, I think that the right hand side of the equation provided might not be correct and may need revision. In any case, this derivation needs to be justified, by invoking an Eulerian integral of the first kind and the Beta function, using either the appropriate reference(s) and providing a succinct appendix.

AGREE. We appreciate the close attention to detail. The form of the beta-binomial likelihood shown in the earlier version was indeed incorrect. We have verified that the correct version was used in our code; the error is one of transcription and not the underlying derivation. We now provide a more detailed form of that derivation in an

appendix, as well as the derivation for the Student's-t likelihood used for the gravity anomaly and magnetic intensity.

Some references may be missing or are miscited. Several studies have been overlooked and have not been cited. This comment is relatively minor but addressing it would be important to show where the presented work stands in the literature. In my view, the introduction should emphasize the fact the idea of geology-geophysics integration is not new but that quantitative integration of both discipline is an area of research that has received more and more attention recently. Some references need to be added, but I will come back to this in my detailed comments of the document. Please check that all the papers you cite as 'in review' are still in review and have not been published.

AGREE. We have added several more references throughout the manuscript, as suggested by reviewer 1 and 2. Only Scalzo et al. was still in review with GMD and, unfortunately, is still in review.

The hypotheses and assumptions that the authors made need to be clarified. They neglect the presence of some geological units on the basis that their relative coverage in previous interpretation amount to only a few percent of the total. This is a simplification that need solid justification as it is sometimes the case that only a very small portion of a rock unit of interest is outcropping. Moreover, the area of the authors' study is known to be prospective for several minerals, the deposits of which is often not born by geological units making up most of the geology of the area, and are covered by regolith and outcrop only at a few locations.

PARTLY AGREE. The other lithologies are less significant volumetrically, appearing primarily near the surface, and are also present areas smaller than our model can resolve. However, we agree that this was not well emphasized in the text. We have focused in this work on mapping the boundary at depth between the two major units. More detailed future work will undoubtedly move towards sampling of more detailed

Interactive
comment

models that can resolve finer features.

An aspect of geophysical joint inversion that needs to be considered is the relative impact of the two different geophysical datasets inverted for when they present largely different spatial coverage and sampling. This is not clearly mentioned in the manuscript and I expect that there would be an imbalance as the magnetic dataset seems to have about 100x more points than the gravity one (see Kamm et al. (2015), Sun and Li (2016)). How do you cope with the fact that in such case your joint inversion may be dominated by magnetic data? State it clearly.

We would consider the imbalance in the sizes of training sets to be a genuine asymmetry in available information from the two sensors. However, the reviewer has a point in that the magnetic data is available on a much finer spatial scale than our actual model parametrization can possibly resolve. This produces correlated residuals from our baseline model that result from model inadequacy instead of from random variation of measured sensor values from an underlying forward model. The effect is reduced by our Students-t likelihood relative to the usual Gaussian likelihood, i.e. by our relatively vague prior about the expected noise variance in each sensor.

The Figures are not all very informative and some could be grouped. Fig. 1 and Fig. 2 both refer to the geology of the area and are referred to next to each other in the manuscript. It may be a good idea to merge them.

DISAGREE. Figure 1 and 2 are difficult to merge into one because they use different colour schemes and have different aspect ratios of maps and cross-sections. We have attempted combining Figs. 1 & 2 previously but found it added more confusion when merged then when separated.

Fig. 3 is not very informative.

AGREE. We have removed Figure 3.

In Figure 5, I think that it would be good to have the line of cross-section X-Y shown.

C4

AGREE. The cross-section line in Figure 5 has been added to each of the panels, and the caption updated.

The authors rely a lot on Scalzo et al., in review, which is a good complement to the manuscript. However, it is submitted to a different journal and is still in review. For this reason, I suggest that they reduce the dependency of their manuscript to Schalzo et al., in review, and explain succinctly key concepts they refer to Scalzo for explanation. This would make the paper more readable and easier to understand as all key elements would be readily available. The manuscript also has a number of sentences or pieces of sentences that are either an exact match or are very close to what can be read in Scalzo et al., in review. This is not an accusation of plagiarism but a mere observation. There are a few occurrences that I noticed when reading Scalzo et al., in review.

The recurring phrases are oversights on our part. We originally had one longer paper which was split at an early stage of editing into "methods" and "applications" papers, and the repeated text elements are left over from this stage. Scalzo et al. focuses largely on sampling from the posterior distribution under different priors, while this work focuses on a specific application and on introduction of the beta-binomial likelihood for lithostratigraphic measurements. We agree that each paper should be able to stand on its own and that no text should be re-used. We have taken steps throughout to reduce the interdependence of the two manuscripts while preserving complementarity.

Below are the detailed comments I have.

2. Detailed comments and technical corrections You mention the fact that you use a global optimization technique only in 2.2. Maybe state it in the intro.

AGREE. We have added a mention to Obsidian's parallel-tempered MCMC scheme to paragraph 3 in the introduction. We re-emphasize there that our model estimation is carried out through sampling, not optimization, which may be important in situations with vague prior knowledge such as that encountered in an exploration setting.

C5

P1 l14: "model results": results of the technique introduced here? Could be clearer.

AGREE. This has been changed to "3D model results" to help the reader understand the link to the previous sentence.

P1 l16-18: "The boundaries between geological units are characterized by narrow regions with <95% certainty, which are typically 400–1000 m wide at the Earth's surface" what is the relation between these values and the sampling of gravity data? If I am correct the data sampling of gravity data guaranteed by GSWA is that there is a data point every 400 to 1000 m in the area?

The nominal station spacing for the gravity data is 2.5 km. Interpolation of these data yield a datapoint every ~400 m but this interpolation is not a Bayesian method. Instead, we have used the original 2.5 km spacing to avoid introducing correlations in the interpolation process.

P1 l18: "Beyond âĹij4 km depth, the model requires drill hole data". You need to be clearer here. Drillhole data that reache below 4km in the area is not likely in hard rock scenarios, although it might be in oil and gas exploration (basin scenarios). I suspect that you mean that for model cells below 4km the addition of constraints at depth such as drillhole data might help constrain the deeper regions better?

PARTLY AGREE. To avoid confusion in the abstract, we have removed the drill hole constraints. We have left the option of seismic data to be able to constrain models at >4 km depth.

P1 l27: "faults or suture zones": how about unconformities in general?

AGREE. This line has been changed to "...via unconformities or structural discontinuities such as faults or suture zones."

P2 l13: You cite Pakyuz-Charrier et al 2018 but this is a conference abstract. There are two journal papers relating to their MC approach for geological modelling that appeared in 2018. Please replace that reference by the most appropriate one (or both if you want

to be broader) of the following: Pakyuz-Charrier et al. (2018a), (2018b).

AGREE. We apologize for referencing this conference abstract. Both Pakyuz-Charrier papers have now been cited.

P2 l13-14: There have been metrological studies published in recent years that tackle the issue of modelling the uncertainty on geological measurements.

As mentioned above we were aware of Pakyuz-Charrier et al 2018(a,b) and references therein, which deal with the propagation of uncertainties on structural measurements and contact point measurements from drill holes. We were unaware of any previous treatment of lithostratigraphic measurements at the surface, which occasions the new likelihood we derive in this work.

P2 l21-22: "However, there is still a paucity of work in fusing solid Earth geological observations and geophysical data in a Bayesian framework to develop robust 3D geological models". True, but you may need to consider Wellmann et al. (2017), who "...address these shortcomings here with an approach for the integration of structural geological and geophysical data into a framework that explicitly considers model uncertainties [...] in probabilistic programming in a Bayesian framework". Please cite this work. This also relates closely to de la Varga et al. (2018), which you cite earlier in this paragraph. Likewise, Jessell et al. (2010), (2014), (2018) highlight the need for robust geology-geophysics integration. I suggest to cite some of these as they strongly advocates for the kind study presented here.

AGREE. All these papers have now been cited in these two sentences: "...(iv) fusion of structural geology data with geophysical datasets (Wellmann et al., 2018). Despite a clear need for Bayesian fusion of solid Earth geological and geophysical datasets (Jessell et al., 2014; Jessell et al., 2018; Jessell et al., 2010), there is still relatively little work in developing robust 3D geological models, particularly at the local and camp-scale."

P2 l31-33: "One useful addition to the current features of Obsidian would be the integration of geological and geophysical field observations made on the Earth's surface, which are vital for surface and near-surface applications (< 1 km)" this statement sort of contradicts the 1st sentence of this paragraph where you say "The Obsidian software package provides a workflow to fuse disparate geological and geophysical data within a Bayesian framework".

AGREE. We have updated the first sentence to emphasize Obsidian's distinctive value-add, which is its distributed MCMC sampler for 3-D geological models conditioned on geophysical survey data. We then say, "Although previous iterations of the Obsidian software package could not fuse geological field observations made on the Earth's surface with geophysical survey data, relatively little amendment to the program is required to make this possible."

P3 l1-2: "geophysical observations" and "geophysical survey data": how is it not the same thing?

AGREE. Geophysical observations has been deleted. This was an oversight during text editing.

P3 4-14: Maybe say somewhere that exploration undercover has been recognised to be important for the future of mineral exploration with a ref or two. The last sentence of this paragraph could also go in conclusion.

AGREE. Added a final sentence to the first introduction paragraph: "In a future where exploration under cover has been recognized as vitally important for the mineral exploration sector (McFadden et al., 2012), developing geological models with accounted uncertainty is pivotal."

P3 section 2.1: Lead authors Johnson and Sheppard are cited a number of times - consider adding work from someone else.

DISAGREE. These two researchers have worked most thoroughly on the tectonic his-

C8

tory of the Capricorn Orogen.

P4 section 2.2: Sambridge and Mosegaard are cited many times here – maybe add some diversity with papers coming from other researchers.

AGREE. This section has now been peppered with many other references.

P4 l14: "single unique" → "unique" is enough.

AGREE. Single has been deleted.

P4 l16: There are also other works you may want to cite when it comes to using infomation derived from geological measurements or modelling directly into geophysical inversion. For instance, Fullagar et al. (2008), Guillen et al. (2008), Scholl et al. (2016) integrate geological information or modelling in their inversion algorithm. Publications relating to works using level-set inversion also rely on geological models (see for example Bijani et al., 2017, and Zheglova et al., 2018, for joint inversion).

AGREE. We have updated this sentence to: "Ways to introduce such constraints include 3D geometry inversion (Fullagar et al., 2008; Guillen et al., 2008), level-set inversions (Bijani et al., 2017; Zheglova et al., 2018) and (cross-)gradient regularization (Giraud et al., 2019; Scholl et al., 2016). However, these techniques are deterministic, yielding a single geological-geophysical inverse model that represents only one scenario."

P4 l15-16: "Ways to introduce such constraints include regularization (Giraud et al., in review) but this technique fails to acknowledge alternative scenarios." This statement in unclear as Giraud et al derive constraints from a collection of geological models, therefore using all realizations from MC sampling of geological model space. But because their inversion is deterministic, they obtain a single geolophysical inverse model, which indeed represent a single scenario. I think that this is what you mean but it needs to be clarified.

AGREE. See previous comment.

P4 l21-30. "The Bayesian framework converts a deterministic model into a probabilistic one by using probability distributions to represent the free parameters rather than using optimal or single-point estimates." To make this clear and unambiguous, you should remind the principle behind Bayesian approaches, when you invoke Bayes' theorem. Solving a problem in a probabilistic way does not necessarily make it Bayesian. In this sentence you may want to stress the fact that you also use the prior distribution and sample the posterior, as it is a major difference with deterministic inversions. The utilisation of the priors is stated in the next sentence but I think that it could be made clearer overall. I would also not cite Oldenburg 2005 here, but perhaps one of A. Tarantola and others' publications which are seminal to many inversion approaches.

AGREE. We have updated the text to include an invocation of Bayes's theorem. We also include descriptions of the Metropolis-Hastings algorithm later on.

P5 l1: I strongly suggest that you add the mathematical foundation of your Bayesian inversion methodology. Just a few of the equations centre to your modelling approach would do and I think be informative to the reader.

AGREE. Equations we have added in response to this comment include: the Metropolis-Hastings criterion, the swap rule for parallel-tempered MCMC, the Crank-Nicholson proposal, explicit derivations of the likelihoods we use, and definitions of the PSRF and Geweke convergence metrics for MCMC sampling.

P5 l10: "convergence can be challenging" rephrase. Something like: "convergence can be difficult to reach"

AGREE. This has been changed.

P5 l13-24: Please add other cites to your citations of works by M. Sambridge and his team's. For instance, they may have brought lots of new ideas and methods to the field but the Metropolis-Hastings algorithm is not by them.

PARTLY AGREE. Both Metropolis et al., 1953 and Hastings, 1970 are already cited in

the next paragraph, where we believe it is most appropriate.

P5 l23: section 2.3. You need to give more information about Obsidian. It is too short and not sufficiently informative. The main reference you use (Schalzo et al in review) is still in review a basic summary of Obsdian should be self-contained in the paper.

AGREE. We have rewritten substantial parts of sections 2 and 3 to provide more specifics relevant to our problem.

P5 l31-32: This sentence is exactly the same as Scalzo et al., in review, beginning of section 2.4. Layers are not discrete if they have smooth boundaries. I suspect that what you try to say is that the outline of a given layer is not angulous (i.e., that its representative functional would be differentiable)? If so this should be made clear.

AGREE. We have updated the text and now specify that the layers are differentiable (as they must be if a square exponential kernel is used).

P6 l2-4: "The layer boundaries are indexed in a strict order of increasing depth in the subsurface but are permitted to cross." The fact that they cross violates basic stratigraphic principles. You need to state explain briefly how you cope with that and how this is dealt with by your algorithm. This sentence is exactly the same as Scalzo et al., in review before they introduce equation 9.

AGREE. This phrasing was unclear. We mean that the ordering constraint among the depth to each layer is not explicit in the parametrization, but is enforced at the stage when the model is discretized for calculation of the potential field forward models. We now state this expressly: "The constraint $z_i(x,y) \leq z_{i+1}(x,y)$ for each layer i is enforced at this stage, allowing layers with no coverage in a particular region to "pinch out" to zero thickness."

P6 l10: "x and y correlation length" you haven't introduced what x and y are, except in Fig. 2. Just say that your RBF is anisotropic and I think that it's enough. The first sentence of this paragraph is very close to the one preceding equation 10 in Scalzo et

C11

al., in review.

AGREE. We now specify that the kernel is anisotropic and that the covariance length matches the lateral grid resolution between control points.

P6 l13-15: please reformulate the last sentence. The parameter alpha and beta haven't been defined.

AGREE. We now defer discussion of these parameters to section 3.4 (on likelihoods) where alpha and beta are defined when they are used.

P6 l18: PTMCMC: this acronym has not been defined yet.

AGREE. This has now been defined in the last paragraph of section 2.2.

P6 l32: "and cross-cuts" you cannot rule out all uncertainty about the fact that it cross-cuts, but you are making the (reasonable) assumption that it does. Please add this information.

AGREE. This has been added.

P7 l1: "ordering of layers". Say "stratigraphy" or "stratigraphic pile".

AGREE. Changed "ordering of layers" to "stratigraphy"

P7 l6-9: typo in line 6. I don't think that a lithology with 3% of occurrence is insignificant. You need to provide more information as to why you do not account for rock units except Halfway Gneiss and Durlarcher Supersuite.

DISAGREE. These lithologies are less significant volumetrically, appearing primarily near the surface, and are also present areas smaller than our model can resolve. We have focused in this work on mapping the boundary at depth between the two major units. More detailed future work will undoubtedly move towards sampling of more detailed models that can resolve finer features.

P7 l14-16: how do your values compare to litterature values, or work reported in the

Interactive comment

area or similar settings? If you have only approx. 100 samples to characterize several rock units through mean and standard deviation you can assume that your uncertainty on these parameters is quite high.

These are all the petrophysical values available from the area.

P7 l24-32 and 1st paragraph in P8: I don't think that it is necessary to give detailed information about the processing of the field measurements. You can keep it, but I recommend to put it in Appendix instead of the full text.

DISAGREE. We think this information is particularly important as it highlights the relatively minor pre-processing of the data, which is non-probabilistic. The interpolation vs. original data is also an important point that, in our opinion, needs to be kept in the main text.

P8 l15: consider replacing 'acquired' by 'obtained'.

AGREE. Replaced.

P8 l21: I think that you should replace ">" by plain text words.

AGREE. Replaced with "...more than 100 age and over 500 samples with..."

P9 l2-3: please explain briefly how the shape parameters were obtained.

AGREE. These parameters constitute a prior elicited from our geological expert collaborators and we have added text to explain this.

P9 l9: please number the equations.

AGREE. This has been done

P9 l13-14: the calculation of this integral is not straightforward. If you want to leave it in the manuscript as it is please add a short appendix explaining how it is calculated, and at least provide the appropriate references.

AGREE. As mentioned above, we now include detailed derivations of both the beta-

binomial likelihood (the reference for the original comment) for the lithostratigraphic observations, and the Student's-t likelihood for the potential-field observations.

P9 l17: please define what 'ID dataset' is.

AGREE. Replaced "field ID dataset" with "field observations dataset"

P9 l25: the 'iGRW' acronym is not used in the rest of the text. Delete.

AGREE. Deleted.

P10 paragraph 1: please add a figure to help the reader to understand how this works. The readership of Solid Earth might not be specialist in MCMC techniques.

AGREE. We now include a new figure (new Figure 3) illustrating the operation of parallel-tempered MCMC, showing trace plots and marginal distributions of each chain operating on an easily visualized example distribution. We hope this will help clarify the algorithm's operation for the reader.

P10 l 13: This information is relevant only if you provide information about the computing resources you used.

AGREE. This section was poorly worded, the true resource use should be measured in CPU-hours and not merely walltime (though arriving at a tractable result in a reasonable walltime is also important). We have revised the text accordingly.

P10 l17: please add in the methodology the definition of the indicators you use to analyse your results.

We have now added more detailed descriptions of the autocorrelation time, the PSRF, and the Geweke score.

P10 l22: I can make an educated guess about what sigma means here, but it needs to be clearly stated.

AGREE. This has been replaced with: "(uncertainties quoted at two standard devia-

tions [2σ] here and throughout)

P11 l13: the reference to Geweke score should come earlier in the text. The ref given might not be the best. I would cite the following instead: Geweke, Evaluating the accuracy of sampling based approaches to the calculation of posterior moments, 1992, Bayesian Statistics 4, pp. 169-193

AGREE. This reference has been replaced here and added to the first paragraph of section 4.1.

P11 section 4.2. Although it's not perfect as an indicator, I'd also give the root-mean-square error. As said above I do not provide review comments on the results section. From here I go straight to section 5.2.

If by "root-mean-square error" the referee is referring to the residuals of the posterior mean forward-model predictions from the data, this is easy enough to add, and we have done so. P14 l20-21: "So, if higher resolution geophysical surveys and/or geological field observations are acquired, the model can then become more precise". This is not necessary as it is obvious.

AGREE. This has been deleted.

P14 l24-25: "Where such regions are under cover and drilling is required to establish formation contacts, our results also aid in constrainingwhich areas should be drilled first to maximize information gain". This is true as a first approximation but not always valid. You can imagine that adding more information in a certain part of the model may improve greatly a portion of the model that is equally uncertain because it is linked to that first structure in a structural or topologic sense.

AGREE. We have added "could also aid" to allow for the fact that our approximation may not always be true.

3. References Bijani, R., P. G. Lelièvre, C. F. Ponte-Neto, and C. G. Farquharson, 2017, Physical- property-, lithology- and surface-geometry-based joint inversion using

Pareto Multi- Objective Global Optimization: Geophysical Journal International, 209, 730–748. Fullagar, P. âĹijK., G. âĹija. Pears, and B. McMonnies, 2008, Constrained inversion of geologic surfaces - pushing the boundaries: The Leading Edge, 27, 98–105. Guillen, A., P. Calcagno, G. Courrioux, A. Joly, and P. Ledru, 2008, Geological mod- elling from field data and geological knowledge. Part II. Modelling validation using gravity and magnetic data inversion: Physics of the Earth and Planetary Interiors, 171, 158–169. Jessell, M., E. Pakyuz-charrier, M. Lindsay, J. Giraud, and E. de Kemp, 2018, Assess- ing and Mitigating Uncertainty in Three-Dimensional Geologic Models in Contrasting Geologic Scenarios: , 63–74. Jessell, M., L. Aillères, E. De Kemp, M. Lindsay, F. Wellmann, M. Hillier, G. Laurent, T. Carmichael, and R. Martin, 2014, Next Generation Three-Dimensional Geologic Mod- eling and Inversion: SEG Special Publication 18, Chapter 13, 261–272. Jessell, M. W., L. Ailleres, and E. A. de Kemp, 2010, Towards an integrated inversion of geoscientific data: What price of geology? Tectonophysics, 490, 294–306. Kamm, J., I. A. Lundin, M. Bastani, M. Sadeghi, and L. B. Pedersen, 2015, Joint inversion of gravity, magnetic, and petrophysical data âAËŸTËĞA case study from a gabbro intrusion in Boden, Sweden: GEOPHYSICS, 80, B131–B152. de la Varga, M., A. Schaaf, and F. Wellmann, 2018, GemPy 1.0: open-source stochas- tic geological modeling and inversion: Geoscientific Model De- velopment Discussions, 1–50. Pakyuz-Charrier, E., M. Lindsay, V. Ogarko, J. Giraud, and M. Jessell, 2018a, Monte Carlo simulation for uncertainty estimation on structural data in implicit 3-D geological modeling, a guide for disturbance distribution selection and parameterization: Solid Earth, 9, 385–402. Pakyuz-Charrier, E., J. Giraud, V. Ogarko, M. Lindsay, and M. Jessell, 2018b, Drillhole uncertainty propagation for three-dimensional geological modeling using Monte Carlo: Tectonophysics. Scholl, C., J. Neumann, M. D. Watts, and S. Hallinan, 2016, Geologically constrained 2D and 3D airborne EM inversion through cross-gradient regularization and multi-grid efficiency: ASEG-PESA-AIG 2016, 1–6. Sun, J., and Y. Li, 2016, Joint-clustering inversion of gravity and magnetic data applied to the imaging of a gabbro intrusion: SEG Technical Program Expanded Abstracts 2016, 2175–2179. Wellmann, J. F., M. de la

Varga, R. E. Murdie, K. Gessner, and M. Jessell, 2017, Un- certainty estimation for a geological model of the Sandstone greenstone belt, Western Australia – insights from integrated geological and geophysical inversion in a Bayesian inference framework: Geological Society, London, Special Publications, SP453.12. Zheglova, P., P. G. Lelièvre, and C. G. Farquharson, 2018, Multiple level-set joint inver- sion of traveltime and gravity data with application to ore delineationâĂŸ ′r: A synthetic study: , 83. âĂČ

Please also note the supplement to this comment:
https://www.solid-earth-discuss.net/se-2019-4/se-2019-4-AC1-supplement.zip

---

Interactive comment on Solid Earth Discuss., https://doi.org/10.5194/se-2019-4, 2019.