Reviewer 1

> Thank you very much for your insightful comments and taking time to review this paper! We agree with your major point about the gravity data noise value, and have re-run all relevant portions of the code used a large amount of noise. This has highlighted gravity noise as a more significant factor in bathymetry inversions than the original manuscript suggested, and we will re-word some of our discussion and conclusion remarks to include this. We have responded inline to each of your comments below with indents.

General comments

This paper aims to provide a robust theoretical background and test for practicality and usefulness of gravity inversion for determining sub-ice shelf bathymetry. Such a paper is useful as it has the potential to guide and optimise future real-world data collection over Antarctic ice shelves. The paper uses a prism-based forward model, coupled with an iterative least-squares approach to provide bathymetric estimates. The test results point towards the importance of higher quality/resolution gravity data in areas of low amplitude background field (simple underlying geology), while direct observations (e.g. seismic or AUVs) become increasingly important where the underlying geology is complex.

Overall the paper is well written and the results appear reasonable. However, I have one specific comment associated with the treatment of gravity errors which I feel should be addressed and a few additional more technical points. This will likely not significantly change the outcome of the paper, but may change the suggested likely minimum achievable error in bathymetry from gravity data.

Specific comments

Around L390 to 395 the authors talk about simulating noise in the gravity data. My understanding of the paper is that the authors simulate noise by first adding random Gaussian noise to the baseline gravity disturbance. This pixel by pixel noise has an amplitude in-line with the errors reported for typical airborne surveys. The initially adulterated data is then re-filtered to achieve a best noise reduction with minimal loss of gravity signal, and the subsequent re-filtered data inverted for bathymetry. However, the data loss from noise and filtering Fig. 10c is consistently below +/-1 mGal, which seems small compared to what would be expected for a real survey.

The authors justify re-filtering the data after adding noise because filtering is a standard method of noise reduction in airborne gravity processing. However, the errors quoted for gravity surveys are after filtering. I therefore don't think this is the best way to simulate noise in a synthetic gravity dataset. I would suggest that a better method would be to create a random Gaussian noise field, which when filtered with a 10 km wavelength filter (to simulate gravity processing) had a 1 mGal standard deviation (equivalent to the error in high quality gravity data). Adding this filtered error field (with likely local maximum amplitudes of +/- 4 mGal) to the baseline gravity disturbance would be more representative of the likely errors in real Antarctic airborne gravity data. Other ways to create realistic noise could be considered. Use of this error field would likely amplify the errors in the recovered bathymetry, giving a higher, but more realistic, estimation of the expected error due to noise in the gravity data.

> This is a good point, thanks for pointing it out. We agree the tested noise value was at the optimal end of realistic noise levels. We have re-ran Tests 2 and 5 using pseudo-random noise with a standard deviation of 3 mGals. This resulting in peak errors of ~12 mGal. Filtering with a 10km Gaussian filter width gave RMSE of ~1 mGal. We have retained the trials of different filtering wavelengths, since we believe these results in themselves are informative. The optimal filtering results was found to be ~24 km, which is generally larger than we would have expected. This reduced noise RMSE to ~0.6 mGal. Additionally, for Ensembles 2-4, we have changed the tested noise limits (pre-filtering) from 0-3 mGals to 0-5 mGals. We will revise the manuscript to reflect these results using the more realistic noise levels. This will result in updates to Figures 8, 9, 10, 13, 14, 15, and 17, as well as all relevant portions of the text.

Technical corrections

L35 and other places in the text (e.g. L277, L343) refer to "regional gravity field strength". It is not 100% clear what is meant by this. My understanding of this in other contexts in the paper is that the authors mean the "amplitude of the variability in the regional field". High field strength could be a uniform value of 200 mGal, but this would have no impact on the inversion quality. I would suggest re-wording.

> Yes you are correct when we refer to `strength` it is the amplitude of the variability, aka the standard deviation. Thank you for point out this confusion. We will update the text to be clearer, for example changing "weak regional fields" to "regional fields with low-variability".

L163 – It is not clear why the sensitivity matrix is populated by the vertical derivative of the gravity. This should probably be justified in a little bit more detail. – I think high gradient areas might have shallower sources so be more sensitive, but this is a guess? This is covered in Appendix 1, which could be cited. However, in the appendix the example of varying density was given. As this is fixed in the inversion then the matrix can be filled just with the gravity gradient. However, the parameter which is varied is the topography, which isn't fixed at each iteration. Therefore is the sensitivity matrix re-computed at each step as well (L191-193)?

> Thanks for pointing out this mistake. We will update the text to read "Each entry in the Jacobian is the partial derivative of gravity produced by a specific prism at a given observation point, with respect to the prism's thickness." I think `partial` as accidentally replaced by `vertical`. And yes because the Jacobian depends on the

current model of bathymetry, it has to be recalculated at each iteration.

L210 – constructing training datasets for Damping value cross-validation. This is done by creating two raster’s – training and testing, which are on meshes with cell size X, shifted by ½ X. In effect taking a mesh with cell size ½ X and considering alternating points. A concern with this is that the mesh size X must leave some ambiguity. For example if you have 10 km wavelength gravity data and training/testing meshes of 100 m both will be in effect identical. Mesh size therefore matters in this case and is related to the wavelengths considered. The mesh size used for generating the observation and test data, or how it could be estimated, should be stated here.

This is not exactly how the training and testing set are created. If the original gravity data was on a 10 km grid, then it is interpolated onto a 5 km grid. The points which fall on the original 10 km grid now make the training set, and the new (interpolated) points which fall between the 10 km grid points are the testing set. With this technique, we don't require an arbitrary mesh size, we just use half of the original cell size. There could be some testing for whether 1/2, 1/4, or 1/8 the mesh size is optimal, but our results (and those of Uieda et al. 2016) appear reasonable, and we think this is out of the scope of this work.

229-237 – Uncertainty constraint. It is not clear if/how the uncertainty is quantified given the control points form part of the inversion, so should have zero offset. Were random control points left out?

We didn't include the depth uncertainty of the constraint points in our uncertainty analyses as there were already many components in our uncertainty analysis.