- **RC2**: ['Comment on essd-2021-304'](#), Jiajia Sun, 01 Nov 2021 [reply](#)

The authors created a data set consisting of 1 million geological models and the associated gravity and magnetic responses using the Noddy package. This is very timely and welcome contribution to the geophysical community. It is broad applications for training machine learning models for predicting geology (including history, geometry, structure, etc) and for testing inversion algorithms (e.g., understanding the non-uniqueness of inversion). As the authors mentioned, the geoscience community suffers from a lack of large, labelled datasets that can be used to validate or train robust Machine Learning and inversion schemes. This contribution is a timely and valid response to this problem. As such, it is my belief that the authors' work fills an urgent need in the geoscience community. I would like to commend the authors for recognizing such a critical need and for developing a first-step solution to it.

The authors also discussed three possible applications of this massive data set in Section 5. They are all highly relevant and deserve future research work. This again highlights the importance of the authors' work documented in this manuscript.

I am also glad to see that the authors recognized the limitations in their current work and discussed several ways to expand and improve the repository of 'real world' geological models.

I also tested the notebook (on mybinder.org) and visited the repository https://cloudstor.aarnet.edu.au/plus/s/8ZT6tjOvoLWmLPx. They both work and are in good shape.

I do not have any major concerns. Below are some minor grammatical and/or clarification suggestions and questions.

Detailed comments

Line 52: 'In the case examined in this study, the total number of publicly available 3D geological models probably numbers less than 10,000000,' What case? Where does this number come from?

This statement was based on an estimate and does not add much to the discussion so has been deleted.

Line 59: 'a very large database of possible outcomes'. Not exactly sure how to understand 'outcomes'. Based on the context, I supposed it means geological outcomes of a series of geological events such as faulting, folding, intrusion, etc. Is that correct?

This has been clarified in the text.

Line 86: Exactly!

Line 87: How is 'implicit modeling' defined? And how is it different from 'explicit modeling' (if the latter exists)?

A definition of implicit modelling has been supplied and compared with CAD-style explicit modelling in line 90.

Line 98-99: Great contribution!

Line 123: 'taken' à 'taking'

fixed

Figure 1: Please double check the 3D visualization in panel (a). Looking from NE to SW, the East face should be in the left and the North Face in the right.

Fixed, it is looking from the NW.

Line 152: 'Monte Carlo sampling'. From the text below, it seems that only Gaussian and uniform sampling were employed when generating the petrophysical and all the other parameters. Does 'Monte Carlo sampling' simply mean random sampling from Gaussian and uniform distributions?

In this case yes, and this has been clarified.

Line 154-157: Great! It is important to make the lithologies consistent with the associated geological events. This is where expert knowledge from geologists can play an irreplaceable role. Just curious about how this was realized. Did the authors develop an automated way of ensuring geological consistency? Manually checking each geological model and evaluating its geological and lithological consistency do not seem practical.

An explanation of the hierarchical grouping of 'associated lithologies' is provided in line 161.

Line 181-183: Please rephrase this sentence, as it is very long and hard to follow.

Sentence has been rephrased

Line 193: 'citations'? *fixed*

Line 196: 'clustering of geophysical fields'. Did the authors mean classification of gravity and magnetic measurements into different classes?

Yes, and this has been specified more clearly.

Line 200: "forward models of the gravity and magnetic response". Not exactly sure what is meant here. Seems to me that 'forward models' is simply a repeat of the 'gravity and magnetic response'. Please rephrase.

This has been rephrased.

Line 211: remove the question mark.

done

Line 219: Remove the word 'and' in the heading.

done

Line 229: suggest replacing 'trial' with 'test'.

done

Line 233-235: Excellent!