

## Response to comments from Wouter van der Wal

184: I suggest to add here or in the conclusion that there is also a potential use case where lateral variations itself might be of interest, for example, where one of the parameters is a factor that converts seismic velocity anomalies to viscosity. In that case it is an open problem how to obtain weights that are general enough that they work for different 3D patterns but still provide accurate emulated results.

To include the suggested idea, the text was revised around line 330 to the following:

... confidence were too low). The second extension would be to include a parameter, or multiple parameters, in the inputs to the ANNs such that some information about the lateral Earth structure is encoded into the networks (e.g., scaling from seismic velocity anomaly to viscosity). This would allow evaluation across multiple realizations of lateral Earth structures without the need for separately trained datasets as was conducted in this investigation. The final extension would be to train the ANNs on multiple ice sheet ...

215: The figure shows the smallest MSE for layer width of 128. Is that a special case and is the smallest MSE for most other cases found for higher widths? This could be clarified when referring to figure S3.

This is indeed a bit of a special case and, unfortunately, not something that generalized to the other datasets. Part of the difficulty in using this figure to compare with the other results presented in the manuscript is that the layer widths in this case were all the same width, and the training and validation datasets were missing the LT dimension (i.e., the LT values were all the same, but still included as a dimension). These shortcomings are noted in the caption of the figure.

---

## Response to comments from AR#1

1) Lines 198/228: What is PV? I guess parameter vector.

Correct; PV is now defined prior to this instance (one line previous).

2) Line 233: Why to the power of "-2"?

Should read to the power of 2; not -2. Text has been updated.

3) Fig2. 2 & 6: Again, you are having sequential data, but are using a divergent color map. Thus, your misfit graphs are difficult to interpret as the lightest color is not related to the smallest MSE. Please consider using a sequential color map, e.g. *batlow* (<https://www.fabiocrameri.ch/colourmaps/>).

While we appreciate the guidance provided; we find the use of our current colour map (vik) to be easier to interpret than the sequential maps provided by Crameri, F. Given the chosen colour palette still meets the requirement to be color-blind-friendly we leave the plots unchanged.

4) Line 333-334: This word construct sounds weird: "The finding is found".

Text has been revised to "This result is generally found across..."

5) Fig. 6: How can you have a total misfit below 0?

The rendering of side-bar triangles on both upper and lower limits is the default behaviour of the software used to generate the figure in question. There were no zero or less-than-zero misfits. The figures (Fig. 6 as well as several supplemental figures) have been updated to remove the lower limit triangle

6) Wu et al. was published in 2013 and not 2012.

It would appear the bibtex data provided by the publisher is providing the publishing date of the article (01 November 2012), not the issue in which the article was published (January 2013). Since the publisher provides the latter as the plain-text citation, our bibtex entry and associated citation has been updated to reflect the 2013 date.