Reviewer 1

Firstly we would like to thank the anonymous reviewer for the time they have taken to read through the manuscript and make their suggestions. The majority of our responses to the reviewer are provided as additional annotations on the reviewer's response (included). Several of the points require a greater volume of detail than is useful in that format, and as such we present an itemized response to these particular comments.

"*This is one of my biggest concerns with this manuscript: Why is this needed? According to lines 104/105 NMSS and Seakon-SS are equivalent.*"
&
"*Coming back to my point above (at lines 127/128): According to lines 104/105, the NM and Seakon model outputs are equivalent for SS structures, so why is there a difference?*"

On lines 104/105 the manuscript states " When lateral variations are not included, Seakon model output is equivalent to the NMSS model (assuming parameter values corresponding to the 3-layer viscosity structure are the same)."

This statement is a simplification which may be the source of confusion for the reviewer. When Seakon is used in the spherically symmetric (SS) configuration, it is functionally equivalent to the NMSS numerical model. Given the significant differences in numerical methods, and subtleties such as the differences in time-stepping and grid-discretization, the model output is not bit-for-bit identical as some readers may interpret the statement on lines 104/105. As such, to avoid introducing any potential source of structural error that may arise from these differences, and also for the ease and expediency of data processing during the investigation, we elected to use Seakon in the SS configuration rather than rely on it solely for the calculations which incorporate lateral variability. We will incorporate a brief summary of the above (as well as the comments made to the other reviewer with respect to this subject) into the text and remove the statement on lines 104/105 to avoid this potential source of confusion.

'*My concern is mainly that ROC of RSL is not used when comparing to observations. ROC of the radial displacement is used, so there it is okay to use the ROC, but I would not use it for RSL.*'

The reviewer indicates that they have an issue with the use of the rate of change (ROC, i.e., the first derivative with respect to time) of relative sea level (RSL) to produce predictions of RSL itself. The reasons for this choice are twofold: Firstly we can easily convert between the two due to the definition of RSL, and secondly, we found reduced misfits using this approach. With respect to the conversion aspect, the reason we can use the ROC RSL in lieu of RSL itself in the emulator is that, by definition of RSL, we always know that RSL at present day is equal to zero. I.e., $RSL(t=0) = 0$. As such, it is a simple matter to integrate the provided ROC of RSL from either the emulator, or actual RSL data itself, to reproduce the RSL timeseries. With respect to the second part, early sensitivity tests revealed nearly an order-of-magnitude reduction in the prediction misfits from the neural networks when using the rate-of-change rather than trying to predict the RSL value directly. As such, we chose to incorporate the ROC RSL into the emulator rather than the direct prediction.

As well, part of the goals of the emulator is to eventually be able to provide input to other numerical models, in which case the ROC of the field is more useful for inclusion into numerical solvers than an absolute value. We will incorporate a brief summary of the above into the manuscript.

“*You discuss the ice model effect of the trained dataset at the end, but how much is the trained dataset depending on the tomography and lithosphere model? Please discuss this as well.*”
&
“*Why these? Why 2/3/5/7 factors of 9? What about 36 and 54? You could test it with N being anything between 1 and 330. Why not testing it for values between 27 and 63 with steps of 1? Of course, this is is time consuming, but could identify where the misfit line deceases most and reaches a "plateau". How is this number also ice, tomography and lithosphere model depending?*”

These values are a result of the initial experimental setup. With the limited compute resources available at the outset of the project we budgeted ~100 3D and SS model parameter vectors (PVs) to explore with Seakon and a limited number of Artificial Neural Networks (ANNs) to train. The subsequent numbers of PVs incorporated into the training datasets of the ANNs were a result of combining two sets of PVs: a core set of PVs and the supplemental PV sets (which have increasing numbers of PVs). The core PVs consist of the 8 most extreme PVs from this initial sampling of the LT/UMV/LMV space (i.e. minimum and maximum of each of the 3 dimensions, resulting in 23 PVs) as well as the center-most PV. The supplemental sets had 9, 18, 36, and 54 members. The lowest bound of 9 PVs was used to determine if we could accurately represent the 3D-SS difference as a function of LT/UMV/LMV with only the core PV set and a single intermediate PV between each of these extreme PVs. From there we doubled the supplemental PV set to 18 and subsequently 36. Finally, the set of 54 was chosen to be half-way between a subsequent doubling and not evaluating the impact of adding another supplemental PV set. The reasoning was that requiring so many members in the training set would indicate this fast-surrogate approach was not sufficiently resource-effective.

This latter reasoning is also why we do not test ANNs with >63 (i.e. 9+54) PVs included in the training dataset. For the purposes of the examples presented, requiring the same order-of-magnitude of computational resources to train the neural networks would indicate that the method employed in this study is ineffective.

As for training and subsequent analysis within the values of 27 and 63 with steps of 1, such an undertaking would be an order of magnitude more expensive to train than what is presented in the investigation. We could potentially add several more intermediate steps, but preliminary work with other tomography models indicates that there is a relationship between the complexity of the 3D Earth structure and the quality of the ANN predictions for a given training dataset size. As such, we argue this scale of investigation would be better suited to a study which explores this approach more fully (i.e., one which varies not only the SS Earth structure, but also ice and tomography). We did do some preliminary investigation work exploring a simpler 3D Earth model, one which uses the same tomography but assumes a spherically symmetric elastic lithosphere, and found that for the same number of PVs included in the training dataset we obtained reduced ANN:model misfits.

We will condense the above reasoning regarding the sizes of the PV datasets and add additional context around line 210. As well, we will include some details regarding the preliminary results of the impact of different 3D Earth structures.

“*Why do you use the NMSS model output here? I understand that the goal is to use it in the future, but here you describe the method and the most obvious would be to compare the* *emulated 3D GIA-SS on SS Seakon to 3D GIA Seakon as these would show you the real differences between emulated and calculated 3D models."*

The choice of using the NMSS results is largely a result of being the eventual workflow with this tool, as pointed out by the reviewer. Previous iterations of the manuscript used the approach (i.e. using Seakon data only) the reviewer specified, but it made no appreciable difference to our results or conclusions. We will clarify this point in the manuscript.

"*A calculation of NMSS+emulated 3D-SS GIA Seakon is okay to have here, but the purpose of doing this is for future application studies. It would be first interesting to see what the differences are between emulated and calculated Seakon models."*

Given the functional interchangeability of the Seakon SS and NMSS output, the differences the reviewer requests are already available in the supplemental materials by comparing the bottom two rows of Figures S10 through to S12 or by comparing Figure S7 to S8.

"*I do not understand why you calculate the emulator:NMSS. For what? To find out the difference between 1D and 3D? But this is not the aim of this paper. It can be a side result, but shouldn't cover that much in the manuscript."*

Calculating the mean-square-error between the emulator and the NMSS is done to demonstrate that, while there are still non-negligible misfits between the emulator and the explicit 3D data, the output of the emulator are significantly more like the 3D output than the NMSS output. We will re-evaluate the amount of discussion dedicated to this comparison.