Dissertation Summary (Discussion, Conclusions, and Future Work)

Notes:

* After taking points out there isn’t a noticeable difference between mean of MSE from cross validation without all points and score seismic constraints function which returns the MSE between the moho estimate and the point constraints for all the points.
* Andes problem should be able to be solved by modelling the subducting slab. The method of cross validation should in theory be very similar in an area without much tectonic activity when compared to South America e.g. Africa, model will just be better due to lack of deep moho. Crust1.0 no model of subducting slab either, but this method still has discrepancies with this. Deeper moho is less variation can be seen in model so leads to underestimation under Andes. Not just subduction zone that causes problems, only one density value that is being constrained for seismic models. Unlikely that estimate of one value will change. In Africa difference won’t be that large.
* Or if were still using south America instead of using a “random” selection for the training set in cross validation by taking say 2/3 of points out based on their geographical location (in blocks) try with Haas method not Uieda 2017.
* Adding in more degrees of freedom in density estimations, like the seismic regionalisation method used in Haas 2020 and this should decrease the mean errors between the model and seismic constraints. However issue arises with manually choosing how many different regions there will be given the exponential increase in computational time in accordance to increasing the number of regions with differing densities.

Summary:

The results attained from the procedure of cross validation indicate that there is not a noticeable difference between the mean of the errors when comparing all the different training sizes. This result is the same for both models with and without the added underplating of Mariani (2013) with variations in both being around 0.1km meaning that for this model the size of the testing set does not matter as all sizes above 2/3 of the data will arrive at the same conclusion. It is worth noting though that the model with the intrusion implemented has on average higher RMS values than that without the intrusion but was initially expected. [add more info here about overall result and what it means]

Modelling a subducting slab to combat uncertainty

Given the fact that the RMS error values give one singular uncertainty on the model this can be skewed by few large disparities between the model and the point estimates. This is very much so the case, in the Andes due to the active plate tectonics in the area leading to a sharp increase in Moho depth when compared to the surrounding depths, something which cannot be modelled properly due to the regularization parameter which keeps the model smooth and without sharp vertical variations over short distances. Given that the regularization parameter was selected through hold-out cross validation it is not plausible to change this value as it can cause instability in other parts of the model, where Moho depths are estimated well when comparing to previous results. One solution that Uieda & Barbosa (2017) has was to implement a separate smoothness regularization for areas in South America such as the Andes or not to have one for these regions at all. Another solution to this problem that may be easier to implement into the code would be to model the subducting Nazca plate using a Slab2 model from rockhound (<https://www.fatiando.org/rockhound/dev/gallery/slab2.html#sphx-glr-gallery-slab2-py>). The only issue with this is that in the model the crust and lithosphere thickness and density would have to be assumed and then these can be mapped onto tesseroids. This procedure will address the shallow Moho under the Andes and the surrounding area and should in theory decrease the MSE values attained from the repeated random sub-sample validation. On top of this it would be straight forward to see how much this method would improve the error estimates on the model by comparing the difference between identically derived models with the only change being one notebook including the Slab2 model and the other not. With the deep Moho values (upwards of 40km) being underestimated in South America for many models including [insert models here, should be found in the literature review] one way to obtain better MSE values would be to change the location of the model to an area that lacks significant tectonic activity. One possible region would be Africa although it does have tectonic activity most notably the East African rift zone although the Moho depths in this region do not surpass 50km. However, if using the same procedure of cross validation here then Africa would present the same issue as South America with the locations and clustering of seismic point data around the coast of the continent. This issue like that of South America would be due to the environment in central Africa along with financial hardship as seismic surveys are expensive to carry out. Although overall the RMS values or errors on the model should be lower as the model can be fit better to the seismic point estimates through regularization keeping the Moho surface smooth.

Even if the focus was turned back to South America the CRUST1.0 model used (Laske et al. 2013) doesn’t consider and model the subducting slab either, though taking this into account the final Moho depth model produced using the code from Uieda & Barbosa (2017) has discrepancies with the CRUST1.0 model [need to expand on this section, ask Leo].

Blocked testing sets

The procedure used here as mentioned used repeated random sub-sample validation to estimate the errors on the model as a result the singular error value attained is the average error on the whole model across the continent of South America. This as a bigger picture doesn’t give a lot of information as the areas of larger and smaller disparity are not known. Rather than using “random” sampling of the data to separate it into the training and testing set instead the data can be split up using a method of blocking. By selecting points for the testing set by geographical area rather than randomly an uncertainty estimate can be calculated for just the region that the seismic point estimate span. This method entitled “Blocking” was used in Haas (2020) to separate South America into 6 regions of different densities with values found using seismological regionalisation. By using this method the likely larger errors seen in the Andes can be separated from the areas of the continent where the model and the point estimates agree very well with each other leading to smaller RMS values in blocked areas with small disparities. However, “blocking” with limited seismic estimates in some regions and large quantities of these estimates grouped in tight clusters may lead to different validating sizes, which as seen in the results is not much of a problem as the 3 different sizes used achieved very similar mean RMS values.