Dissertation Results

Notes:

* Figures from cross validation loop, along with the text files for the data put everything in GitHub repository and send in zip file of that for submission.
* Adding in Intrusion from Mariani paper increases Mean Error but only by a few hundred metres.
* Still large discrepancies between model and seismic moho estimates in Andes

Results:

To test the cross-validation (CV) approach to error estimation this code has been added into the Uieda, 2017 synthetic-crust1 with Moho depth information extracted from the CRUST1.0 model (Laske et al. 2013). In addition this method will be added to the south-america-moho notebook to see if there are any freak disparities between the two notebooks, not that there should be. Both notebooks use the same method of calculating the best model of the Moho from a combination the best fitting hyperparameters and Bott’s method, mentioned as part of the overview of the inversion method from Uieda, 2017.

The cross-validation approach used is [insert type here] and as mentioned in the methodology randomly splits the full seismic data set into a training and testing (validating) set, with the training set compared to the solution to attain cross validation values, the best solution is then selected from the smallest cross validation value which is then scored against the testing set to attain the Mean Square Errors and subsequently the square root of these values which are the difference between the model and the point estimates. This error gives an indication to the average uncertainty in the overall model depth, and mainly to how good the model is where seismic data is not present which is largely the case for South America as most seismic point estimates are situated near the coast.

Cross Validation results from the synthetic-crust1.0

In this trial run, we will split the seismic point estimates into a training and testing set, for a range of different proportions; these include the training size being 2/3, 3/4, and finally 4/5 of the full data. For each training size the data that makes up this training size will be selected randomly from the full data set for 100 iterations. The data that was not selected for each iteration will be put into the validating set and held back for later scoring. For this data set in particular the total number of seismic point estimates was 937 and the locations of these can be seen in Figure [X insert figure here from synthetic1.0 Ln[29]]. Therefore when splitting the data these proportions need to be rounded to the nearest integer hence 2/3 is 625 data points, 3/4 is 703 data points and finally 4/5 rounds to 750 data points. For just this section of the notebook (the cross validation and plotting of the MSE attained from it) it took 2hrs 49mins and 2hrs 48mins for notebooks with and without the underplating added from Mariani (2013) respectively, this was performed on a laptop computer with an AMD Ryzen 5 3500U 2.1GHz processor.

The data attained from the cross validation on the model without the underplating lead to somewhat of a normal distribution with the mean of all the different training sizes being around 2.3-2.4km, marked by the black dashed line on the plots in Figure [X insert histograms for all sizes for no intrusion]. Generally as the training size increased the error on the x-axis decreased albeit not by much with the mean values for training sizes 625, 703, and 750 being [insert mean values here for each training size, no intrusion, could put this as a table]. However these values ranged from between 1.9 to 2.7km in places, although these values reached were of low frequency and are likely from the code selecting a set of values that skew the result to make it either higher or lower than the average error achieved.

Cross validation results from south-america-moho