

Section 2: Extreme value check

In reviewing some of my data with Michael and Nik, several anomalously low values in both Vp/Vs estimates and Mooney's Vp demanded further attention. In the case of the Vp/Vs estimates two stations have been removed for being too noisy, while one has been judged as clean enough to stand as a deviation. The low Vp estimates in the Mooney database were, in all but one cases, data taken from Continental Shelf and Oceanic Plate environments, which have now been removed from future calculations. All of the values were being parsed and computed correctly from the database. New averages will be computed with this refined data set.

Section 3: Split by Azimuthal Cluster

M. Bostock asked that the data be separated by source region and independently calculated to see how close the source region estimates are to the aggregate. I split each station data into two k-mean clusters centred over Japan and Chile and reprocessed. The results show strong inconsistencies in more data than I expected.

Figuring out what to do with this information is difficult so I have written a logical query against this data to provide a first draft of how I should be using it.

$$| \text{japanR} - R | < 2 * \text{std}(R) \quad (1)$$

$$| \text{chileR} - R | < 2 * \text{std}(R) \quad (2)$$

$$| \text{japanR} - \text{chileR} | < 0.055 \quad (3)$$

Where $R = Vp/Vs$ calculated for the dataset as a whole and japanR and chileR and the regional estimates, $\text{std}(R)$ is my calculated bootstrap error estimate. I am using two standard deviation since I believe my error estimates to be low.

The value 0.055 is my arbitrary cutoff number, Vp/Vs values with a standard deviation higher than this are not included in my calculations, so I have used the same number here for consistency.

Total stations examined: 133

For stations failing (1) or (2) and (3) there is a likely significant lateral heterogeneity or some processing flaw and the station may be marked as unusable.

number of stations: 37

For stations failing (1) or (2) but passing (3) the data may be reasonable but my bootstrap error estimate is much too optimistic. Readjustment of error required.

number of stations: 44

For stations passing (1) and (2) and (3), Great.

number of stations: 50

For stations passing **(1)** and **(2)** but failing **(3)** the regional Vp/Vs difference may be above the arbitrary cut-off of ± 0.055 but could be included provided the std Vp/Vs error is not above some other arbitrarily chosen cap (Vp/Vs ± 0.1) and there are visible reflected phases.

number of stations: 2

These numbers raise some questions. Can I report error estimates for the data which is lower than the difference between the Japan source region Vp/Vs estimate and the Chili/Peru source region Vp/Vs estimate? Removing the 37 stations listed above which fail either condition **(1)** or **(2)** and also fail **(3)** is going to damage the representativeness of my data for some Geological Provinces and for coverage of Canada as a whole.

The last figure (shown below) is the same as the first but this time I have included the regional Vp/Vs estimates and increased my error bars to 2 standard deviations. You can see that most show strong correlation with the Vp/Vs estimates I have calculated though some fall outside 2 standard deviations.

I would like to thank you for taking the time to read this review. In the next few days I will publish and email out histograms and maps showing the state of my data so far. Any recommendations, comments and criticism are most welcome. Especially some comment on how to treat my low error estimates and on how to deal with the information provided by the source region separated data test.