Investigating the Yellowstone mantle plume using hierarchical clustering

Matthew Harding

mh826@cam.ac.uk, Homerton College, Supervisor: Sanne Cottaar

May 12, 2019

Abstract

The Yellowstone volcano is a famous example of a mantle hotspot, and is located in the western US. In this paper, a new clustering and stacking algorithm is developed to characterise the behaviour of the Yellowstone plume at mantle discontinuities.

These discontinuities are mapped using P-to-s converted waves. However, since converted phases are usually invisible under the level of the noise, data across a whole region must be stacked together in order to observe a signal. Most stacking methods choose these regions somewhat arbitrarily, so smooth over local variations. These techniques therefore cannot be used to probe small-scale lateral variations in the mid mantle. The stacking method developed here developed uses hierarchical clustering to group data points together based not only on geographical distance, but also on the statistical similarity of their waveforms. This allows small-scale variations in mantle topology to be better resolved.

Applying this technique to a dataset of receiver functions from Yellowstone park, we observe a collection of clusters centred on the Yellowstone plume. They vary a lot from the average background waveform, showing a strong signal from around 340km depth, and a suppressed 410km peak. The signal at 340km is interpreted as being due to a high pressure silica phase transition.

The thickness of the mantle transition zone can be used as a thermometer for the mantle. The resulting temperature map produced closely resembled others produced from previous studies. Using MTZ thickness also gives an estimate of 200K excess temperature for the plume compared to the surrounding mantle, although when correcting for velocity anomalies this goes down to only 50K excess temperature.

A strong signal around 300km is also seen from the west side of the data, which is interpreted as the influence of the subduction of the Juan de Fuca plate leading to the presence of high pressure silica phases in this area of the mantle.

1

Contents

1 Introduction

The Yellowstone volcano is one of the archetypal examples of a volcanic mantle hotspot. The scientific consensus is that the Yellowstone hotspot is fed by an underlying mantle plume [5], but the nature of said plume has been extensively mapped and debated. Figure 1 gives an example of this mapping. Usually, the simple deep-mantle plume hypothesis is invoked to explain hotspots, but in Yellowstone the simple plume hypothesis alone cannot account for all geological observations – for example the lack of evidence for a significant temperature anomaly [19]. This observation, amongst others, must be explained by auxiliary hypotheses. These disputes demonstrate the complexity of the Yellowstone plume, and indicate there is much about its nature that remains to be probed and discovered.
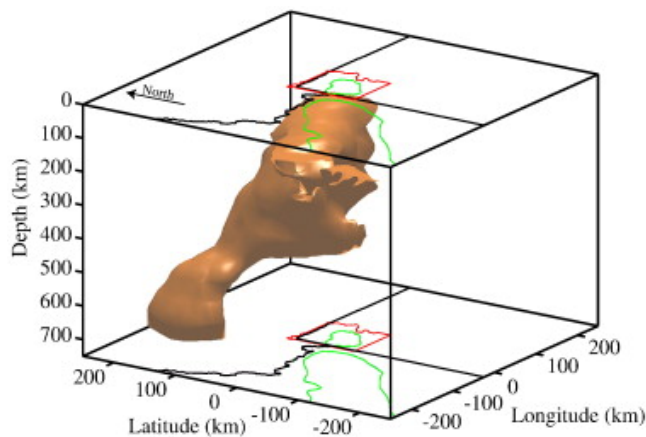

Figure 1: Isosurface of P-wave seismic image of the Yellowstone plume [5]. Reproduced with permission.

One useful method of imaging the mantle is using converted PdS waves. We can use a signal processing technique called receiver functions to investigate these waves and examine discontinuities in the mantle. In particular, in this report we examine the 410 km discontinuity (the P410S phase) and the non-global X-discontinuity, located about 300 km deep in the mantle. The signal from these PdS phases is small, and often invisible under the level of noise. Therefore, in order to make the signal visible, data from multiple sources must be partitioned into bins and then the data in each bin stacked together. Most stacking techniques stack over somewhat arbitrary areas, and therefore will smooth over any lateral variations. If there are

small-scale changes in the mantle topography, these stacking methods will render them invisible by combining data from across these disparate areas.

This report develops a new stacking technique that clusters points together based on both their geographical proximity and the statistical similarity of their waveforms. By clustering similar waveforms together – and therefore exaggerating any interesting features – this technique may be able to distinguish small scale lateral variations in mantle topography where other stacking techniques cannot. Applying this to a data distribution taken from Yellowstone park will yield more evidence on the mantle discontinuities underneath Yellowstone, and on the nature and composition of the Yellowstone mantle plume.

Section 2 will cover the theoretical background that underlies the rest of the report, Section 3 will cover the data distribution and the methods used to process the data, Section 4 will present and interpret the results and Section 5 will discuss any possible further work and Section 6 will summarise the report's conclusions.

2 Theoretical Background

2.1 The Mantle Transition Zone (MTZ)

There are several seismic velocity discontinuities in the mid-mantle, most notably at 410km and 660km. These discontinuities are observed globally, and happen due to changes in temperature and pressure inducing phase transitions in the dominant mantle mineral. The 410km discontinuity is thought to be due to a phase transition from olivine to wadsleyite, and the 660km is thought to be a ringwoodite to bridgmanite and periclase transition. Since these mineral phase transitions can be studied and understood in a laboratory setting, probing the depth of these discontinuities can yield information about temperature variations within the mantle. The Clapeyron slopes for these transitions predict that the 410km transition depresses in hot regions and the 660km is uplifted. Therefore thinning of the MTZ is an indicator of higher temperature [9] and this means, to a certain extent, the MTZ thickness can be used as a mantle thermometer. However other phase transitions in the region (for example, the marjorite garnet transition) with different Clapeyron slopes may complicate the interpretation of the signal [13].

4

Another discontinuity in the range of depths considered is the so-called X-discontinuity. The X-discontinuity refers to any seismic discontinuities observed between 250–350km. This is a non-global discontinuity [1] that is not observed everywhere, and its exact cause is not universally agreed upon. Hypotheses include formation of hydrous phase A in subduction zones, a crystallographic transition in pyroxene from orthorhombic to monoclinic structure, and the transformation in silica from coesite to stishovite [13].

2.2 Converted PdS Waves

When a compressional (P) wave hits a seismic discontinuity at depth d, there is a shear (S) wave refraction. This is illustrated in Figure 2. The resulting shear wave is labelled a PdS phase. Since S waves are slower, this second wave arrives after the initial strong P wave signal on a seismogram. Examining the variation of the characteristics of this wave can help map mantle topography, including determining MTZ thickness. However, the small PdS signals are often difficult to see on a seismogram, and the use of receiver functions is one way to pick out the PdS phases. This is a necessary signal processing step even before stacking is applied, and is discussed in more detail in the Methods section.

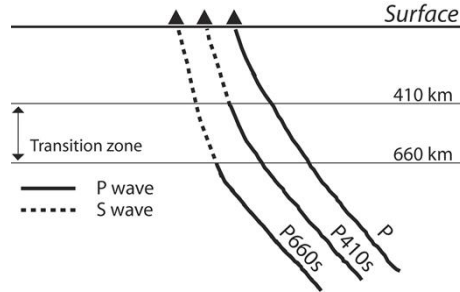Figure 2: Illustrative diagram of the seismic rays hitting discontinuities in the mantle [3].

2.3 Hierarchical Clustering

Hierarchical clustering is a common data science technique. The basic algorithm defines a distance between two data points, for example Euclidean distance. Then, the closest two points are merged together into a cluster, and the rest left as individual points. The process then repeats until all

points are clustered. The final result is a hierarchical tree. This can be visualised as a dendrogram, as seen in Figure 3. This tree can then be split into a chosen number of flat clusters, where the tree structure is discarded and a set of groups containing closely related points is returned as a result. There are many criteria for splitting the tree, for example a maximum cophenetic distance between any two points in a cluster, or a maximum number of clusters [4].
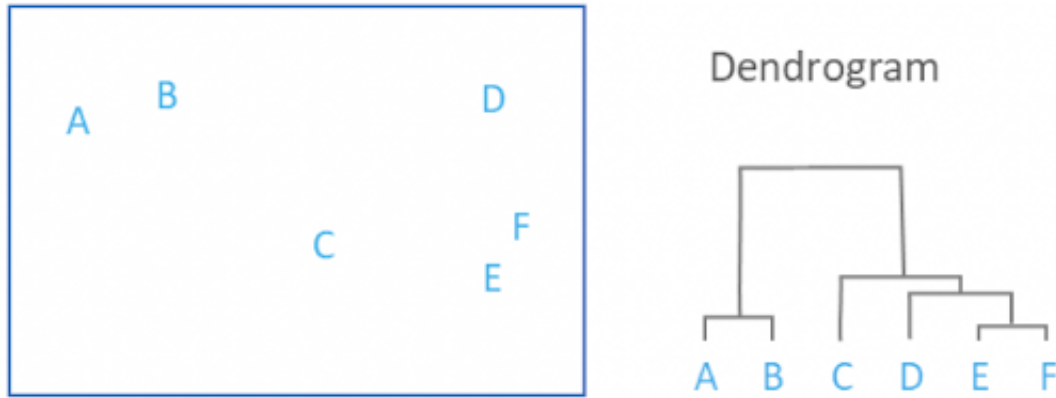

Figure 3: The left hand side shows a set of points, with the dendrogram that would result from applying hierarchical clustering on the right[2]. The points are compared using Euclidean distance.

For this procedure to work, there has to be some way to compare clusters to clusters – often by using the average location in a cluster, but other possibilities include using the minimum or maximum distance between points within the clusters, or taking the centroid [4].

3 Methods and Data

3.1 Data Distribution

The data set is made up of 19292 receiver functions, drawn from January 1976–January 2018 for earthquakes of Mw 6-8 and at epicentral distance from 30-90. It comes from 473 separate stations, most of them from the Transportable Array Experiment (http://www.usarray.org/researchers/obs/transportable). This data ranges across a geographical distribution of about -104 to -118 degrees longitude and 38 to 50 degrees latitude.

3.2 Receiver Functions (RFs)

Receiver functions are a signal processing technique. They can be understood by considering the fact that a seismogram is essentially a convolution of three signals: the earthquake source $S(t)$, the response of the earth $E(t)$ and the response of the instrument $I(t)$, plus a noise signal $n(t)$ [1]. Therefore the signal $X(t)$ from a three-component seismogram, with radial, transverse and vertical components of motion, can be defined as the following, where $\star$ represents a time-domain convolution.

$$X_Z(t) = S(t) \star I(t) \star E_Z(t) + n(t) \tag{1}$$

$$X_R(t) = S(t) \star I(t) \star E_R(t) + n(t) \tag{2}$$

$$X_T(t) = S(t) \star I(t) \star E_T(t) + n(t) \tag{3}$$

The earth response functions are also known as receiver functions (RFs), and the radial RF in particular is usually interpreted as information about discontinuity structure [1]. By approximating $E_Z(t)$ as a delta function, we can deconvolve $X_Z(t)$ from $X_R(t)$ (using an algorithm like iterative time domain convolution) to obtain the radial RF $E_R(t)$. Figure 4 demonstrates how a converted phase at a discontinuity can show up in an RF, but be almost impossible to see in a seismogram.

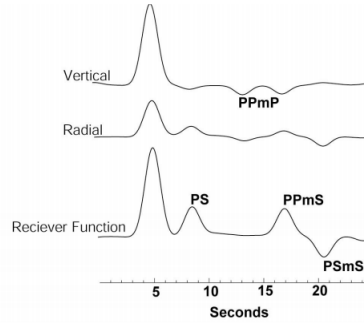
Figure 4: The vertical and radial components of a seismogram, placed above the reciever function that would result from deconvolving one from the other. Signals from the converted phases are marked, and this diagram illustrates how converted phases (in this case, the signal from a p-wave converting to an s-wave at the Moho discontinuity) can show up in an RF, but not in the raw seismogram [8].

3.3 Time-to-Depth Conversion

After generating RFs from the dataset, an arrival time-to-depth conversion is done. This requires choosing a particular earth velocity model to ascertain which depth each arrival time corresponds to. In this paper, we consider two different velocity models: the spherically-symmetric PREM model and the EMC-US-SL-2014 model, a P and S teleseismic body-wave tomography of the mantle beneath the United States [10]. This gives two resulting datasets for each model, each containing a set of RFs that display amplitude relative to the original P-wave along the y-axis, and depth into the mantle along the x-axis.

3.4 Cluster Stacking

Cluster stacking was developed in this project. It is a purely statistical method, rather than being based on any seismological theory. This groups points together based on not only the distance between points, but also the similarity of waveforms, and should therefore be more effective than most stacking methods at isolating small-scale areas in the mantle that are interesting or anomalous in some way.
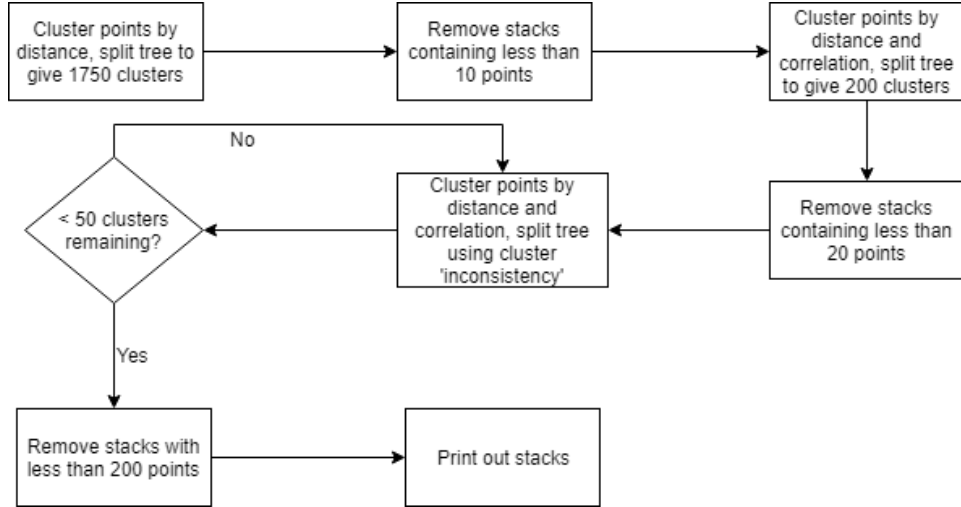

Figure 5: Flow chart showing the overall algorithm for clustering.

Figure 5 shows the process of the clustering algorithm. Essentially, it contains three different clustering phases, separated by removal phases. In

each clustering phase, the metric used to compare points and also the algorithm used to split to final dendrogram (an example of a final dendrogram is shown in Figure 3) are varied, for reasons that will be discussed in the next few subsections. At the end of each clustering phase, all the data in each resulting cluster is stacked, and the next phase begins. In the removal phases, small stacks are taken out of the data. These steps are performed in order to reduce the effects of noise, and to weed out any stacks that are anomalous and not meaningful.

3.4.1 Distance Metric

The distance metric considers both the geographical distance between stations and the L2 correlation between the waveforms. The geographical coordinates of each waveform are the piercing points at 410km, i.e. the coordinates of where the ray crossed the 410km boundary in the mantle. The distance between these points is calculated using the Haversine equation [12]. The distance metric can also include the L2 correlation function. For two waveforms, u and v, the L2 is given by the following equation, where i runs over all depths in the receiver function.

$$L2 = \Sigma_i (u_i - v_i)^2$$

Many other measures of statistical similarity were considered, including L1, L3, Pearson correlation and cross-correlation. In the end, L2 was chosen for pragmatic reasons – the algorithm functioned most effectively with L2. There is a balance to be found when choosing the metric in hierarchical clustering. If the 'distances' between points are too small on average, the algorithm will try to cluster them all into one large cluster, and if the distances between points was too large on average, then the algorithm will try to put each point in its own individual cluster. Ultimately, L2 was the most successful in balancing the metric, with a large range of variation between similar and disparate waveforms.

In the first clustering step in Figure 5, the metric used to compare the 'distance' between clusters is just geographical distance. This first step reduces noise, and gets a small averaging effect before the shapes of the waveforms are compared. Including this step results in much more coherent results. In all subsequent clustering steps, the 'distance' metric is the product of geographical distance and L2 correlation. Using the product of the two rather just L2 correlation on its own is essentially a sanity check – it ensures coincidentally similar RFs on opposite sides of the map don't get clustered

together. Clustering using L2 alone resulted in a lot more discontinuous and scattered clusters, and therefore less meaningful results.

3.4.2 Splitting the Tree

At the end of each clustering phase, the algorithm needs to split the tree into a number of discrete clusters. In this algorithm, there are two methods used to split the tree: *maxclust* and *inconsistent*. The first and second phases use *maxclust*, and the final stage uses *inconsistent*. The *maxclust* method minimises the total distance within each cluster while keeping the total number of clusters below a given limit. This doesn't necessarily find 'natural' clusters.
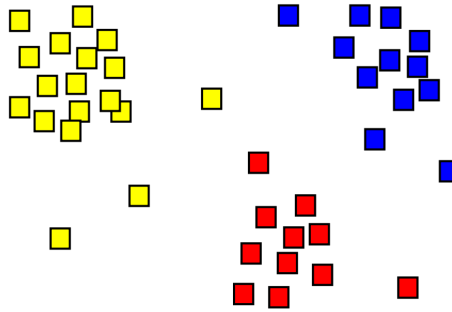

Figure 6: A dataset that can be partitioned into three obvious clusters, marked in red, blue and yellow [20].

For example, if the data contains three very clear clusters, as in Figure 6, and you give the *maxclust* algorithm a cluster limit of two, then however the algorithm decides to split the tree, it won't be in the most natural way.

Inconsistent cuts the cluster such that the inconsistency within a cluster remains below a given threshold [4]. The inconsistency value for two nodes is given by $I = \dfrac{d - \mu}{\sigma}$, where d is the distance between nodes, μ and σ are the mean and standard deviation of link distances, respectively. Essentially, it tries to find natural breaks between the clusters. This algorithm would take the points in Figure 6 and easily split them into three clusters, however with large amounts of data it doesn't reduce the total number of clusters by

10

very many, and so is inefficient and slow.

Therefore, *maxclust* is used to predictably bring the number of clusters down in the first two phases, and *inconsistent* is used to find 'natural' clusters in the last phase. This makes the final set of clusters more robust and resistant to small changes in input data.

3.4.3   Comparing Clusters

The method of comparing one cluster to another cluster during a clustering phase must also be considered. In this algorithm, the 'distance' between clusters is calculated using by finding the average 'distance' using the equation below, where u and v are clusters with cardinalities $\mid u \mid$ and $\mid v \mid$, made up of individual points labelled by i and j.

$$d(u, v) = \sum_{ij} \frac{d(u[i], v[j])}{\mid u \mid \star \mid v \mid}$$

Let's consider the implications of this for making comparisons using L2 correlation. This means that when comparing two clusters using L2, we compare the average correlation between all the constituent waveforms making up a cluster with every single waveforms in the other cluster. Another possible way to compare cluster correlation would be to stack all the waveforms in each cluster and compare the overall stacks. Figure 7 demonstrates the difference between the two methods. While both methods will be effective in clustering if used consistently, the latter is a better method. This is because it is more efficient, and also because comparing large stacks rather than individual waveforms also reduces the effects of noise. The individual RF waveforms that we start with can vary wildly from each other, so noise is an issue. Unfortunately the constraints of SciPy's algorithm necessitated doing it the first way. With more time, I would certainly write my own hierarchical clustering algorithm (or edit SciPy's) in order to permit this.
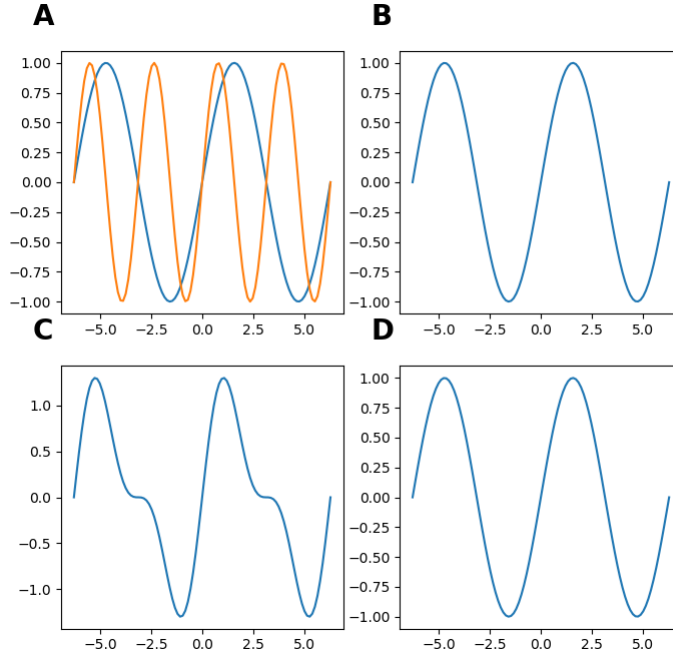
Figure 7: A - two seperate waveforms that make up a cluster. B - a second cluster, containing only one waveform. C - the stack resulting from stacking the waveforms in A. D - the same cluster as in B. All waveforms contain 100 points. If we use L2 correlation to compare each individual waveform in A to the waveform in B, then take the average, we get a value of 49.5. If we take the renormalised stack of the two waveforms (as shown in C) and compare it to the same waveform (shown in D), we get a value of 12.375. The two methods clearly yield different values.

3.4.4   Code Details

All scripts were written in Python, and come up to approximately 400 lines of code. All hierarchical clustering was done using SciPy's clustering package, giving an algorithmic time complexity of $O(n^2)$ [4], where n is the initial number of data points. Wherever possible, the routines in the program were vectorised for efficiency.

All scripts used in the algorithm can be found at the public repository https://github.com/stevinus1/Stacking˙Algorithm.

3.5 Other Stacking Methods

A brief overview of some other common stacking methods.

3.5.1 Epicentral Distance

All data within a certain epicentral distance from a given event is stacked together, and so on in a series of concentric circles outward from the event.
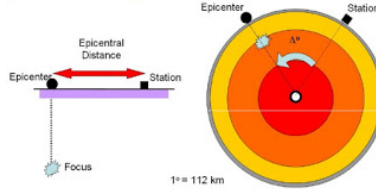
Figure 8: Diagram showing epicentral distance across the surface of the Earth [6].

3.5.2 Common Conversion Point

Common Conversion Point (CCP) stacking back-propagates rays down into the mantle, and stacks receiver functions from around the ray path [13]. Usually, every RF in the stack is weighted by a factor, for example one based on the Fresnel zone half width of the wave [7].

4 Results and Discussion

4.1 Overall Results

The preliminary results of applying the algorithm to the PREM-converted dataset are shown below.
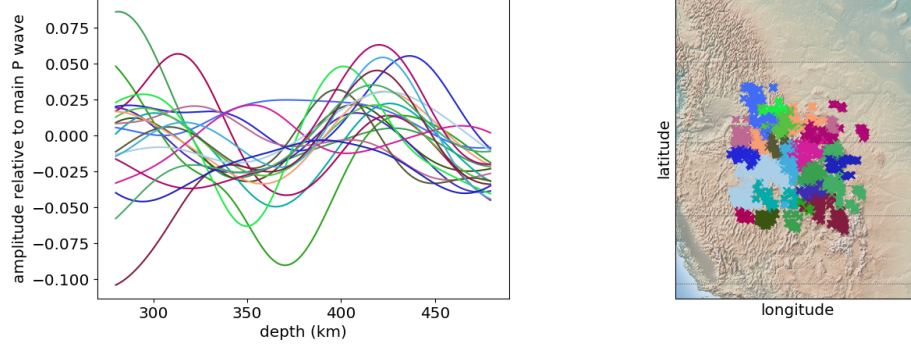
Figure 9: Results of applying cluster stacking to our dataset, for 280km to 480km depth in the mantle. The waveforms corresponding to each cluster are shown on the left, with the location of clusters indicated in the corresponding colour on the right. The RFs are converted from time to depth using the PREM velocity model.

Figure 10: Results of applying cluster stacking to our dataset, for 280km to 780km in the mantle. Results shown as in Figure 9, and conversions are done using PREM.

We now apply the method to a set of velocity corrected data. Instead of using the spherically-symmetric PREM velocity model to do time-to-depth conversions, we use the US-SL-2014 velocity model, which includes very small scale variations in P and S wave speed across the US.
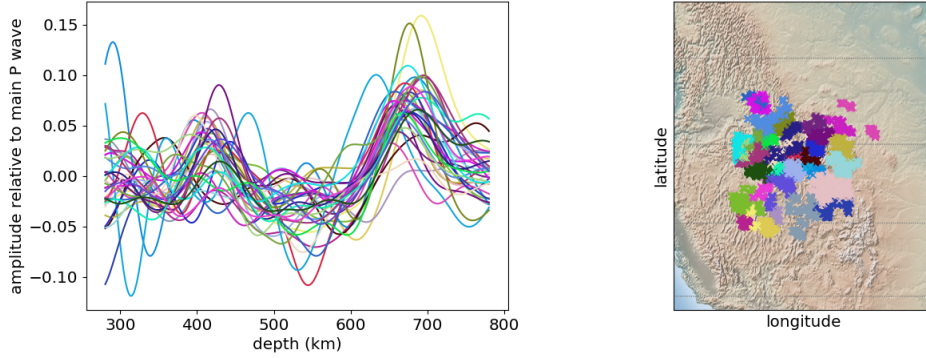
Figure 11: Results of applying cluster stacking to our dataset, for 280km to 780km in the mantle. Results shown as in Figure 9, and conversions done using US-SL2014, depth range 280 to 780km.

In general, all these results look reasonable, and the clustering method seems to work. While there is a lot of variation, most stacks in Figure 9 show a clear peak at around 410km, and Figures 10 and 11 show peaks around 410 and 660, as expected by our knowledge of discontinuities in the mantle. There are some discontinuous clusters in both figures, but in general the distance weighting results in them being contiguous.

Comparing Figures 9, 10 and 11, we can see that they contain quite different final clusters. Since the final step of the algorithm doesn't choose a fixed number of final clusters, Figure 10 also contains more small clusters than the other two. This variation between figures is expected, as the algorithm is inherently chaotic. The results each clustering step is contingent on the results of the previous step, the procedure will naturally produce very different outputs from small changes in input. All these figures are difficult to read any interpretations into initially, but more detailed comparisons can be made and interesting results will be examined in later sections.

4.1.1 Qualitative comparison to CCP stacking

Figure 12 shows that cluster stacking gives far more variation between the different waveforms, with the 410km peaks at different depths, and more variation in the shapes of the waveforms. This is as expected, as this method is intended to pick out more small-scale variation. One thing to note is that the cluster stacks can vary in the amount of waveforms that make them up, so some of the really anomalous waveforms could simply be smaller stacks
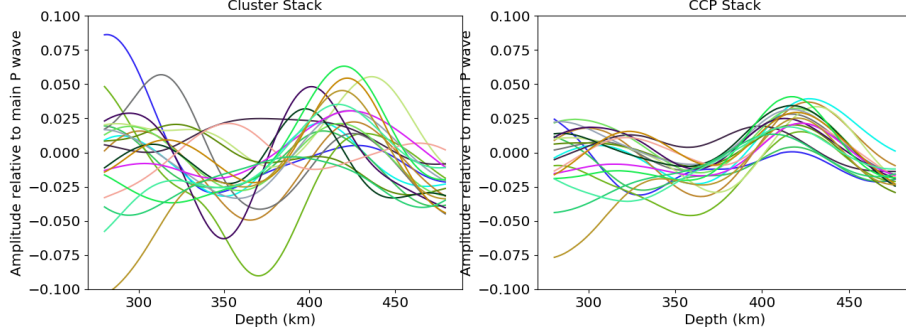
Figure 12: Same stacks seen in Figure 9 on the left, with the corresponding CCP stacks on the right. Depth range between 280km to 480km. The average location of the cluster stacks is used to evaluate the CCP stacks, i.e. using the centre of the cluster as a starting point for back-propagating the ray.

that still have a significant contribution from noise. However, the final removal step – shown in Figure 5 – is intended to reduce this, and the fact that a large amount of variation remains even after this step is promising.

The cluster stacks also have a larger waveform amplitude, on average. This shows the method works as expected, as clustering similar waveforms together will result on more constructive addition and less cancellation, so a greater amplitude than for CCP stacking.

4.1.2 MTZ thickness and temperature map

We can use MTZ thinning as a thermometer, as discussed in the theoretical background. We assume Clapeyron slopes of +2.7 MPa/K and 1.3 MPa/K for the 410 and 660 transitions, and use the same method as Maguire, Ritsema and Goes to convert from MTZ thickness to temperature [15]. The US-SL2014 data is used as a basis for these temperature estimates. This is a good way to get a more intuitive understanding of the many waveforms that make up these results.
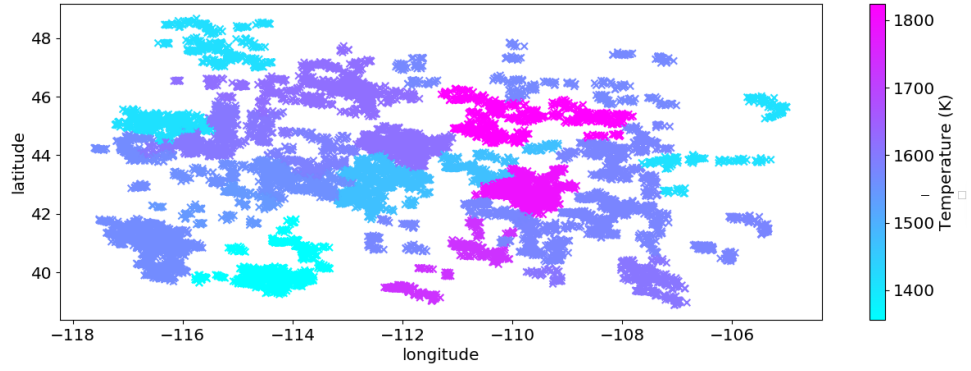
16

Figure 13: Temperature heat map calculated using the data shown in Figure 11. Temperatures in Kelvin.
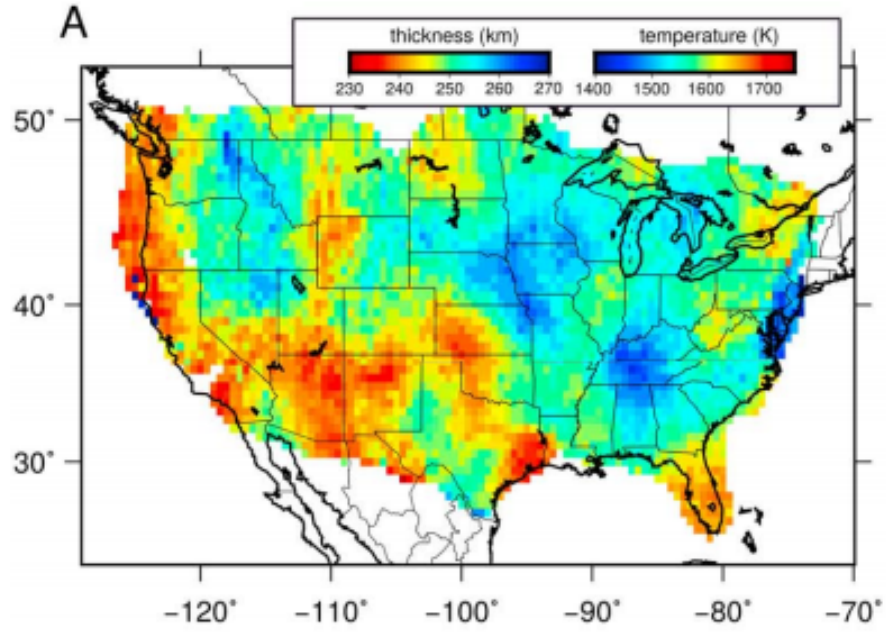

Figure 14: Map of transition zone thickness and temperature, made using P410S and P660S arrival times [15]. From a previously mentioned study by Maguire, Ritsema and Goes. Reproduced with permission.

Comparing Figure 13 to Figure 14 is promising – there is a very similar

17

hot plume of about 1800K running through the data at about -100 degrees longitude, and very cool areas at about 1400K around 40N, 114W and 48N, 116W in both our results and theirs.

Most of Figure 13 sits around 1500K, but a hot plume around 43N, 110W stands out from the background, and apart from this no areas seem particularly anomalous. This will be considered in the next section.

4.2   The Yellowstone Plume

On the surface of the Earth, the Yellowstone caldera is centred at 44.4123 N, 110.7232 W. Since we're using the 410km piercing points as our measure of distance, it is valid to assume the plume is still located close to 44N, 111W at this depth in the mantle [5]. Examining the individual clusters from Figure 9 shows that for PREM converted data, there is an anomalous waveform centred at that point.
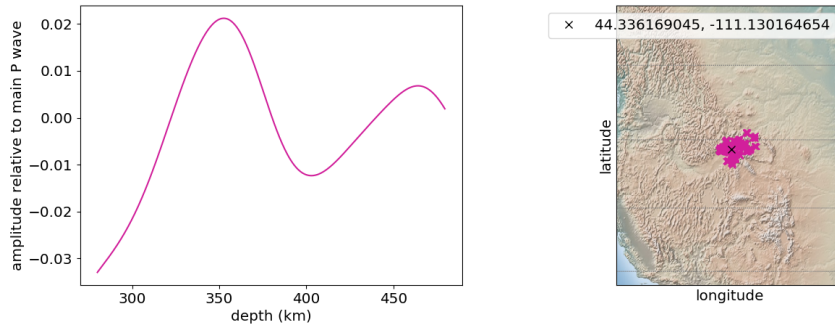

Figure 15: A single cluster isolated from the results shown in 9, located at the Yellowstone caldera. Waveform shown on the left hand side, cluster location on the right hand side.
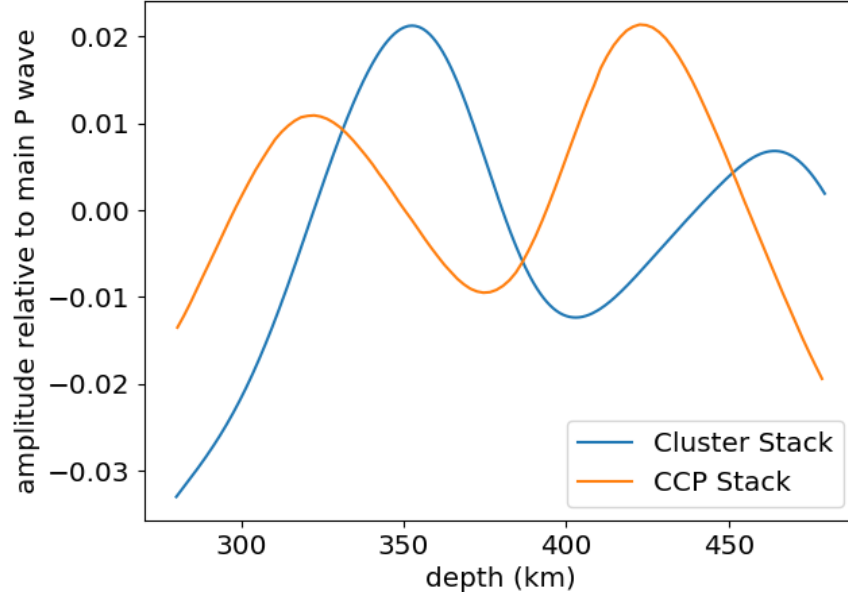
Figure 16: Left hand side is a plot of the plume stack, as shown on the LHS of 15, and on the RHS there is CCP stack, centred at the same location.
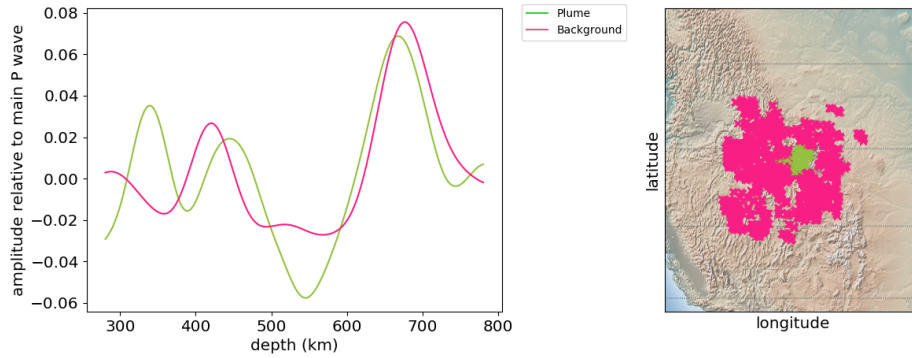

Figure 17: The waveform of the stack centred on the plume, shown in light green, compared to the background average, shown in pink. The RHS shows where the clusters correspond to on a map. This figure was constructed by isolating the clusters centred on the Caldera in Figure 10, stacking them, and plotting them against a stack of all other data.

The waveform shown in Figure 15 is distinctive. It has a 410km peak that is deeper than average – at around 450km rather than at 410km – and a large peak around 350km. Figure 16 shows the difference between this waveform and the results of CCP stacking – these distinct peaks don't show up on the CCP stack at all.

Extending our analysis out to a larger range of depths (280-780km) and stacking together all the clusters centred near the caldera, we get the results shown in Figure 17. This figure compares to plume waveform to the average of all the other RF data. Examining it, we see the same phenomenon – a suppressed 410km peak and a strong peak around 330km.

4.2.1   Excess plume temperature

We can use Figure 17 to compare the plume temperature to the background. Calculating MTZ thickness for both these waveforms gives 223km for the plume, and 256km for the background. The average global MTZ thickness is estimated to be 242km [11]. The background MTZ is therefore measured to be about 10km thicker than the global average. This concurs with the results of a study by Maguire, Ritsema and Goes from 2008, which suggests that the MTZ is thicker than average in this region [15], apart from the anomalous hotspot at the plume. This is shown in Figure 14 – the blue area across most of the data range of interest (-104 to -118 degrees longitude and 38 to 50 degrees latitude) in that figure corroborates our results.

The plume MTZ is about 20km thinner than the global average. A thin MTZ is expected, but this is 10km thinner than a previous result for the plume suggested by Maguire, Ritsema and Goes [15]. This gives a plume temperature of about 1800K, which is around 150K than the global average, and 200K higher than the background. In 2009, Smith and O'Connell et al estimated the excess temperature of the Yellowstone plume to be 70-120K at 410km depth, using a geodynamic model of the plume to obtain these estimates. This significantly lower than the 200K estimate obtained here, so this data may suggest the Yellowstone plume is stronger and hotter than previously thought. These estimates of excess temperature compare to values of about 250K for the Hawaii and Iceland mantle plumes [5].

4.2.2 Interpretation of the plume waveform

This deep 410 peak and large peak around 350 are similar to the results seen for the hotspot mantle plume in Hawaii in a currently unpublished paper by Kemp and Cottaar, et al [13]. This paper shows a similar suppressed 410km peak and a strong peak around 340km for Hawaii. It is suggested in this paper that this is due to the X-discontinuity in the mantle, and puts forward the silica phase transition hypothesis (coesite to stishovite) as the most likely cause. This hypothesis also seems plausible for the Yellowstone plume, and apart from Hawaii there is evidence of a 300km discontinuity at many other mantle hotspots, including Iceland, Samoa and Cobb Seamount [17]. The silica transition hypothesis suggests that the plume is thermo-chemical in nature, rather than just thermal. In other words, the anomalous nature of the plume may be due to compositional differences from the surrounding mantle as well as a temperature anomaly.

4.2.3 Effects of velocity correction

Next, we isolate the clusters centred on the caldera in Figure 11 and plot them. The resulting plume for the velocity-corrected data is centred at about 44N, 112W.
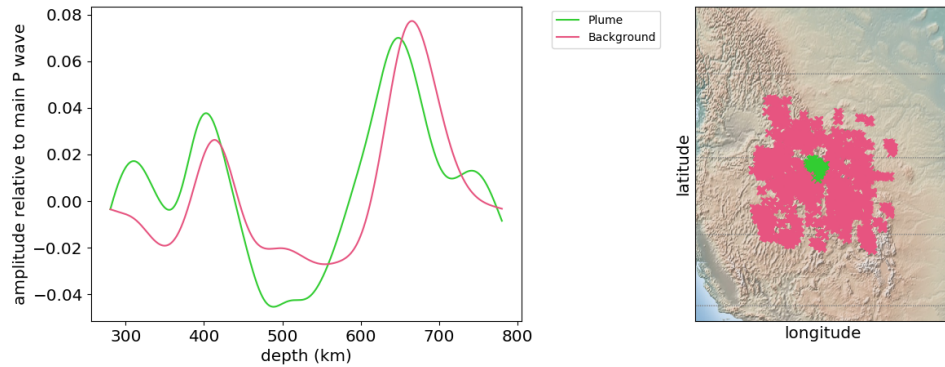

Figure 18: Comparison of plume waveform, shown in lime green, to the background data, shown in pink. All depth-time-conversions are done using the SL2014 velocity model.

Figure 18 shows that the plume cluster has retains the peak around 300km, although it is now smaller than the 410km peak. Both peaks have been shifted to shallower depths. The shallower peak was previously at

about 340km, and is now much closer to 300km. The 660 peak has also moved shallower, to about 620km. This change is expected, as the Yellowstone plume is known to be an area of anomalously low velocity [10]. This also provides further suggestion that the shallower peak is due to a silica phase transition, as that particular phase transition is usually found closer to 300km than the peak we initially observed [17].

However, in this case we don't see the same temperature difference between the background and the plume. The difference in MTZ thickness for this data 6km, which would indicate a plume temperature excess of about 50K, significantly less than our previous estimate of 200K. This is more in-line with previous analyses of the Yellowstone plume, which don't show a significant temperature anomaly [5]. It is also notable that the isolated plume has shifted west by about one degree compared to the PREM data, indicating that some of the distinctiveness of the waveform shown in Figure 15 is due to velocity anomalies, and that the clustering has clustered anomalous results together, resulting in the exaggerated temperature estimate. Alternatively, this change could show a flaw with the clustering method – perhaps in this set of data, the plume hasn't been totally isolated from the surroundings. That would result more of an averaging effect, resulting in a smaller difference of the plume from the background. This might also be why the 410 peak here is larger compared to the results for the PREM data.

4.3   X-Discontinuity distribution

Next we examine the geographical distribution of signals from the X-discontinuity within our data. To do this, all clusters that showed a peak at 300km that had a comparable amplitude to the peak at 410km were stacked together to obtain a map estimating the presence of the X-discontinuity across the data.
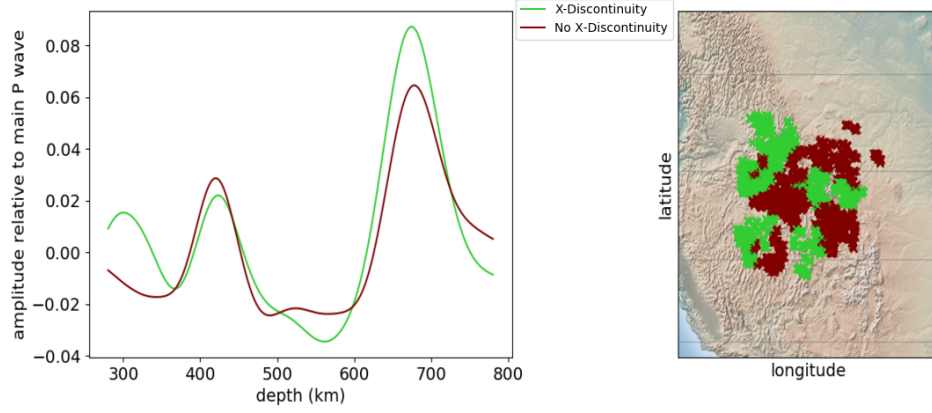
Figure 19: Average waveform with a strong peak at 300km plotted in green, with average waveform lacking a strong peak at 300km plotted in maroon. Time-to-depth conversions done using PREM velocity model.
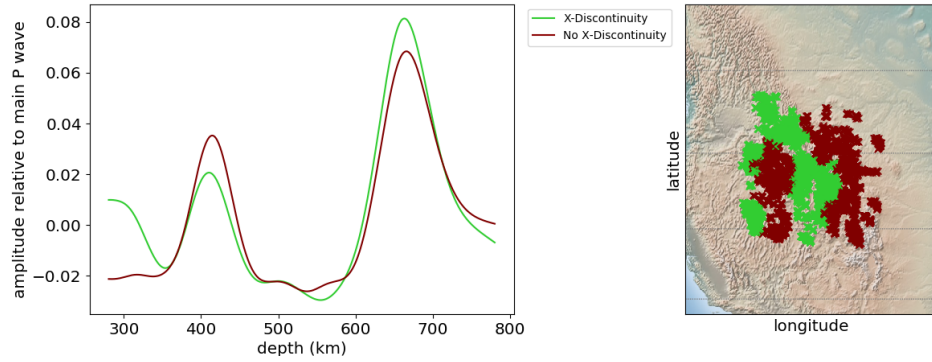

Figure 20: Average waveform with a strong peak at 300km plotted in green, with average waveform lacking a strong peak at 300km plotted in maroon. Same as Figure 19, but with conversions done using SL2014 velocity model.

The distributions shown in Figure 19 and Figure 20 are not at all identical, but they have a lot of commonalties. The X-discontinuity is present in both at the Yellowstone plume, and also in many clusters to the west side of the data.

4.3.1 Interpretation of the X-discontinuity distribution

I hypothesise that this strong presence of the X-discontinuity between -118 and -116 degrees longitude (seen in both Figure 19 and 20) may be due to the Juan de Fuca microplate subducting under this area. In a 2016 study by Tauzin and Reynard, receiver function imaging of the Juan de Fuca plates postulates its subduction zone dropping down to 400km in the range of about -120 to -116 degrees lonitude across the whole range of latitudes considered in this report [18]. High pressure silica phases are major mineral components of oceanic crust like the Juan de Fuca plate, and the ecolgitic silica phases necessary for this transition are usually associated with old subducted plate [17]. This lends credence to the hypothesis that material from the Juan de Fuca plate contributes to the presence of an X-discontinuity.

5   Further Work

This clustering method could be extended further and applied to other data sets, to investigate mantle discontinuities for many different areas. The method could also be extensively improved upon. Many of the numbers seen in Figure 5 are arbitrary and tailored to work the most effectively with this particular dataset. Running a few clustering steps at the start and seeing the results might give enough information for the algorithm for automatically tailor itself to an arbitrary dataset, with an arbitrary distribution of data. As mentioned in the section on Comparing Clusters, the algorithm could also be retooled to compare the similarity of waveforms in a more efficient way.

Further work could also be done in confirming these results. More detailed study of the 200-400km range of depths could characterise the X-discontinuity is these areas further. Other estimates may also be made from it - for example the relative size of peaks in our data could be used to estimate the basaltic composition of the mantle [15].

6   Summary and Conclusions

We apply a new stacking method to 19292 receiver functions to image seismic discontinuities beneath Yellowstone. This reveals the unique characteristics of the converted phases in the Yellowstone plume. A signal from around 300km in the area of the Yellowstone caldera could be evidence of a silica

phase transition, and the thermo-chemical nature of the plume.

Using MTZ thickness to estimate temperature gives a temperature map for this area that concurs with the results of previous studies. Comparing the plume to the background gives an estimate for excess temperature of about 200K, however this decreases to 50K when velocity correction is applied to the data. Apart from the plume, the X-discontinuity is particularly prominent on the west side of the data. This could be due to subduction from the Juan de Fuca plate resulting in high pressure silica phases in this area.

References

[1] *The seismic signature of the Iceland mantle plume from crust to mid-mantle*, Jennifer Jenkins, Department of Earth Sciences, University of Cambridge

[2] *What is Hierarchical Clustering?* Tim Bock, (URL: https://www.displayr.com/what-is-hierarchical-clustering/).

[3] J. Ritsema, P. Cupillard, B. Tauzin, W. Xu, L. Stixrude, C. Lithgow-Bertelloni, *Joint mineral physics and seismic wave traveltime analysis of upper mantle temperature*, Geology , 37 (4): 363366, DOI: https://doi.org/10.1130/G25428A.1

[4] *scipy.cluster.hierarchy*, Copyright 2008-2019, The SciPy community. Last updated on Feb 10, 2019. Created using Sphinx 1.7.2. (URL: https://docs.scipy.org/doc/scipy/reference/generated/scipy.cluster.hierarchy.html#scipy.cluster.hierarchy)

[5] R.B. Smith, et al. (2009), *Geodynamics of the Yellowstone hotspot and mantle plume: Seismic and GPS imaging, kinematics, and mantle flow*, Journal of Volcanology and Geothermal Research 188 (2009) 2656, DOI: 10.1016/j.jvolgeores.2009.08.020

[6] *Basic of Civil Engineering*, Dr. Mohd. Zameeruddin, Assistant Professor, MGM's College of Engineering, Nanded, (URL: http://mzsengineeringtechnologies.blogspot.com/2015/08/element-of-civil-and-environmental.html)

[7] *Largescale mantle discontinuity topography beneath Europe: Signature of akimotoite in subducting slabs*, S. Cottaar A. Deuss (2015),

JGR: Solid Earth 28 December 2015, Volume 121, Issue 1, DOI: 10.1002/2015JB012452

[8] *Introduction to Receiver Functions*, Incorporated Research Institutions for Seismology

[9] *Seismic Observations of Mantle Discontinuities and Their Mineralogical and Dynamical Interpretation* A. Deuss, J. Andrews, E. Day (2013), John Wiley and Sons, Ltd. (2013), pp. 295-323, Chapter 10

[10] *P and S teleseismic body-wave tomography of the mantle beneath the United States*, Schmandt and Lin (2014), Geophys. Res. Lett., 41, DOI: 10.1002/2014GL061231.

[11] *Global mapping of topography on transition zone velocity discontinuities by stacking SS precursors.*, Flanagan, MP, Shearer PM (1998) Journal of Geophysical Research - Solid Earth, DOI: 103:2673-2692.

[12] *Calculate distance, bearing and more between Latitude/Longitude points*, Movable Type Ltd, 31 York Terrace, Cambridge CB1 2PR, (URL:https://www.movable-type.co.uk/scripts/latlong.html)

[13] *X-discontinuity and mantle transition zone structure beneath the Hawaiian hotspot using P-to-s conversions* M. Kemp, J. Jenkins, J. Maclennan, S. Cottaar, Bullard Laboratories, Department of Earth Sciences, University of Cambridge

[14] *Determination and analysis of long-wavelength transition zone structure using SS precursors*, C. Houser, G. Masters, M. Flanagan, P. Shearer (2008), Geophysical Journal International, Volume 174, Issue 1, Pages 178194, July 2008, DOI: 10.1111/j.1365-246X.2008.03719.x

[15] *Evidence of Subduction-Related Thermal and Compositional Heterogeneity Below the United States From Transition Zone Receiver Functions*, R. Maguire, J. Ritsema, and S. Goes (2018) Department of Earth and Environmental Sciences, Geophysical Research Letters, 45, 89138922, DOI: 10.1029/2018GL078378

[16] *Variations in Global Transition Zone Thickness*, G. R. Foulger and B. R. Julian, MantlePlumes, last updated March 2nd, 2004, (URL: http://www.mantleplumes.org/TransitionZone.html)

[17] *Ancient subduction, mantle eclogite, and the 300 km seismic discontinuity*, Q. Williams, J. Revenaugh, Geology 2005;33;1-4, DOI: 10.1130/G20968.1

[18] *Multi-mode conversion imaging of the subducted Gorda and Juan de Fuca plates below the North American continent*, B. Tauzin, T. Bodin, E. Debayle, J. Perrillat, B. Reynard (2016), Earth and Planetary Science Letters Volume 440, Pages 135-146, DOI:10.1016/j.epsl.2016.01.036

[19] *Yellowstone*, G. R. Foulger, MantlePlumes, last updated 9th February, 2006, (URL: http://www.mantleplumes.org/Yellowstone.html)

[20] *Cluster analysis*, Wikipedia, the Free Encyclopedia, https://en.wikipedia.org/wiki/Cluster_analysis