Available online at www.sciencedirect.com

ScienceDirect

Geochimica et Cosmochimica Acta 265 (2019) 443–467

Geochimica et Cosmochimica Acta

www.elsevier.com/locate/gca

Statistics and segmentation: Using Big Data to assess Cascades arc compositional variability

Bradley William Pitcher [a,*], Adam J.R. Kent [b]

[a] Vanderbilt University, Nashville, TN, United States
[b] Oregon State University, United States

Received 10 July 2018; accepted in revised form 28 August 2019; Available online 16 September 2019

Abstract

Primitive and mafic lavas erupted in the Cascades arc of western North America demonstrate significant patterns of along-arc heterogeneity. Such compositional diversity may be the result of differences in mantle melting processes, subduction geometry, regional tectonics, or compositions of the slab, mantle, or overlying lithosphere. Previous authors have partitioned the arc into four geochemically distinct segments in order to assess the importance and relative roles of these processes (Schmidt et al., 2008). However, despite the significant amount of data available from the Cascades arc, no previous study has utilized a statistical approach on a comprehensive dataset to address arc segmentation and its petrogenetic causes. To better characterize the heterogeneity of the entire arc, we first compiled >235,000 isotopic, major, and trace element analyses (glass and whole rock) from over 12,000 samples, which range in composition from mafic to felsic, and include data from arc-front and back-arc centers. We focus on the 2236 mafic arc-front samples in our compilation in order to assess potential causes for along-arc differences in less differentiated magmas, and to potentially lessen any effect of crustal assimilation. To minimize inherent sampling bias – the effect where well-studied volcanoes heavily weight conclusions – we use a weighted bootstrap Monte Carlo approach in which the probability of a sample being selected to the posterior distribution was inversely proportional to the number of samples within its 0.25° latitude bin. This methodology produces a more uniform and unbiased distribution from which we can assess regional, rather than local, compositional variability in the Cascades arc. Using a multivariate statistical approach, we demonstrate that the four segments designated by Schmidt et al. (2008) are, in fact, statistically distinct. However, using a modified hierarchical clustering mechanism, we objectively divide the arc into six regions which have geochemical differences that are up to 6.3 times more statistically significant than in the previous scheme. Our new, more robust segmentation scheme includes the Garibaldi (49.75–51°N), Baker (48.5–49.75°N), Glacier Peak (47.75–48.5°N), Washington (45.75–47.75°N), Graben (44.25–45.75°N), and South (41.25–44.25°N) Segments. By partitioning the arc into the most statistically distinct segments and calculating unbiased mean compositions for each, we explore the petrogenetic causes for the regional-scale differences in primitive lava compositions. These bootstrapped mean data indicate significant inter-segment differences in fluid-flux signature, mantle fertility, and depth and degree of melting. We suggest that differences in subduction geometry, regional tectonics and mantle heterogeneity are the primary causes for these intra-arc differences.

Keywords: Cascades Arc; Segmentation; Basalt; Geostatistics

1. INTRODUCTION

The composition of magmas erupted at arc volcanoes bear evidence of the complex interplay between geochemical contributions from subducted oceanic crust, sediment,

* Corresponding author.
E-mail address: bradley.w.pitcher@vanderbilt.edu (B.W. Pitcher).

https://doi.org/10.1016/j.gca.2019.08.035
0016-7037/

and liberated fluids, and from the mantle wedge and overlying lithosphere. The composition and mass contributed from each of these varies worldwide and even within a single arc (Hildreth and Moorbath, 1988; Elliott, 2003; Carr et al., 2004; Schmidt et al., 2008). This diversity is made even more complex by the fact that other parameters such as the rate and angle of subduction, slab age, mantle flow patterns, and thickness and tectonics of the overlying lithosphere may also greatly affect the composition of arc magmas (Patino et al., 2000; Syracuse and Abers, 2006; Till et al., 2013). Comparing along-arc compositional changes to variability in these parameters provides an excellent means by which to identify those factors that may be most responsible for production of heterogenous magmas within a single arc system. This, in turn, provides a more comprehensive understanding of the complex magmatic processes that occur within arcs worldwide.

Systematic intra-arc changes in volcanic rock compositions have been demonstrated for many systems around the world, and numerous causes have been proposed for each. For example, in the Central American volcanic arc, systematic trends towards lower La/Yb and higher U/Th and Ba/La in mafic lavas from the Nicaragua portion of the arc may indicate higher degree partial melting and greater contribution from the slab (Carr et al., 2004). This trend may be the result of a steeper slab angle in the central arc which could act to concentrate fluid-flux (Patino et al., 2000; Shaw et al., 2003). Wörner et al. (1994) explore geochemical variations in volcanoes from the Central Andes (17.5–22°S) and suggest that crustal age is the predominant factor producing heterogeneity since crustal thickness, sediment supply, slab depth, and distance from trench are invariant over this portion of the arc. Along-arc changes in the Indonesian Sunda arc have been attributed to mantle heterogeneity and slab depth (Whitford et al., 1979) or differing sediment compositions and degree of melting (Turner and Foden, 2001). Thus, since such along-arc trends are common worldwide, understanding the causes for such diversity is a crucial question in the field of igneous petrology.

Numerous authors have also demonstrated along-arc compositional variability of the Cascades volcanic arc (Bacon et al., 1997; Green and Harry, 1999; Mullen et al., 2017), as well as differences in the spacing between volcanoes, volumetric production, and eruptive style (Guffanti and Weaver, 1988; Sherrod and Smith, 1990; Hildreth, 2007; Till et al., 2019). Based on vent distributions, Green and Harry (1999) suggest that the Cascades arc can be subdivided into seven arc segments, including one with the Canadian Garibaldi Belt and one with Mount Baker and Glacier peak in northern Washington state. The study compares the geochemistry of these proposed segments based on a dataset of 398 olivine-normalized basalt compositions and suggests that the age of the subducting plate is a major driver for the compositional variability. Schmidt et al. (2008) instead suggest that the Cascades arc can be partitioned into four segments based on trace element and isotopic data and the relative abundance of several primitive lava types (i.e. calc-alkaline, tholeiitic, intraplate, and absarokite). The four segments are: The North (Mt. Meager to Glacier Peak), Columbia (Mount Rainier to Mount Jefferson), Central (Three Sisters to Medicine Lake), and the South Segments (Mount Shasta to Lassen Peak). These boundaries were defined such that the geochemical composition of primitive lavas is visually similar within a segment but are relatively distinct between segments. The definition of these segments has been adopted by numerous subsequent authors working on Cascades volcanoes. However, although this work made great strides in understanding the geochemical variability of the arc, it is based on a relatively small data set ($n = 390$) that is strongly spatially biased towards just 14 of the over 3400 Quaternary volcanic vents (Hildreth, 2007) of the Cascades arc (Fig. 1, red columns) and subjectively identified different segments without use of statistical methodology.

To make robust inferences about the causes of large-scale variability within an arc requires that there is adequate geochemical data to fully characterize the compositional range within each region and provide comprehensive spatial coverage of the entire arc. The Cascades is one of the most highly studied arcs in the world, and the tremendous amount of geochemical data that exists for this arc has not been utilized in its entirety by previous studies. This provides a unique opportunity to apply more advanced statistical methods that were developed for larger datasets to address Cascades arc variability in a novel way.

However, while a larger dataset mitigates some problems of incomplete sampling, this sampling is still inherently disproportionate; certain volcanoes within an arc tend to be oversampled, while other less popular edifices remain poorly studied. This is a problem for studies that seek to address large scale questions such as variability over an entire arc, since the oversampled volcanoes overwhelm the overall trends, while compositions erupted at the rest of the volcanoes may not be adequately represented. For example, in the Cascades arc, there are 190 different Quaternary mafic vents between 45.75°N and 46.25°N (the vicinity of Mount St. Helens, Mount Adams, and Indian Heaven); although this is less than 9% of the total vents of the Cascades arc (Hildreth, 2007), data from this region accounts for 26% of all mafic data from Cascades arc literature (~3 samples/vent) (Fig. 1). In contrast, the region between 44.25°N and 45.25°N (includes Mount Jefferson and Three Sisters), which contains 325 Quaternary mafic vents, averages less than a half a sample per vent (Hildreth, 2007). Thus, simple calculations involving a comprehensive arc-wide dataset are highly skewed towards the compositions of over-sampled regions such as Mount St. Helens and Mount Adams, and inferences regarding Cascades arc mafic compositions are therefore biased towards local, rather than regional processes. In this case statistical treatment of these data can help avoid the effects of sampling bias and allow for more robust interpretations of larger, more representative datasets.

For our study, we compile major and trace element and isotopic data on over 13,000 volcanic samples from nearly 250 different publications and use multivariate statistical techniques to more thoroughly assess geochemical variability along the Cascades arc. Our study consists of two parts. We first re-examine the segments proposed by Schmidt

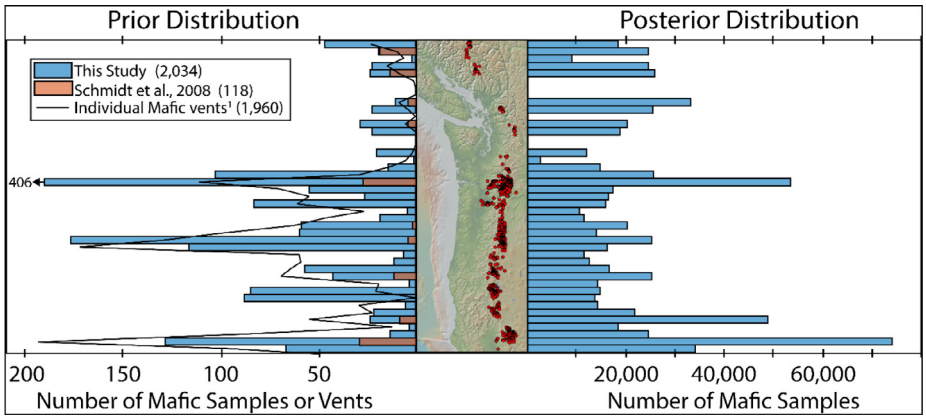

Fig. 1. The histogram on the left shows the distribution of mafic samples used by Schmidt et al., 2008 (red) compared to the mafic arc-front dataset used in this study (blue). Bins are 0.25° latitude. The black line gives the number of Quaternary mafic vents in that 0.25° latitude bin ([1]Hildreth, 2007), plotted on the same axis as number of samples in our data compilation. Locations of arc-front mafic samples used in our study are depicted on the map. The histogram on the right shows the posterior distribution after the Monte Carlo simulation with bootstrap resampling. Note that the posterior distribution is much more evenly distributed than the prior distribution, thereby reducing the inherent sampling bias of the comprehensive dataset. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

et al. (2008) with this more representative dataset, utilizing a Monte Carlo bootstrapping approach to reduce sampling bias. We then use multivariate statistics to quantify differences among the segments, and to test whether they are statistically distinct from each other. In the second part of our study, we develop a multivariate clustering technique to objectively establish a more statistically robust segmentation scheme.

To our knowledge, this study is the first of its kind to use rigorous multivariate statistics to address along-strike geochemical variability within a single arc. By compiling and utilizing a more representative dataset, reducing the influence of oversampled volcanoes, and partitioning the arc into regions that are most geochemically distinct from each other, we can better assess the petrogenetic causes of geochemical variability along the Cascades arc.

2. BACKGROUND

The Cascades arc is well-suited for a study of this kind, due to the wealth of published data and the compositional variability that has been revealed by numerous previous authors (Bacon et al., 1997; Conrey et al., 1997; Green and Harry, 1999; Leeman et al., 2005; Schmidt et al., 2008; Mullen et al., 2017). In addition, the Cascades arc contains over 2300 volcanic vents in a span of approximately 1200 km (Hildreth, 2007), making it particularly well-qualified to assess regional trends rather than those that are localized to widely-spaced stratovolcanoes. Along-arc compositional variability in the erupted volcanic rocks of the Cascades arc may be the result of observed differences in regional tectonics, subduction geometry, or the composition of slab, mantle or overlying lithosphere that occur over the 1200 km arc (Guffanti and Weaver, 1988; Leeman et al., 1990; Hildreth, 2007; Mullen et al., 2017). These along-arc changes are briefly summarized below:

2.1. Subduction geometry

Subduction transitions from nearly orthogonal in the north to highly oblique in the south. The age of the slab that is subducted ranges from 4 to 5 Ma in the northernmost and southernmost regions of the arc, to 10 Ma near the California-Oregon border (Wilson, 2002). Green and Harry (1999) suggest that the age subducted Juan de Fuca plate beneath the Canadian volcanoes is much younger (14–16 Ma) and thus may be up to 200 °C hotter than the age of the slab beneath the High Cascades to the south (16–20 Ma). This may have a significant effect on the dehydration reactions of the subducting slab (Green and Harry, 1999; Green and Sinha, 2005). The slab itself may also be highly segmented, with differing angles, and slab gaps and tears that may allow enriched sub-slab mantle into the overlying wedge (Porritt et al., 2011; Gao and Shen, 2014). Hydration of the Juan de Fuca Plate is highly variable along-strike, as a result of several fracture zones and psuedofaults (Wilson, 2002; McCrory et al., 2012). The entire Gorda Plate in the south is also observed to be highly fractured and the down-going slab has been demonstrated to have concave-down flexure from the trench to 20 km depth, which may affect the release of fluids in the forearc (McCrory et al., 2012). Finally, the depth of the slab beneath the arc volcanoes range from <70 km under Mount St. Helens, to >90 km for Glacier Peak, Lassen Peak, and rear-arc volcanoes (McCrory et al., 2012).

2.2. Subducted slab composition

The Gorda plate, which is subducted beneath the California portion of the arc, has a more depleted MORB composition than that of the Juan de Fuca plate (Davis et al., 2008; Gill et al., 2016). The Gorda plate and the northernmost portion of the Juan de Fuca plate are more intensely

fractured, which may introduce more subduction fluid to the mantle wedge beneath the northern and southern portions of the arc than the central arc (Schmidt et al., 2008). Furthermore, sediment abundance and composition also vary significantly from north to south on both the Juan de Fuca and Gorda plates (Carpentier et al., 2014).

2.3. Mantle heterogeneity

Some authors suggest that the mantle beneath the arc is heterogeneous and contains many isotopically distinct domains, which are sampled throughout the arc (e.g. Bacon et al., 1997). However, recent high precision isotopic data from Mullen et al. (2017) suggest that there are only three isotopically distinct mantle compositions which are sampled by the northernmost Garibaldi segment, the High Cascades, and the rear-arc Simcoe Volcanic Field. In addition, mantle flow patterns may be toroidal around slab edges, and the degree of melting may be highly dependent on position relative to slab edges or gaps (Long, 2016, and references therein).

2.4. Lithospheric heterogeneity and regional tectonics

The Pacific Northwest is a patchwork of accreted terranes and thus the basement beneath the Cascades arc changes significantly along-strike. In the northernmost Garibaldi Belt (~49°N to 51°N), the arc overlies the variably metamorphosed mid-Triassic to mid-Jurassic Bridge River Terrane, which transitions southward into the upper Paleozoic volcanics and schists of the Northern Cascades (~48°N to 49°N) (Righter, 2000). The middle portion of the arc (~44°N to 48°N) is built atop the Columbia Embayment (Trehu et al., 1994), which is assumed to consist of relatively young accreted oceanic terrains; unfortunately the basement is not exposed along the axis except in the Goat Rocks area (Mullen et al., 2017). The forearc is comprised of the 50–55 Ma Siletz Terrane, an accreted oceanic plateau (Phillips et al., 2017), but it is unclear whether the terrain extends beneath the arc (Finn, 1990; Gao et al., 2011; McCrory and Wilson, 2013; Bedrosian and Feucht, 2014). South of ~44°N, the arc overlies the lower Mesozoic Klamath Mountains as well as the Paleozoic island arc Sonomia Terrane (Righter, 2000). Moho depths beneath the arc largely range between 30 and 45 km, with no systematic north-south trend in crustal thickness (Eagar et al., 2011; Gao et al., 2011; Liu et al., 2012). The thickness of lithospheric mantle beneath the Cascades is still highly debated and unconstrained, but most studies suggest only a thin mechanical lithosphere, if it exists at all. Recent studies have suggested the lithospheric mantle is generally less than 10 km thick (Till et al., 2013), and Liu et al. (2012) suggest lithospheric mantle thickness changes rapidly from less than 5 km beneath Mount Shasta to approximately 15 km beneath Lassen Peak.

Regional tectonics also change from overall compressional in the north to extensional in the south (Wells and McCaffrey, 2013; Brocher et al., 2017). In addition, a marked increase in crustal heat flow and hydrothermal heat discharge occurs within the arc southward of 45°15′ (Blackwell et al., 1990; Ingebritsen and Mariner, 2010). This increase in heat flow within the southern half of the Cascades is likely the result of an elevated flux of mafic magma (Till et al., 2019).

3. METHODS

3.1. Data compilation, filtering, and categorization

To fully characterize the geochemical heterogeneity of the Quaternary Cascades arc, we carefully compiled a comprehensive dataset from the literature (Supplementary data S1). To do this, we used three online data repositories: EarthChem, GeoRoc, and the U.S. Geological Survey (USGS) National Geochemical database. Data from these databases were carefully concatenated such that no repeat analyses were included. We manually added data from 98 other sources, including many theses and dissertations, that were not found by the aforementioned databases. Only whole rock or glass analyses of Quaternary volcanic rocks were used. The study was limited to the Quaternary since significant changes have occurred in some portions of the arc since the initiation of the current High Cascades arc axis in the late Neogene (e.g. Pitcher et al., 2017). We compiled all analyses for a given sample into one line of data; for each element, priority was given to the most recent data from the most precise analytical method (e.g. decreasing priority methods for major elements are X-ray fluorescence, wet chemistry, electron microprobe, and then atomic emission and absorption spectrometry). We removed analyses that were collected before 1970 (n = 245 samples). We converted all iron data to FeO*, removed all samples with analytical totals <90%, as suggested by WoldeGabriel et al. (2005), and normalized all major element data to totals of 100%.

Because the processes of magma generation and differentiation within the rear-arc may differ from those of the volcanic front (e.g. Pearce and Stern, 2006), we created a separate dataset for the rear-arc volcanic centers (i.e. Simcoe Volcanic field, Newberry Volcano, and Medicine Lake). USGS geological maps were used to constrain the longitudinal boundary for these three rear-arc provinces. In total, our dataset includes 3683 samples from the rear-arc, which accounts 30% of all Cascades samples (Table 1).

In addition, since compositional differences in felsic rocks may indicate the influence of very different processes than along-arc variability in mafic rocks, we separated the dataset into three compositional groups: mafic (<52 wt.% SiO_2), intermediate (52–62.99 wt.%), and felsic (≥63 wt. %). For the samples that did not have SiO_2 data, we used the rock names given by the author. In addition, a separate dataset was created for more primitive samples (MgO > 7wt.% and Mg# >57.5) (ca. Schmidt et al., 2008).

Although a major goal of this study is to fully characterize the compositional range of the Cascades, outliers greatly affect the results of multivariate statistical techniques such as those implemented in this study (Tryon et al., 2011). Thus, we utilized a two-stage Chauvenet's criterion for rejection to remove outlier elemental or isotopic determinations. Unlike more traditional methods, such as removing

Table 1
Number of samples and individual elemental or isotopic determinations in our Cascades arc compositional data compilation. Note that samples classified as primitive also include high-Mg basaltic andesites and are thus included in the counts for mafic or intermediate samples.

	Arc-front	Rear-arc	All data	Determinations
Primitive	1382	785	2167	37,756
Mafic + Primitive	2034	2624	4658	88,309
Intermediate	5177	1024	6201	122,523
Felsic	1727	443	2170	40,678
Total	8938	4091	13,029	251,510

values beyond 2σ from the mean or 1.5 times the interquartile range from the median, the cutoff point for Chauvenet's criterion is also dependent on the number of analyses used (Taylor, 1997). This is important because the effect of a single outlier on larger datasets is much less significant than for a smaller one. Therefore, we may be more confident in keeping unusual points if we have a larger dataset (Taylor, 1997). We first calculated the mean and standard deviation for each element and isotope and applied the following criterion to each: suspect values were removed only if they were more than τ standard deviations away from the mean, where τ is the critical Z-score for which the probability $P(\tau) = 1 - \frac{0.25}{n}$ and n is the number of analyses. This means that for a dataset with 200 analyses, outliers are defined as being more than 3σ from the mean, but for a dataset with only 10 analyses, outliers are those that are 1.96σ from the mean. Since mean values of each element differ widely between compositions, we applied the Chauvenet's criterion separately for each of the eight subsets of data (four compositions—primitive, mafic, intermediate, and felsic—each with and without rear-arc data). We also did two iterations to account for the effect that especially large outliers have on the mean and standard deviation. It should be noted, however, that many elements do not follow a normal distribution, and perhaps some accurate compositions were erroneously removed as outliers. However, this effect will be minimal, as only an average of five determinations per element were flagged as outliers from the mafic dataset.

After the removal of outliers, our data compilation includes over 250,000 elemental and isotopic determinations on over 13,000 different samples. Of these, 36% are mafic samples, 48% are intermediate, and 17% are felsic (Table 1). To focus on processes related to generation of mafic magmas in arcs, this study is based only on the 4658 mafic samples within the data compendium.

3.2. Reducing sampling bias: Weighted bootstrap resampling

To better characterize along-arc variability and to reduce sampling bias associated with well-studied locations we used a Monte Carlo method with a weighted bootstrap resampling. Bootstrap resampling, or bootstrapping, refers to any statistical technique that involves iterative random sampling from a population, *with replacement* after each sample is drawn. Repeating this sampling many times (n > 10,000), referred to as a Monte Carlo technique, increases the likelihood that the mean or median of the sample set accurately represent that of the entire population. This technique is especially powerful in reducing sampling bias when an inverse weighting scheme is used, such that each analysis from under-sampled regions are given a much higher probability of being selected during bootstrapping, compared to those from oversampled regions. Thus, analyses from under-sampled regions are "pulled up by their bootstraps" to create a new "posterior distribution" that is much more uniform than the original "prior distribution" (Fig. 1).

For our study, we first separated the data into bins of 0.25° latitude ("latbins"). Then, we assigned each sample a weight, or probability of selection ($0.05 \leq W \leq 1$), that is inversely proportional to the number of analyses within its latbin. The weight given to each element within each latbin (W_i) was calculated by dividing number of analyses of a given element within that latbin (N_i) into that of the bin that contains the least samples for a given segment (N_{min}), $\left(W_i = \frac{N_{min}}{N_i}\right)$. This was done separately for each compositional group (e.g. mafic and primitive). Since samples often did not contain the full suite of major and trace elements, the final weight given to all samples within a latbin (W_{latbin}) was calculated by taking the median of the weights of all elements within that latbin: $W_{latbin} = median\left(\frac{\min(N_{SiO2})}{N_{SiO2}}, \frac{\min(N_{TiO2})}{N_{TiO2}}, \cdots, \frac{\min(N_{Yb})}{N_{Yb}}\right)$. We used a minimum weight of 0.05, such that the locations that are most highly sampled (i.e. Mount St. Helens) are not completely disregarded (Keller and Schoene, 2012). Thus, the improvement in the posterior distribution was limited to 20 times (Fig. 1). This weighting scheme is effective at reducing sampling bias between latbins but relies on the full compositional variability within each latbin to be represented in published literature. It should be noted that if a latbin is poorly sampled, such that certain compositions are not sampled or published, our bootstrap resampling methodology would act to amplify the effect of this bias against the unsampled compositions.

Bootstrap sampling was implemented via a MATLAB code (Supplementary Material S2.1). Our procedure was as follows:

(1) For each bootstrap re-sampling iteration, a random number, r, between 0 and 1 is assigned to each sample in the dataset. MATLAB uses a Mersenne Twister pseudorandom number generator, which has been

shown to be sufficiently close to true randomness (Matsumoto and Nishimura, 1998). New r values are generated for each iteration.
(2) Each sample for which W > r, was "chosen" for that round.
(3) For each element of the chosen sample, a random value is drawn from a Gaussian distribution formed by μ = reported value, σ = analytical error, and all these values for a chosen sample are added to the bootstrap subset.
(4) Step (3) is completed for each chosen sample (W > r) of that bootstrap iteration.
(5) The bootstrap subset from this iteration is concatenated to the subsets from all previous iterations into a single Monte Carlo iteration set. This simulates random sampling with replacement—a fundamental aspect of the bootstrap method.
(6) Steps 1–5 are repeated until the number of samples within the re-sampled dataset is equal to the original number of samples.
(7) For each element and isotope, the mean, standard deviation, and median of this iteration's resampled set is calculated and stored in an array.
(8) Steps 1–7 are repeated for 10,000 iterations.
(9) The bootstrapped mean and median, as well as the 95% confidence intervals for both statistics are calculated directly from the distribution of these 100,000 values.

Step 3 is necessary because the bootstrap subset must have a continuous distribution. We used an analytical uncertainty of 2%, as suggested by Keller and Schoene (2012). To test the effect of this choice of analytical uncertainty, we did two additional test runs, one in which all elements had a 4% uncertainty, and one where major elements had 2% uncertainty and trace elements had 4%. We found that this made almost no difference in the final bootstrapped averages of all elements, and the effect on the bootstrapped confidence intervals was small enough that it could not be visually observed on most bivariate plots.

3.3. Calculating confidence intervals for bootstrapped means

To best approximate the unbiased mean composition of each segment, we calculate the overall mean of the average values from each of the 10,000 Monte Carlo iterations, as in Keller and Schoene (2012). We also calculate 95% confidence intervals around the mean (Supplementary Material S4). Although the Central Limit Theorem states that with a large sample size, the distribution of samples means will approach a Gaussian distribution, we chose to avoid the assumption of normality when calculating these confidence intervals since data is sparse for some elements and isotopes, and others have multimodal distributions. Instead we calculate the 2.5th and 97.5th percentiles of the distribution of bootstrapped means (N = 10,000) to derive the 95% confidence intervals of the bootstrapped mean directly (Carpenter and Bithell, 2000; Haukoos and Lewis, 2005; Pek et al., 2017). Thus, confidence intervals tend to be slightly asymmetric for our bootstrapped means. In addition, we calculate the bootstrapped median for each element and isotope (Supplementary Material S4). However, we choose to focus on the bootstrapped means, as these represent the bulk composition that would be produced if one were to physically mix all mafic samples from the literature, but in proportions such that each 0.25° latitude had approximately the same number of samples.

3.4. Testing the robustness of segmentation schemes

Segmenting the arc should be done in a way that the segments are most geochemically dissimilar to one another, so that the compositions can be compared, and inferences can be made as to the causes for the differences. If two segments are relatively similar, then they should be combined. Thus, to test whether segments are geochemically distinct, we use a multivariate technique called the Hotelling's T^2 test. This test is the multivariate equivalent of the Student's t test and is a "post-hoc test" that follows a multivariate analysis of variance (MANOVA). Hotelling's T^2 test essentially evaluates whether the multivariate means ("mean vector") of two segments, are different enough in n-dimensional space to reject the null hypothesis that the two are the same. Before testing this, the variables (elements) are transformed by the MANOVA technique, such that the new transformed variables minimize the within-segment variation and maximize the between-segment variation. Thus, instead of just simply comparing the multivariate distances between mean values of each segment, the Hotelling's T^2 compares the means of each element, variance of each element, and multivariate covariance between elements (i.e. multivariate trends) of the data in each segment.

For large populations (N > 30), such as the dataset in our study, the Hotelling's T^2 statistic follows a X^2 distribution with k degrees of freedom, where k is the number of geochemical elements, allowing us to calculate a p value. Small p values indicate that two segments are more different than would be expected for two samples drawn from a single population. Specifically, if the $p < 0.05$, then we reject the null hypothesis, and the two segments are considered statistically distinct. The Mahalanobis distance, which is closely related to the Hotelling's T^2, provides the multivariate distance between the mean vectors of the two groups. For each segment, we utilized a bootstrapped resample set of 10,000 (using steps 1–6 above) to complete the statistical test. We used the Hotelling's T^2 test on both the segmentation scheme of Schmidt et al. (2008) and our new statistically-based scheme and compared the Mahalanobis distances between segments to assess which created more distinct groups.

3.5. Establishing new segments: modified Hierarchical clustering mechanism

To determine a quantitative segmentation scheme that best separates the Cascades arc into geochemically distinct regions, we developed a new, modified hierarchical clustering technique. For each step in a classic hierarchical clustering analysis, the two individuals or groups that are most similar (e.g. have the shortest distance between their mean

values) are combined into a new group and this process is repeated until the desired number of groups are reached or all individuals are combined. Thus, the process clusters data such that the within-group similarity is maximized. This process can be represented by a dendrogram, in which all individuals start out as separate branches on the y axis, and as two individuals are merged together, the distance between their means (distance of fusion) are shown on the x axis (e.g. Fig. 4). We had to modify this traditional hierarchical clustering technique to reduce sampling bias, deal with problems associated with missing data, ensure that clusters consisted of regions that were geographically proximal to one another, and account for correlation between elements (i.e. differentiation trends). These modifications will be discussed in further detail below, and our MATLAB code is given in Supplementary Material S2.2.

To account for the significant covariation expected between many elements in geochemical data (e.g. correlation with SiO_2) we chose to use the Mahalanobis distance (MD) as the measure of dissimilarity between groups in our hierarchical clustering method. In contrast to the more common centroid linkage approach, which uses Euclidian distance to measure the distance between clusters and assumes a spherical distribution for all elements around these centroids, MD measures the distance between the centroids of two groups while also considering the dispersion around the centroids. This dispersion is measured by a dataset's covariance matrix. The MD between groups i and j is given by: $MD_{ij} = \sqrt{(\bar{x}_i - \bar{x}_j)^T S^{-1} (\bar{x}_i - \bar{x}_j)}$ where S is the covariance matrix and $\bar{x}_i$ is the mean vector for sample i that contains the means of each element. To calculate the covariance matrix, we used the Monte Carlo bootstrapped dataset for each latbin. Because most samples had data for only a subset of elements, missing data was common. The covariance matrix of a dataset can become singular if it has excessive missing values. Thus, in order to calculate the covariance matrix for our dataset, we used only a subset of well-represented elements (n = 15 or 24, see below) and utilized the Expectation Conditional Maximization (ECM) algorithm of Meng and Rubin (1993) to calculate a non-singular covariance matrix. This algorithm imputes values for these missing data, based on the mean and variance of that element, and then iteratively changes these imputed values until the log-likelihood function is maximized. Further explanation of the theory can be found in Little and Rubin (2014).

We wanted to reduce the effects of sampling bias when creating our new segments. Thus, every time two groups of latbins were combined, we performed a new Monte Carlo analysis with bootstrap re-sampling to find that new group's bootstrapped mean elemental concentrations. This portion of our procedure was almost identical to the one described above in Section 3.2.

To create the clusters (segments), we used mafic data (<52 wt.% SiO_2) because primitive data (SiO2 < 52 wt.%, MgO > 6.5 wt.%, Mg# > 55) are relatively sparse in some parts of the arc. We also chose to exclude rear-arc data, because this allows us to focus on the processes that may be responsible for along-strike variability, without adding complications from the additional processes involved in rear-arc magma generation.

We completed the clustering process in two stages. The initial stage was used to combine the seven latbins with little to no data for some elements (n < 4) with the neighboring latbin that was more similar. Since Mahalanobis distances are highly skewed by data with few observations, for this initial stage we used only the 15 elements for which all latbins had at least four analyses (SiO_2, TiO_2, Al_2O_3, FeO, CaO, MgO, K_2O, Na_2O, Cr, Rb, Sr, Y, Zr, Ba, Ce). After merging those sparsely-sampled latbins with a neighbor, there were 24 elements for which there were at least four analyses in all latbins of data (Addition of MnO, P_2O_5, Zn, Nb, La, Tb, Yb, Hf, Th). Thus, we completed the second stage of clustering using these 24 elements. We wrote a MATLAB script for our modified hierarchical clustering technique, which is summarized in Fig. 2.

To objectively determine the step at which to end the clustering, we use the upper-tail stopping rule of Mojena (1977). This rule states that the clustering should be cut when, for the first time, $\alpha_{j+1} > \bar{\alpha} + k(s_\alpha)$, where $\bar{\alpha}$ and s_α are the mean and standard deviation of all the j previous distances of fusion. The k value is the upper-tail critical value for a t distribution that is defined by the number of original separate groups, and a choice in confidence level. Since we start with 25 separate latbins in the second round of clustering (after combining the ones with little to no data), the critical t value at the 97.5% confidence level is 2.064. In other words, we stopped the hierarchical clustering once the "closest" segments are separated by a MD that is ($2.064*s_\alpha$) higher than the average of all previous steps' distance of fusion.

4. RESULTS

4.1. Statistical test of the previous segmentation scheme

Our Hotelling's T^2 test results indicate that all four of the segments defined by Schmidt et al. (2008) are statistically distinct for mafic and primitive compositions (Table 2). For all neighboring segments, $p \ll 0.05$, indicating that the samples within two segments differ too much to have come from the same population. Bootstrapped MD values range from 2.5 to 3.1 for mafic compositions and 2.3 to 4.4 for primitive compositions. These MDs indicate that for mafic and primitive compositions, the Central and South Segments are less distinct than the other pairs (MD = 2.5 and 2.31, respectively). The North and Columbia Segments are most dissimilar with MD values of 3.1 and 4.4 for mafic and primitive compositions, respectively.

4.2. The new statistically-derived segmentation scheme

Based on the upper tail stopping rule of Mojena (1977), we terminate the second stage of clustering after 12 steps (Fig. 3), at which point the Mahalanobis distance of fusion (4.9) exceeded the critical value (4.8). At this step, we have seven clusters of data, with MD values that range between 3.3 and 15.9. Cutting the clustering process at this step leaves one solitary latbin without a cluster, the region that

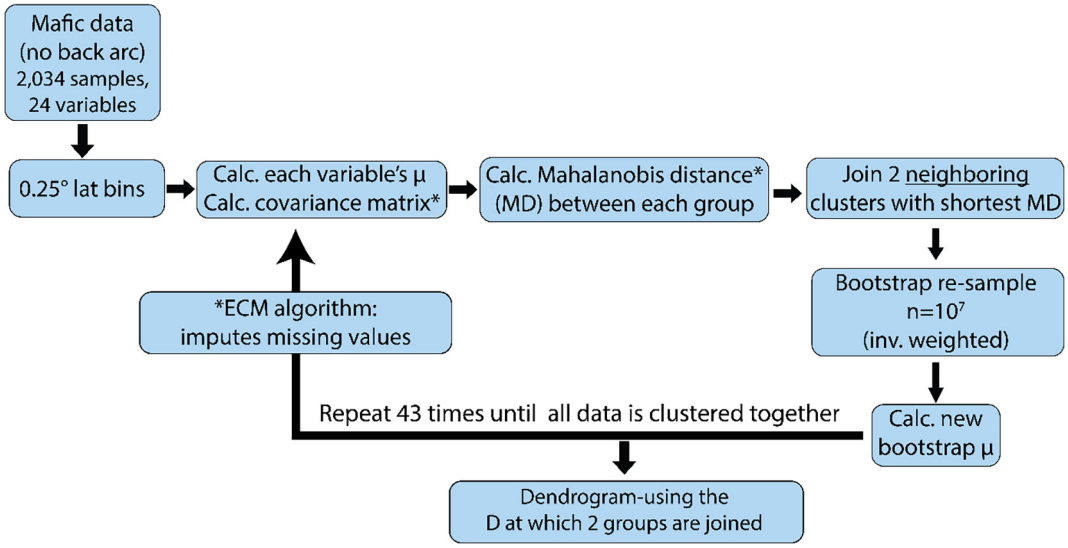

Fig. 2. Flowchart of the new modified hierarchical clustering methodology developed in this study. Note that this methodology was completed in two stages, one with 15 elements to combine the sparsely sampled latbins, and the second with 24 elements.

includes the Columbia River (45.5–45.75°N). To avoid an unusual segment that spans only a quarter degree latitude and includes only 25 samples, we chose to add it to the cluster to the south, as it produces larger MD between the two neighboring segments (MD = 3.6) than if we had added it to the north (MD = 3.1).

Using our modified hierarchical clustering technique, we establish a new statistically-based segmentation scheme for the Cascades arc (Fig. 4), with six segments, hereafter referred to as the Garibaldi (49.75–51°N), Baker (48.5–49.75°N), Glacier Peak (47.75–48.5°N), Washington (45.75–47.75°N), Graben (44.25–45.75°N), and South (41.25–44.25°N) Segments. Hotelling's T^2 tests indicate that for all segment pairs, $p \ll 0.05$, and thus, these new segments are all statistically distinct from each other (Table 2). MD values for mafic compositions range between 3.6 and 16.0, with the largest difference between the Baker and Glacier Peak Segments (MD = 16.0) and Garibaldi and Baker Segments (MD = 10.0). The smallest distinction is between Washington and Graben Segments, with a MD of 3.6. Our new scheme has inter-segment MD that are up to 6.3 times greater than that of Schmidt et al., 2008. This will be discussed further in Section 5.1.

Although using a different subset of elements tends to change the segmentation boundaries in detail, the broad scale features of segmentation remain overall similar, regardless of the elements used. Furthermore, we find that regardless of the subset of elements used, our new scheme is statistically robust.

4.3. Within-segment compositional heterogeneity

Although our methodology acts to maximize the multivariate difference between each segment, while minimizing the within-segment variability, it should be noted that significant intra-segment compositional variability inherently persists. Using MATLAB, we employed a principal component analysis (PCA) of our bootstrapped dataset to assess, in multivariate space, the degree of compositional variability within each segment. This PCA transforms our data onto three orthogonal axes that are the linear combinations of the original major and trace elements that represent the greatest variability in the dataset. This has the advantage that we can determine which segments are most heterogeneous in n-dimensional space and also allows us to assess which elements or oxides vary the most within the Cascades arc. Prior to the PCA, we log-transformed the bootstrapped mafic arc-front data, centered it using Z-scores, and used an alternating least-squares technique during PCA to avoid issues with missing values. Full PCA results are given in Supplementary Material S3.

The first three principal components account for 70% of the variance of all Cascades arc mafic data (Fig. 5). The first principal component axis (PC1) is primarily the result of variability in MREE, LREE, and LILE (particularly Ba, Sr, and Rb). This means that the largest degree of mafic compositional heterogeneity is in these elements. PC2 mostly represents variability in HREE (Yb), HFSE (Nb), and K_2O, and PC3 primarily gives the variance of MnO, TiO2, Zr, La, Rb, and Y. In the first two principal components, the Graben Segment has the highest variance, whereas in PC3 the South Segment has the highest variance (Fig. 5). Glacier Peak and Baker Segments tend to have the lowest variance for most principal components. Our PCA corroborates compositional variability within each segment and indicates that this compositional heterogeneity is particularly high in the southern segments of the arc. While these southern segments do have a larger number of samples, there is not a direct correlation between the number of samples and the principal component variance within a segment.

Table 2
Hotelling's T^2 test results for neighboring segments of our new segmentations scheme (left) compared to those of Schmidt et al., 2008. We used the same data for each scheme and used the same Monte Carlo bootstrap methods. Number of samples (n) in each segment is given. All neighboring segments have $p \ll 0.01$, indicating that both schemes create statistically distinct segments. However, T^2 values, F-values, and Mahalanobis distances (MD) between segments are much higher in the new scheme indicating that it creates more statistically distinct segments.

	New segments						Schmidt et al., 2008		
	Garibaldi vs. Baker	Baker vs. GP	GP vs. WA	WA vs. Graben	Graben vs. South	North vs. Columbia	Columbia vs. Central	Central vs. South
n	157	98	686	887	1191	1035	1565	999
MD	10.0	16.0	7.3	3.6	4.5	3.1	2.6	2.5
T^2	3142	6229	2671	2392	4124	1651	2678	1244
T^2 critical	40	40	40	40	40	40	40	40
F	97	169	95	86	149	60	98	45
F critical	1.6	1.6	1.5	1.5	1.5	1.5	1.5	1.5
p-value	≪0.01	≪0.01	≪0.01	≪0.01	≪0.01	≪0.01	≪0.01	≪0.01

5. DISCUSSION

5.1. Comparison of segmentation schemes

The primary objective of partitioning the arc into segments is to explore the causes for the geochemical differences between them. Thus, it is most useful if the partitioning is done such that the geochemical differences between segments is maximized. Green and Harry (1999) proposed partitioning the arc into seven segments, based entirely on vent density and spacing. Schmidt et al. (2008) used a small dataset of primitive lavas (n = 390 samples) to subjectively segment the arc into four regions such that the compositions of these lavas are visually similar within a segment but are relatively distinct between segments. We demonstrate using our new expanded dataset (n = 2034) that the segments defined by both Green and Harry (1999) and Schmidt et al. (2008) are statistically different from each other, however, our new statistically-derived segments are up to 5.7 times more geochemically distinct than the former study and up to 6.3 times more distinct than the latter (Fig. 6). Thus, our new objective scheme better maximizes the differences between neighboring segments, and we believe it is better suited for partitioning the arc for compositional comparisons.

The new segmentation also provides new insight into longstanding issues related to Cascades arc volcanism and magma sources. For example, the marked change in the strike of the arc near Mount Baker and Glacier Peak is typically used as a basis to subdivide the arc into the Garibaldi Volcanic Belt to the north and the High Cascades to the south (Mullen and Weis, 2015). Recent high precision Sr, Nd, Pb, and Hf isotopic work finds significant differences between the Garibaldi and Cascades arcs and suggests that they are fed by isotopically distinct mantle sources and should thus be considered separate arc segments (Mullen et al., 2017). Our study, which utilizes a much larger major and trace element dataset confirms this distinction, showing statistically significant differences between the Garibaldi volcanic belt (north of the US-Canada border) and the High Cascades to the south. While multiple studies consider Glacier Peak to be part of the Garibaldi belt (Hildreth, 2007; Schmidt et al., 2008; Mullen and Weis, 2015), Mullen et al. (2017) choose to reclassify it to be part of the High Cascades based on isotopic data. However, our results indicate that Glacier Peak and Mount Baker are compositionally distinct from both the Garibaldi Segment and the Washington Segment of the High Cascades to the south, as well as from each other (Fig. 6). In fact, these are the most statistically distinct portions of the arc (MD = 16 and MD = 10). This distinction is important as it implies that these two transition volcanoes result from different processes or source conditions than those to both the north and the south; this will be explored further in the following section.

In addition to considering Mount Baker and Glacier Peak to be disparate segments, our scheme differs from that of Schmidt et al. (2008) by several other key features. Our results indicate that volcanic rocks north of the Columbia river (Washington Segment) and those south of the river

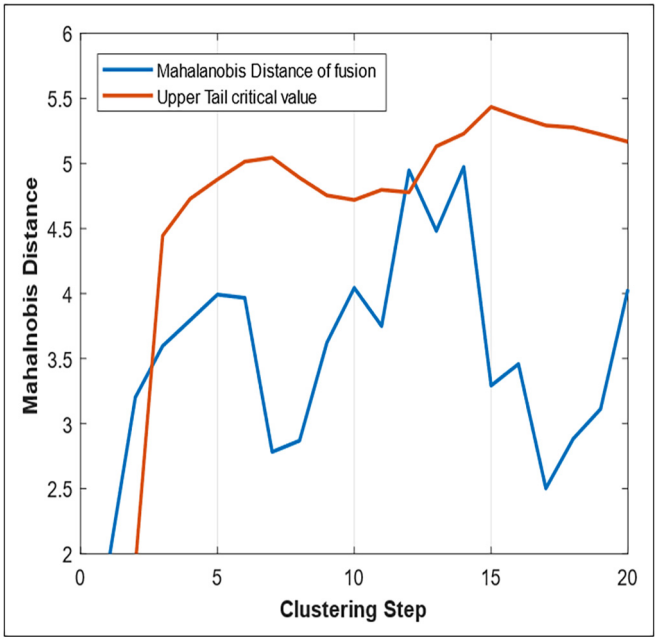

Fig. 3. Methodology used to determine the step at which to end the hierarchical clustering. The Upper-tail stopping rule of Mojena (1977) states that we should end the clustering after the 12th step, as that is the first time that the MD of fusion (blue line) is greater than the critical value (red line). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

are statistically distinct from one another (MD = 3.6), despite having been grouped together within the Columbia Segment of the previous study (Fig. 4). Furthermore, because we used a comprehensive dataset that includes data from the many smaller mafic vents between the major stratovolcanoes, we can better define the boundaries between all segments. For example, while mafic rocks in the vicinity of Mount Rainier were not included within any segment by the previous study, our results indicate that they are statistically more similar to those to the south and can therefore be considered part of the Washington Segment. Finally, although Schmidt et al. (2008) drew a boundary near the Oregon-California border based on data from rear-arc Medicine Lake, there was no data from the over 70 vents (Hildreth, 2007) located within the 140 km region between Crater Lake and Medicine Lake. Our study, which utilizes nearly 200 samples from that region, finds that they are statistically similar to those that surround Crater Lake. In fact, we find that bootstrapped mean compositions of all latbins from Lassen Peak to the Three Sisters Region are similar enough to warrant placing them within the same segment (Fig. 4).

We also note that whereas Schmidt et al. (2008) used data from rear-arc volcanoes to partition the arc, we chose to exclude these data. This continues a longstanding pattern in the Cascades where many authors consider Simcoe, Newberry, and Medicine Lake to be rear-arc or back-arc volcanoes that should be distinguished from those the arc-front (Hildreth, 2007; Donnelly-Nolan et al., 2008; Long et al., 2012), but others treat these as part of the arc-front (Guffanti and Weaver, 1988; Blakely et al., 1997; Schmidt et al., 2008; Mullen et al., 2017). Our Hotelling's T^2 tests indicate that all three are statistically dissimilar to the adjacent arc-front segments that lie to the west (Table 2), with bootstrapped MD values of 4.2, 3.6, and 4.0 for Simcoe, Newberry, and Medicine Lake, respectively (Fig. 6). Thus, we suggest that future regional studies treat these as rear-arc volcanic centers that have significantly different compositions, and thus processes of magma generation, compared to the adjacent arc-front volcanoes.

5.2. Distribution of primitive Cascades arc endmembers

Our new segmentation scheme indicates that there are distinct along-strike changes in the mean compositions of Cascades arc mafic lavas. These data indicate that the Cascades arc does not exhibit simple north-south trends in composition. Instead, the arc is segmented into irregular intervals, indicating that geochemical variability is controlled by more regional differences in the subduction system. In the following two sections we first demonstrate intersegment differences in proportions of primitive endmembers and then discuss potential causes for the geochemically complex and segmented nature of the Cascades arc.

Several end-member primitive basalt compositions have been proposed for the Cascades including arc-typical calc-alkaline basalts (CABs), low-K tholeiites (LKTs) (sometimes referred to as high alumina olivine tholeiite; e.g. Bacon et al., 1997; Hart et al., 1984), intraplate-type basalts (IPBs) (sometimes called ocean island, within-plate, or HFSE-type basalts; e.g. Conrey et al., 1997; Schmidt et al., 2008), absarokites/shoshonites (SHO), and high-Mg basaltic andesites (HMBA). A detailed description of each of these Cascades arc end-members is given in Mullen et al. (2017). The differing major and trace element compositions of these endmembers are suggested to be the result

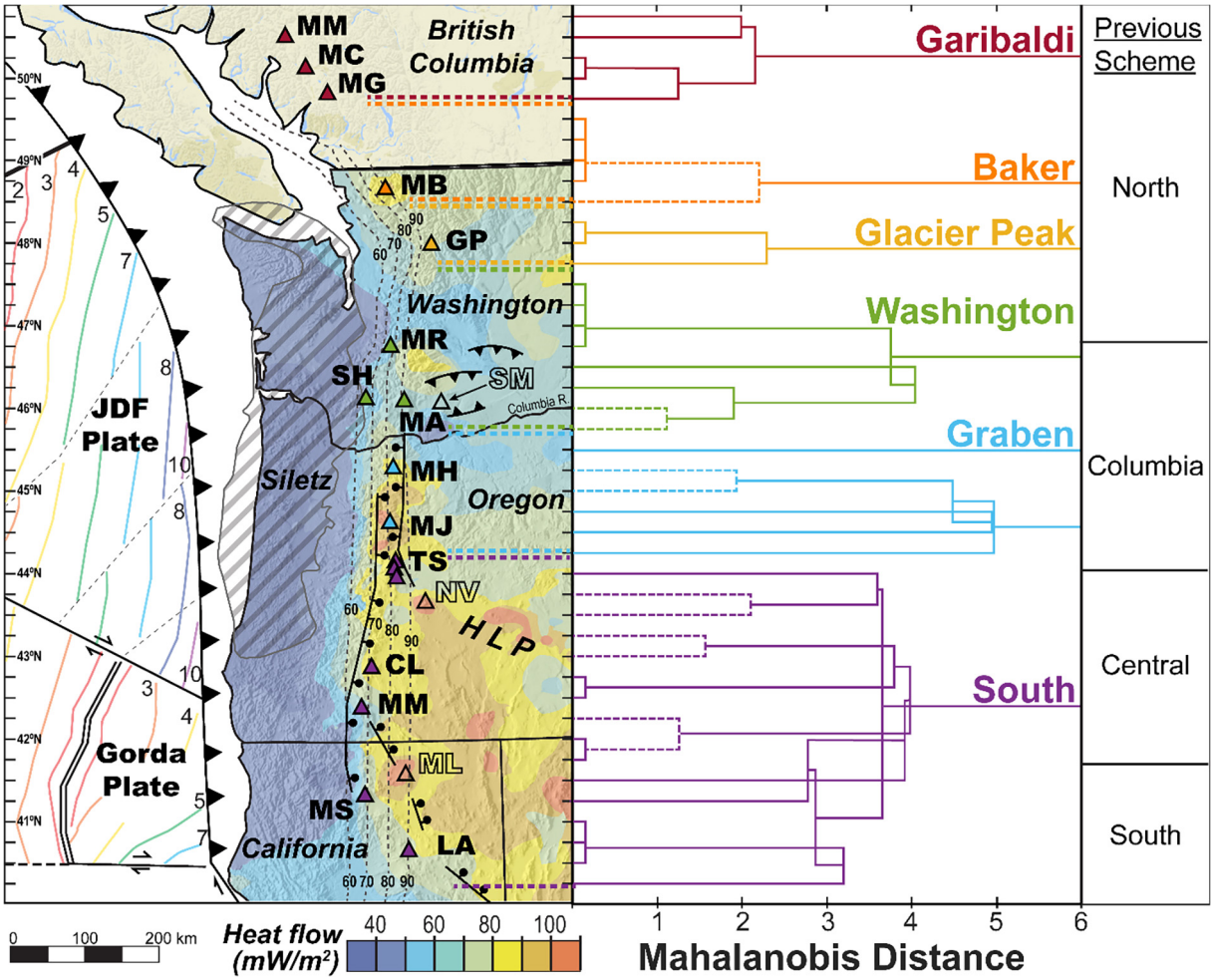

Fig. 4. Dendrogram **(right)** showing the MD at which each latbin was clustered during the modified hierarchical clustering. Latbins that were combined in the first round of clustering (15 elements) are designated by dashed lines. In some cases, after a new cluster was formed at a higher MD, the resulting cluster became more similar to another and was then combined at a lower MD than the prior step. This leads to crossing dendrogram branches. Final segments are designated by color and can be compared to the previous scheme of Schmidt et al. (2008) on the far right. The map **(left)** summarizes various along-strike differences that may lead to compositional variability of the Cascades arc. Approximate age of Juan de Fuca (JDF) and Gorda plates (colored lines) are from Wilson (2002). Estimated extent of the Siletz Terrain is from Phillips et al. (2017). Heat flow data is from Ingebritsen and Mariner (2010). Depth of the subducting slab is estimated by McCrory et al. (2012). General locations of major faults are from Schmidt et al. (2008). General location of the High Lava Plains (HLP) is also shown. Major stratovolcanoes (triangles), are colored by the new segmentation scheme. From north to south: Mount Meager (MM), Mount Cayley (MC), Mount Garibaldi (MG), Mount Rainier (MR), Mount St. Helens (SH), Mount Adams (MA), Mount Hood (MH), Mount Jefferson (MJ), Three Sisters (TS), Crater Lake/Mazama (CL), Mount McLoughlin (MM), Mount Shasta (MS), and Lassen Peak (LA). Rear-arc volcanoes (empty triangles), including Simcoe volcanic field (SM), Newberry (NV), and Medicine Lake (ML), which were not used to create the new scheme. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

of different mantle processes or compositions (Leeman et al., 1990; Bacon et al., 1997; Conrey et al., 1997; Leeman et al., 2005; Schmidt et al., 2008; Rowe et al., 2009; Moore and DeBari, 2012; Mullen et al., 2017; Carlson et al., 2018). Although these studies have demonstrated that most endmember compositions exist within each portion of the arc, no previous study has ascertained the relative proportion of these compositions using a comprehensive dataset that is statistically treated to mitigate sampling bias.

We classified 1571 of the 2167 primitive samples (including HMBA and SHO) in our dataset based largely on the compositional definitions (Table 3) given in Schmidt et al. (2008) and Leeman et al. (2005). Many samples were mixtures of the endmembers and could not be definitively classified as a single endmember type. One common mixture were basalts that had LKT-like low K_2O (<0.4 wt.%) and K_2O/TiO_2 (<0.4) but were enriched in subduction fluid-mobile elements such as Ba and Sr similar to CABs. Due to their abundance, we chose to create a separate category of "High-Ba LKTs" when calculating proportions. Supplementary Table S6 contains the compositions of the primitive basalts that we could uniquely classify, as well as the mean, median and standard deviation of each type within

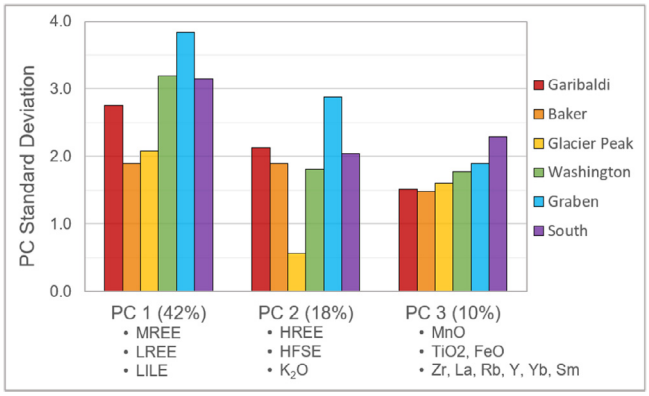

Fig. 5. Standard deviations of the first three principal component scores of each segment. Percentage of total variance of the log-transformed mafic data that is explained by each component is given in parentheses. Elements and oxides that contributed most to each principal component are listed.

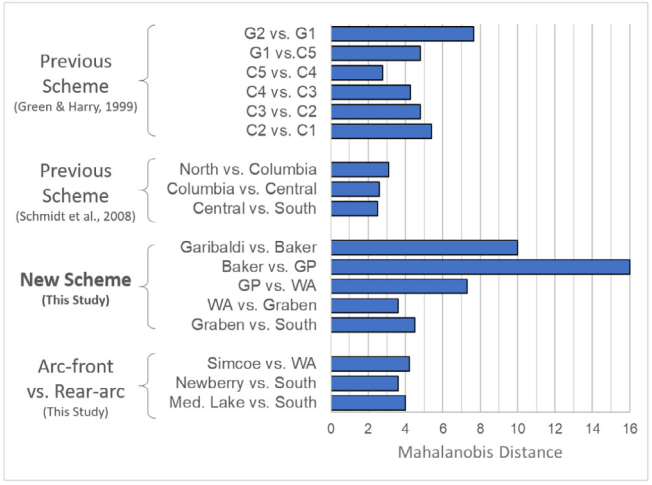

Fig. 6. Comparison of Mahalanobis distances (MD) between neighboring segments from two previous studies, using our new data compilation for each. Our unbiased objective methodology created segments that are much more statistically distinct from each other (higher MD) than those of Schmidt et al. (2008), and more distinct than most segments created by Green and Harry (1999). High MD between the rear-arc volcanic fields and adjacent front-arc segments indicate that they are statistically dissimilar from the arc-front, and thus, should not be considered part of the Cascades arc sensu stricto.

each segment. Note that although we include discussions of the relative proportions of HMBA and SHO primitive lavas in this section, we did not include these compositions in our calculations of bootstrapped means (Section 5.3), as many elemental determinations were flagged as outliers (e.g. Sr, Ba, Yb) particularly for SHO, and therefore highly skewed the mean compositions. We include bootstrapped means that include these compositions in Supplement S5.

Utilizing the Monto Carlo bootstrap methodology to minimize sampling bias, we observe significant inter-segment differences in the distribution of primitive lava types (Fig. 7). The Garibaldi Segment consists of primarily (67%) of IPB-type basalts, while these types are completely absent in the Baker and Glacier Peak Segments. In these latter two, pure LKT-types are also absent, and instead, High-Mg basaltic andesites, Ba-enriched LKTs, and more arc-typical CAB-types dominate. The lack of IPB-type basalts in these two segments is consistent with the southward flow and progressive melting of enriched mantle that becomes too depleted by the time it reaches the Baker segment to allow for the formation of IPB-type basalts (Mullen and Weis, 2015). Decompression melting and the generation of LKT basalts may also be stifled by the compressional tectonics of the region (Wells and McCaffrey, 2013; Brocher et al., 2017). We suggest that even if IPB or pure LKT mantle melts are produced beneath these two segments, overprinting of these compositions by fluid flux signature may be the reason for the paucity of IPB or true LKT endmember final compositions erupted (Schmidt et al., 2008; Rowe et al., 2009; Moore and DeBari, 2012; Mullen et al., 2017). Furthermore, the abundant high-Mg basaltic andesites in these two segments may be primary hydrous mantle melts (Grove et al., 2002; Sas et al., 2017), but may also result from interaction with

Table 3
Criteria used to classify primitive samples into Cascades endmembers. Similar criteria were used by Leeman et al. (2005) and Schmidt et al. (2008). Criteria for absarokite/shoshonite is based on Morrison (1980).

Primitive type	SiO2 (wt.%)	MgO (wt.%)	Mg #	K_2O (wt.%)	K_2O/TiO_2	Ba/Nb	Sr/Y	Nb/Zr	$(K/Nb)_N$
CAB	<52	>6.5	>57	0.5–1.5	0.4–1.0	>20	>15	<0.09	>0.2
High-Ba LKT	<52	>6.5	>57	<0.5	<0.4	>20		<0.09	>0.2
LKT	<52	>6.5	>57	<0.5	<0.4	<20	<20	<0.09	>0.2
IPB	<52	>6.5	>57	>0.25	0.4–1.0	<20		>0.08	<0.25
SHO	<54	>6.5	>57	>1.5	>1.0	>35	>50	<0.09	>0.30
HMBA	<57	>7.5	>60	<1.5	<1.0				

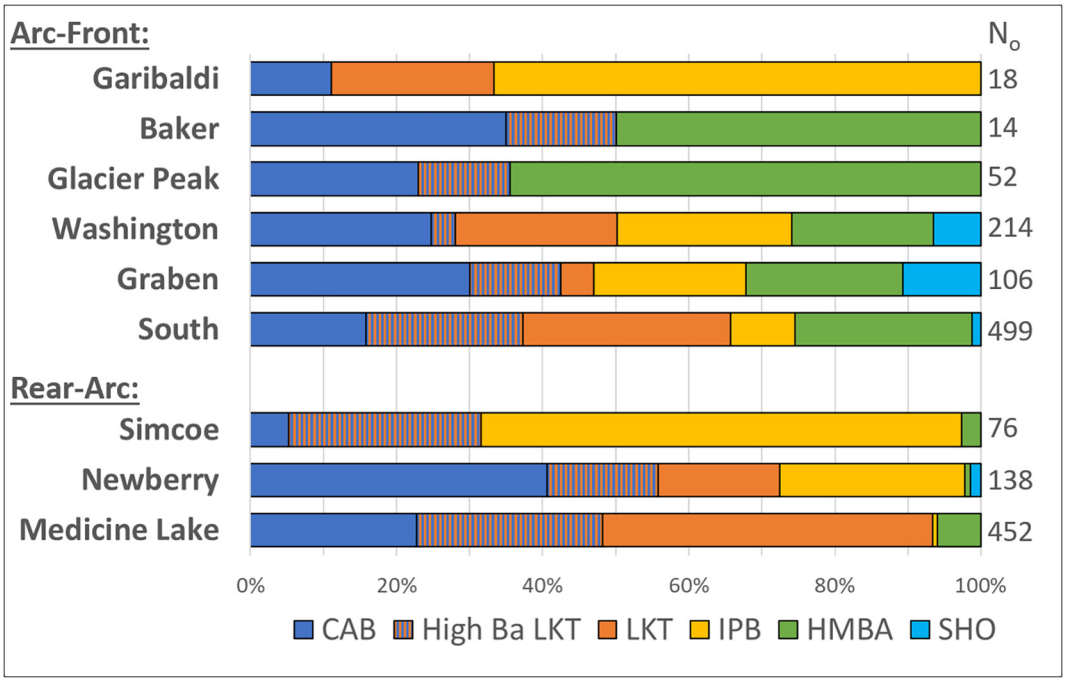

Fig. 7. Posterior distribution of primitive lava types for the new arc-front segments and rear-arc fields. Abbreviations are as given in the text and classification criteria is given in Table 3. The total number of samples in the prior dataset (No), are shown on the right.

slab-derived melts (Moore and DeBari, 2012; Sas et al., 2017), or magma mixing with dacites and assimilation of ultramafic xenoliths (Streck and Leeman, 2018).

Although true LKT and IPB basalts are entirely absent in the northern portion of the High Cascades, both types become quite common in the southern half of the arc, which is undergoing overall extension (Wells et al., 1998; McCaffrey et al., 2007). This extension, which is also associated with higher amounts of mantle derived magma in the crust, higher heat flow (Fig. 4), and lower crustal seismic velocities, may also allow for a higher flux and eruption rate of decompression-driven LKT-type melts (Schmidt et al., 2008; Ingebritsen and Mariner, 2010; Mullen et al., 2017; Till et al., 2019). Absarokite/shoshonite (SHO) primitive compositions are absent in the northern three segments but are present in small proportions in the southern half of the arc, particularly in the Graben Segment, where they represent 14% of primitive lavas. The SHO endmember is characterized primarily by high K_2O, Ba, Sr, and La/Yb, which may be produced by deep low-degree hydrous partial melting of metasomatized (perhaps phlogopite-bearing) and refractory mantle (Morrison, 1980; Conrey et al., 1997; Rowe et al., 2009). Perhaps extension in the southern portion of the Cascades allows for melts that are variably-enriched and subduction-modified from throughout the mantle wedge, thereby creating more diversity in primitive compositions (Figs. 4 and 7). We discuss these implications further in the following sections.

While some authors suggest that the mantle beneath the Cascades may be inherently heterogenous (e.g. Bacon et al., 1997), Mullen et al. (2017) find that there are no isotopic differences between these endmembers and suggest an isotopically homogeneous mantle source for all endmember types, except for IPB basalts in the Garibaldi Segment and in the Mount Adams and Simcoe rear-arc region (Leeman et al., 1990; Mullen et al., 2017). Several studies demonstrate that Cascades CABs and LKTs can both be produced from the same initial composition, simply by varying the flux of fluids and degree of melting (Reiners et al., 2000; Rowe et al., 2009; Moore and DeBari, 2012). Thus, compositional differences between Cascades arc endmembers are likely to be primarily the result of differing

processes of mantle melting and variable degrees of modification by the subduction components but can be, in some specific locations, influenced by enriched mantle that may be isotopically distinct (Reiners et al., 2000; Mullen et al., 2017). However, to explore these potential causes and their relative importance throughout the arc, we must examine the compositions of all mafic/primitive magmas erupted in the Cascades arc, not just the subset of endmember compositions.

5.3. Compositional differences between segments

In addition to comparing the relative proportions of these endmembers, our methodology allows us to compare the bootstrapped mean compositions (Supplemental data S4), which can be thought of as a representation of the unbiased bulk mixture of various primitive lava types, and thus, representative of the relative contributions of the processes that lead to their formation. Although previ-

Table 4
Summary of interpretations of bootstrapped mean compositions of the new Cascades arc segments Although different basalt types in each segment are the result of different depth, degree, and type of melting, the mean composition suggests the most prevalent processes. See text for further discussion.

	Flux melting	Melting depth, degree	Mantle fertility	Interpretation
Garibaldi	**Lowest:** Lowest Ba/Nb, Li, Sr/P **Slab melt?** High Sr/Y, Th/U High % adakites (felsic)	**Deep:** Lowest Y/Zr, Yb, Y, Cr, Sc High Dy/Dy* and Dy/Yb **Lower F:** Steep REE, High K_2O, % IPB	**Highly Enriched:** Highest Nb/Yb, Nb/Zr Highest % IPB	– Young, hot slab dehydrates prior to arc – Hot slab also causes slab melting – Toroidal flow of enriched sub-slab mantle around northern slab edge
Baker	**Moderately high:** High Ba/Ce, Ba/Nb, Pb/Ce Low Nb-Ta, HFSE High % CAB; No LKT, IPB	**Shallow:** Highest Y, Yb, Y/Zr Lowest Sr/Y Lacks IPBs	**Highly Depleted:** More depleted than JDF MORB Lowest Nb/Zr, no IPB	– Older, colder, and slower (more oblique) slab delivers higher fluid-flux – Stagnant depleted mantle – Distinct from Garibaldi segment
Glacier Peak	**Moderately high:** Mod. Ba/Ce, Ba/Nb Lowest Nb, Ta, HFSE	**Deep Fluid:** Highest Sr/P, Sr/Y, Th/Nb, U High Pb/Ce (not sed melt) **Mod. shallow melt:** Mod. Y, Yb, Y/Zr	**Highly Depleted:** Low Nb/Zr, no IPBs More depleted than JDF MORB	– Located further from trench = deeper slab (95 km) vs. rest of arc (avg. 80 km) – Deeper, hotter supercritical fluids – Still relatively shallow mantle melting
Washington	**Low:** Low Ba/Nb, Sr/P, Pb/Ce Lacks Nb-Ta anomaly	**Moderately Deep:** low Y/Zr, mod. Dy/Yb, **Low F? (but fertile):** High LREE, mod. HREE High % SHO (wet, low F)	**Enriched:** High Zr, Ta, Nb/Zr More enrichment eastward; highly enriched Simcoe	– Flow of enriched mantle from back-arc (slab roll-back or slab gap) – No extension = less decompression melts of metasomatized mantle vs. Graben – Fertile mantle = still productive w/low F
Graben	**Moderately low:** Mod. Ba/Zr, Sr/P, Nb-Ta High % SHO, mod % CAB **Sed melt?** Highest Cs/Rb La/Sm, K/Rb; low Pb/Ce	**Moderately Deep:** Low Y/Zr, HREE Steep REE, High K_2O High % SHO (wet, low F)	**Enriched:** High Nb/Yb, Nb/Zr For all basalt types	– Extension = melts metasomatized mantle(=larger fluid signature than WA) – Higher flux = more diversity – Enriched mantle (roll-back or slab gap) – Sediment (melt?): deep offshore sed
South	**Highest:** Highest Ba/Nb, Pb/Ce, Sr/P **Deep + shallow fluids** **Slab melt not seg-wide** High HREE, Y, Cr, Sc; low Sr/Y	**Shallow:** Highest Y/Zr, lowest Dy/Yb **Highest F**: Shallow REE, Lowest K2O (+geophysical data)	**Mod. depleted:** mod. Nb/Zr HLP-depletes mantle Gorda MORB is more depleted vs. JDF	– Fractured Gorda = high fluid delivery – Greatest extension and basalt flux – More decompression melting of variably enriched and fluid-modified mantle

ous studies have explored the causes of compositional variability within the Cascades arc using individual lava compositions (e.g. Bacon et al., 1997; Green and Harry, 1999; Leeman et al., 1990, 2005; Mullen and Weis, 2015; Schmidt et al., 2008), our study differs in that we are examining differences between the unbiased mean compositions of statistically distinct segments to elucidate causes for these differences.

The four most important parameters that control mantle melt compositions in arcs are: subduction fluid contribution, depth and degree of mantle melting, and mantle fertility (Pearce and Peate, 1995; Conrey et al., 1997; Till et al., 2013). In the following sections we will interpret inter-segment compositional differences with respect to these parameters and infer how along-arc changes in regional tectonics, mantle flow, and the age, composition, and geometry of the subducting slab may be responsible for these distinctions. We summarize these in Table 4.

5.3.1. Subduction fluid signature

The Baker, Glacier Peak, and South Segments demonstrate the largest influence of fluid-flux melting. These are each characterized by an arc-typical negative Nb and Ta anomaly (Fig. 8), as well as a relative enrichment in LILE, such as high Sr/P, Ba/Nb (Fig. 9) and Ba/Ce, which are indicators of the influence of slab fluids (Pearce, 1982; Borg et al., 1997; Pearce and Stern, 2006; Ruscitto et al., 2010). Despite differences in the age of the slab at the convergent margin (Wilson, 2002), almost all volcanoes of the Washington, Graben, and South Segments (besides Mount St. Helens) overlie 16–17 Ma subducted lithosphere (Green and Harry, 1999). Therefore, slab age and temperature are not likely the cause for the fluid flux differences that we observe between these segments. Intense fracturing and internal deformation of the subducting Gorda Plate due to motion along the Mendocino Fracture Zone (Wilson, 2002) may lead to increased fluid penetration (McCrory et al., 2012) and higher delivery of subduction fluids into the South Segment (Grove et al., 2003; Schmidt et al., 2008). Other studies have previously demonstrated the more hydrous nature of lavas and melt inclusions near Lassen Peak (Borg et al., 1997) and Mount Shasta (Sisson and Layne, 1993; Grove et al., 2002). It has also been suggested, based on radiogenic Sr, Nd, and Pb isotopes in these primitive lavas, that some of the fluid-flux signature could be from prior metasomatism of the mantle by much older subduction fluids (Borg et al., 1997; Borg et al., 2002; Carlson et al., 2018). Subduction rates were much faster in the past 40 Ma (Verplanck and Duncan, 1987), which would have caused dehydration of the slab to take place further inboard of the subduction zone, thereby modifying the composition of the mantle wedge (Leeman et al., 2005). Modern extension, with higher rates in the south could allow for higher degrees of decompression melting of this metasomatized mantle.

In addition, the South Segment may record the overall largest degree of mantle melting, as its bootstrapped mean has the flattest REE (lowest Ce/Yb, Fig. 9), which is in stark contrast to the neighboring Graben Segment, that has the steepest REE pattern (Fig. 10). This is consistent with thermobarometry results of Ruscitto et al. (2010) that indicate higher degrees of partial melting (>20%) of a more hydrous source (>0.7 wt.% H_2O) in the Mount Shasta region compared to volcanoes of the Graben Segment (9–11% melting, 0.4–0.56 wt.% H_2O). Note that Streck and Leeman (2018) argue that these Mount Shasta compositions (High Mg basaltic andesites to andesites) can also be produced by mixing between primitive calc-alkaline basalt and high proportions of resident dacite. The relatively flat REE patterns of the South Segment are not likely to be simply the result of shallower melting, as HREE and Y abundances are similar to Graben segment, and previous

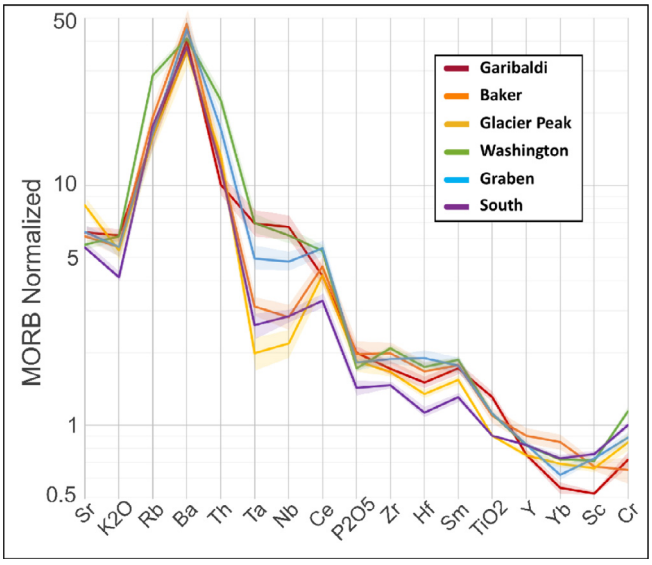

Fig. 8. "Spider diagram" showing the MORB-normalized (Pearce, 1983) bootstrapped mean concentrations of minor and trace elements. Light shaded fields give the 95% confidence intervals of the means. Segment color scheme is the same as in Fig. 4. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

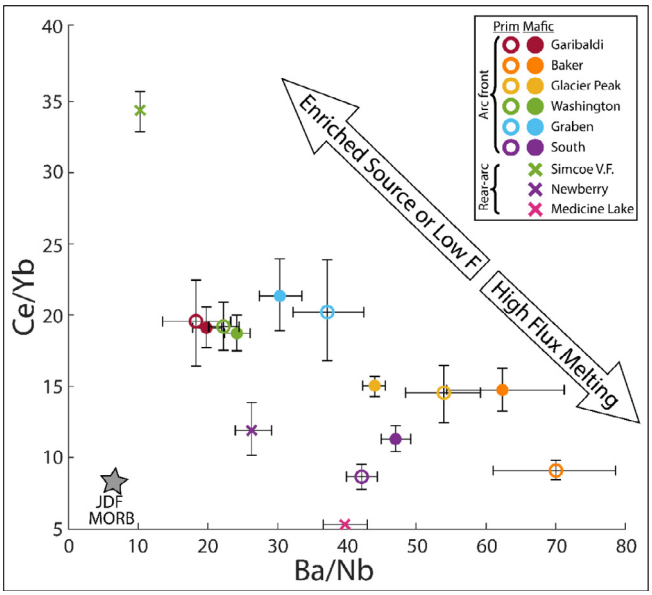

Fig. 9. Ba/Nb vs. Ce/Yb diagram of bootstrapped mean values and 95% confidence intervals of the mean for primitive (open circles) and mafic (closed) compositions of the new segments. Mafic compositions for rear-arc volcanoes are also plotted (x). The mean value for Juan de Fuca MORB (Gill et al., 2016) is plotted. Gorda MORB has Ce/Yb value (3.5) that could not be shown on this scale.

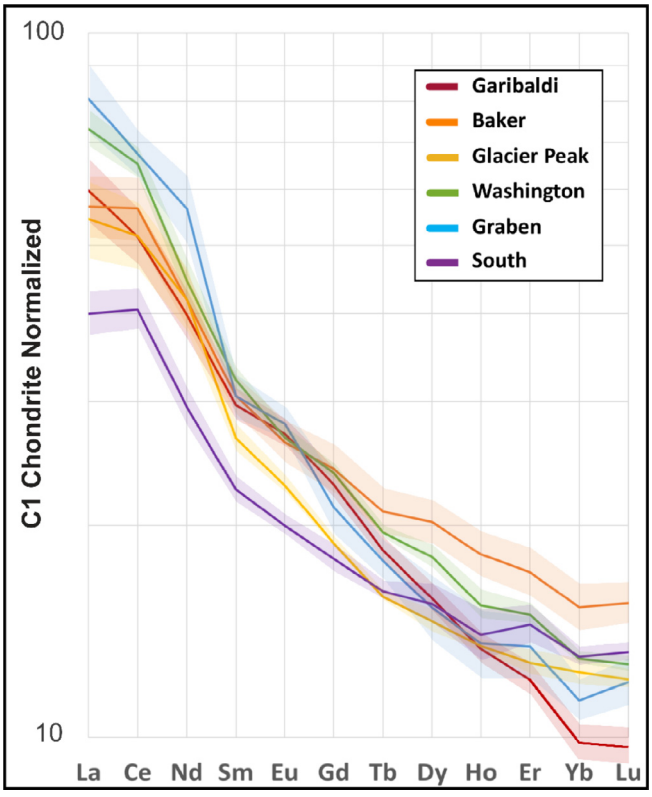

Fig. 10. REE plot of the C1-Chondrite normalized (Sun and McDonough, 1989) bootstrapped mean values for the new segments. Lightly shaded fields give the 95% confidence intervals of the means.

studies have shown similar melt segregation depths for volcanoes from these two segments (Elkins Tanton et al., 2001; Leeman et al., 2005; Ruscitto et al., 2010; Till et al., 2013). Furthermore, beneath the southern portion of the arc, geophysical studies find a large region of supra-slab magnetotelluric (MT) conductivity (Wannamaker et al., 2014) as

well as 50–80 km deep low shear velocity (Vs) (Liu et al., 2012), both interpreted to represent highest degree of flux melting in the Cascades arc (Wannamaker et al., 2014). High-degree mantle melting is consistent with heat flow measurements, geophysical data, and petrologic heat budget estimates that indicate a much larger magmatic flux in the southern half of the arc (Ingebritsen and Mariner, 2010; Till et al., 2019; Janiszewski et al., 2019).

Primitive lavas from the Glacier Peak region are the most enriched in Sr, Th, U, and Pb, both in terms of the elemental concentration, as well as in ratios where a high field strength element such as Nb or Ta is the denominator. These particular elements have been demonstrated to have fluid-rock partition coefficients that are highly temperature dependent, such that Sr, Th, and U are fluid-immobile (D = 0.3, 0.1, 0.1 respectively) at lower temperatures (i.e. 700 °C at 4 GPa) but then become highly fluid-compatible (D = 22,23,17, resp.) at higher temperatures (1100 °C) (Kessel et al., 2005). Thus, the enrichment of these elements in Glacier Peak, which lies further east than any other Cascades volcano, may be the result of the input of deeper and hotter supercritical subduction fluids (Pearce and Stern, 2006). In fact, McCrory et al. (2012) estimate that the volcano is situated approximately 95 km above the Juan de Fuca slab, whereas the neighboring volcanoes, Mount Baker and Mount Rainier are 80 and 70 km above the slab, respectively. This composition-depth relationship can also be seen at Medicine Lake, a southern rear-arc volcano that lies approximately 85 km above the slab (McCrory et al., 2012) and has higher Th, U, Th/Nb and Th/Ta than the arc-front South segment.

The Garibaldi Segment exemplifies the lowest fluid-flux signal in the arc (lowest Ba/Nb, Li, Th/Nb, Sr/P), which may result from the slab being younger (Wilson, 2002) and hotter than the sub-arc slab to the south (Syracuse et al., 2010) which would cause dehydration of the slab significantly trench-ward of the arc (Green and Sinha, 2005; Schmidt et al., 2008). The Washington Segment also has a reduced fluid signature, lacking the arc-typical Nb-Ta anomaly (Fig. 8) and the second lowest Sr/P, Ba/Nb, Pb/Ce, and Th/Ta ratios.

Although the Graben Segment has a reduced fluid component compared to that of the Glacier Peak, Baker, and South Segments, it does have a small Nb-Ta anomaly and significantly higher Ba/Zr and Sr/P than the neighboring Washington Segment. The segment also has the highest proportion of absarokite/shoshonite primitive lavas, indicative of low-degree melting of highly metasomatized and refractory mantle, as discussed in Section 5.2 (Morrison, 1980; Conrey et al., 1997; Rowe et al., 2009). In addition, the Graben Segment has the highest primitive Cs/Rb, K_2O/Rb, and La/Sm, indicating that there may be a larger contribution from subducted sediment (Borg et al., 1997; Labanieh et al., 2012), as offshore drill cores of Cascadia sediment have elevated values for these ratios (Carpentier et al., 2014). Although isotope data is limited (n = 8), the segment tends to be more radiogenic in Sr and Nd than all segments to the north, possibly indicative of a sediment signature (Carpentier et al., 2014; Mullen et al., 2017). The basement rock is inferred to be young and mafic and a lack of correlation between SiO_2 and Sr and Nd isotopes for the Graben Segment is inconsistent with crustal contamination being entirely responsible for the more radiogenic signature. Furthermore, low Pb/Ce indicates that the sediment contribution is likely as melt rather than aqueous fluid (Brenan et al., 1995; Kelemen et al., 2003), as Pb would be more strongly partitioned into the fluid. However, Ce and Pb become equally compatible in the fluid above temperatures of 1000 °C (at pressures of 4 GPa) (Kessel et al., 2005).

Our segmentation scheme places the boundary between the Graben and Washington Segments at the approximate latitude of the pole of rotation of the Cascades fore-arc block which causes overall compression in the Washington Segment and extension in the Graben Segment (Wells et al., 1998; McCaffrey et al., 2007; Labanieh et al., 2012). A marked increase in hydrothermal heat discharge and crustal heat flow (Fig. 4) south of 44.75°N latitude (near Mount Jefferson) is consistent with extensional stresses (Ingebritsen and Mariner, 2010) and increased flux of basalt and magmatic heat from the mantle (Till et al., 2019). Extension in the Graben Segment has been suggested to induce decompression mantle melting (Conrey et al., 2002; Conrey, 2004) or focus upwelling mantle that is already undergoing decompression melting directly into the arc (Till et al., 2013). The increased fluid-flux signature in the Graben Segment relative to the Washington Segment could be the result of slightly higher regional focusing of decompression melted mantle (lithospheric or asthenospheric) that was previously metasomatized by ancient subduction fluids and sediment melt. A similar process for the genesis of CAB-like signatures has been proposed for the South Segment (Borg et al., 1997; Borg et al., 2002; Leeman et al., 2005), in the back-arc (Carlson et al., 2018), and in the Great Basin (Harry and Leeman, 1995). This provides a mechanism for increasing the fluid signature in the Graben Segment relative to the Washington Segment despite being a similar distance from major slab termini and similar subducted slab age and depth (Green and Harry, 1999; Wilson, 2002; McCrory et al., 2012). This ancient metasomatism would also help explain the radiogenic Pb and Sr isotope signature of the primitive lavas in the region, although it should be noted that the Washington Segment has a slightly more radiogenic Pb isotopic signature, and isotopic data for the Graben Segment are scarce (n = 8). Such a process may be important at other hot arcs worldwide, such as the western Mexico, Marianas, or Aleutian arcs (Righter, 2000; Borg et al., 2002; Syracuse et al., 2010), where slab dehydration may largely occur trench-ward of the arc and decompression melting of subduction-modified mantle wedge may be required to explain the presence of fluid signatures in calc-alkaline basalts.

5.3.2. Depth of mantle melting and slab melting

Differences in REE abundances between segments also suggest there are differences in the overall mean depth of melting and the degree to which slab melts contribute. The Garibaldi Segment of the arc has a bootstrapped mean composition that suggests greater involvement of garnet as

a residual phase, thereby indicating deeper mantle melting than elsewhere in the arc. The segment has the lowest concentration of Y, Yb, Sc, Cr (Fig. 8), and has the steepest HREE depletion (Fig. 10). Furthermore, the bootstrapped mean occupies a unique position on a plot of Dy/Yb vs. Dy/Dy* that is separate from the rest of the arc (Fig. 11). The ratio of Dy/Dy*, which gives the ratio of measured Dy over the value interpolated by La and Yb, describes the concavity of the REE pattern (Davidson et al., 2013). By comparing this with Dy/Yb, the slope of the middle and heavy REE, one can infer the degree of LREE/HREE enrichment or depletion compared to MORB, and whether presence of amphibole (or clinopyroxene) or garnet had had a greater role during differentiation (Davidson et al., 2013). Fractionation or mantle melting with amphibole (or clinopyroxene) residual would lead to a more concave MREE pattern (decrease in Dy/Dy*) with only a minor decrease in Dy/Yb. A garnet residuum leads to HREE depletion relative to MREE (large increase in Dy/Yb), with little change to the concavity of the REE (Dy/Dy*) (Davidson et al., 2013). While the rest of the arc lies within the variably enriched MORB-like field, the bootstrapped mean for the Garibaldi Segment lies at much higher Dy/Yb, indicating that it alone has an average composition that is characterized by deeper melting.

The Garibaldi Segment also has the highest Nb/Zr and Nb/Yb (Fig. 12) of any segment, which may indicate a more enriched, IPB-like source (Pearce and Stern, 2006). After accounting for sampling bias, our bootstrapped analysis of primitive compositions suggests that approximately 70% of primitive lavas in the Garibaldi portion of the arc are IPB, with less than 10% CAB compositions. Modeling by Mullen and Weis (2015) suggest that the depth of mantle melting increases with latitude by up to 2 GPa from Mount Baker to the northernmost part of the Garibaldi segment, where the authors suggest basalts segregated at up to 1510 °C and 3.5 GPa—much hotter and deeper than predicted for the Cascades mantle wedge (Syracuse et al., 2010). Thus the hot, deep, and IPB-like signature that is prevalent in the Garibaldi Segment is interpreted to result from toroidal flow of sub-slab enriched mantle around the slab edge (Mullen and Weis, 2015).

In stark contrast, IPB lavas are completely absent in the two neighboring segments to the south, which seem to lack the deep and enriched mantle melting trace element signatures. The Mount Baker and Glacier Peak regions have the highest Y, Yb, and Y/Zr values and relatively low Dy/Dy*, Nb/Yb and Nb/Zr (Figs. 11 and 12) indicating a more depleted source in which garnet is not likely a residual phase, consistent with the conclusions of Sas et al. (2017). Deep melting of an enriched mantle source seen in the Garibaldi Segment becomes progressively less southwards (Green and Harry, 1999; Green and Sinha, 2005; Mullen and Weis, 2015). Our results indicate that the magnitude of southward depletion is statistically significant over the one-degree latitude that separates Mount Garibaldi and Mount Baker.

Slab melting could also contribute to the geochemical signature of the Garibaldi Segment. In addition to having the highest Dy/Yb, the segment has the second highest Sr/Y ratio of any portion of the arc, a signature which has been used by many authors to suggest slab melting in arcs worldwide (Defant and Drummond, 1990; Pearce and Peate, 1995; Stern and Kilian, 1996; Kelemen et al., 2003). Slab melting would be a reasonable process in this portion of the arc, since the slab is much deeper (McCrory et al., 2012), younger (13–16 Ma Green and Harry, 1999), and is the global endmember for slab surface temperature (Syracuse et al., 2010). Additionally, toroidal flow of hot mantle has been suggested to cause thermal erosion of the edges of slabs (Thorkelson and Breitsprecher, 2005). Although previous authors suggest a slab melting

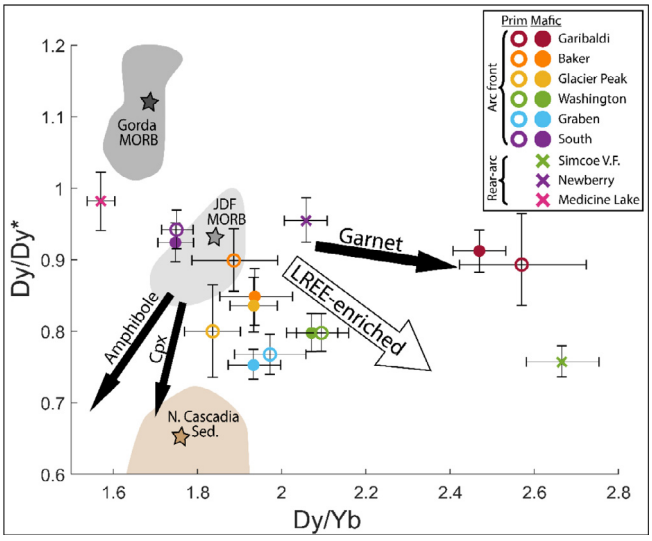

Fig. 11. Dy/Yb vs. Dy/Dy* as defined by Davidson et al. (2012). Bootstrapped means of primitive and mafic compositions are shown. Fields of data from Gorda MORB (Davis et al., 2008), Juan de Fuca MORB (Gill et al., 2016), and offshore North Cascadia sediment (Carpentier et al., 2014) are also plotted, along with the mean value (stars). Effect of enrichment of LREE is shown. Arrows indicate the effect of amphibole, clinopyroxene or garnet as residual phases during melting (Davidson et al., 2013).

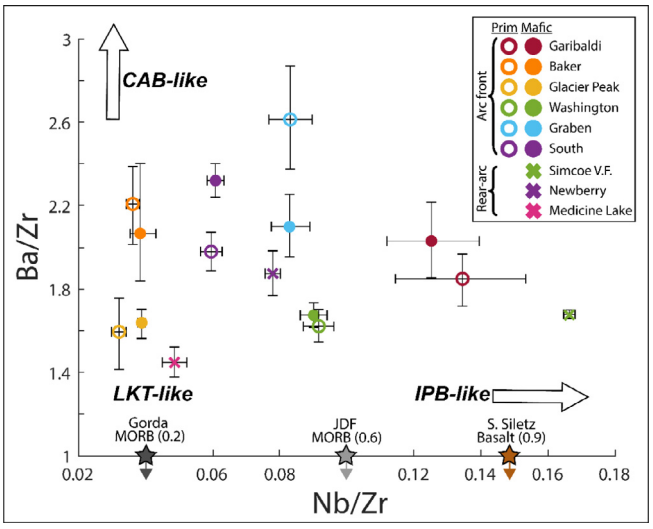

Fig. 12. Nb/Zr vs. Ba/Zr of bootstrapped means of new segments and rear-arc volcanoes. Mean Nb/Zr values for Gorda MORB, JDF MORB, and Siletz basalt (south of 47°N) are depicted, and Ba/Zr values are in parentheses (Davis et al., 2008; Gill et al., 2016; Phillips et al., 2017). Approximate locations of the three most common Cascades endmembers are shown.

origin for one HMBA unit erupted in the Baker Segment (Moore and DeBari, 2012; Sas et al., 2017), our bootstrapped mean compositions for the Segment (with or without HMBA) have too low Sr/Y and Dy/Dy* and too high Yb to suggest that slab melting is common throughout the Segment. While most of our discussion is focused on mafic and primitive magmas, it is interesting to note that in our data compendium, 34% of all Garibaldi Segment of dacite to rhyolite rocks have adakite-like characteristics (Sr/Y $\geq$ 20, Sr $\geq$ 400 ppm, Y $\leq$ 18 ppm, La/Yb $\geq$ 20, Yb $\leq$ 1.9 ppm, and Al2O3 $\geq$ 15 wt.%), and the bootstrapped mean Sr/Y is four times greater than the Baker Segment immediately to the south. These adakite compositions may be indicative of direct slab melts (±mantle peridotite), as suggested for Archean TTGs and at other unusually hot modern subduction zones (Defant and Drummond, 1990; Richards and Kerrich, 2007; Moyen, 2009). Although the term adakite has perhaps been misused, as suggested by Moyen (2009), Garibaldi Segment dacites-rhyolites have high Mg# (50), and the highest Nb/Th (2.93) and lowest Rb/Sr (0.04) of any segment, indicating that the adakite-signature is not simply due to deep crustal melting, as with the misnomer "continental adakites" (Moyen, 2009).

Although Walowski et al. (2015) demonstrated that slab melting likely occurs in the southernmost Cascades beneath Lassen Peak, our bootstrapped data indicate that this may not be a dominant process throughout the entire South Segment, since the bootstrapped mean has high HREE, Y, Cr, and Sc, and the lowest Dy/Yb and Sr/Y of any segment. This is true even if we calculate bootstrapped mean of only the California portion of the Cascades. Thus, it is possible that slab melting in the South Segment is only localized to the region beneath Lassen Peak, where larger slab depths and hot toroidal flow may allow for more slab melting.

While our approach provides new insight into petrogenesis within the Cascade arc, it is important to remember, however, that our bootstrapped means represent a mixture of various basalt compositions and thus record variable combinations of the petrogenetic processes that produce those diverse basalts. Since most segments have all of the different endmember basalt types, but in different proportions, a wide variety of depth, degree, and type of melting are simultaneously occurring beneath a particular segment, but in different proportions. For example, LKT-type basalts tend to be the result of shallow high-degree mantle melts while IPB-types tend to be the result of deeper melting at low degrees (Leeman et al., 2005; Rowe et al., 2009; Till et al., 2013). Thus, our mean compositions help to infer which of the processes was most prevalent in each segment.

5.3.3. Degree of mantle melting

The degree to which the mantle is partially melted beneath an arc has been suggested to vary significantly between different primitive endmember magmas. Alkalic, shoshonite, and IPB-type basalts are interpreted to be the result of low degree partial melting, while LKTs and CABs are likely the result of much higher degrees (Rowe et al., 2009). One signature of high degree of partial melting is a relative depletion of incompatible trace elements (Green and Sinha, 2005; Pearce and Stern, 2006; Rowe et al., 2009). The South Segment, which we suggest demonstrates a large fluid-flux melting signature (Fig. 9), has bootstrapped mean abundances of incompatible elements such as Ti, Nb, Ta, Zr, and LREE (Fig. 10) that may indicate high-degree mantle melting. This is consistent with basalt flux from the mantle being the highest in the South Segment, as suggested by Till et al. (2019) based on calculations of heat flux from volcanic volumes, thermal modelling, and seismic phase velocities.

However, differing degrees of mantle partial melting cannot explain all of the trends that we observe. The Washington and Graben Segments have the second and third highest concentrations of Nb, Zr, Ta (Fig. 8) and Nb/Yb, Nb/Zr (Fig. 12), as well as the steepest REE patterns (Figs. 9 and 10) of the arc-front. One could interpret these trends to be the result of low degree melting. In addition, these two segments have the highest proportion of shoshonitic basalts, which have been suggested to be the result of very low-degree partial melting (Morrison, 1980; Bacon, 1990; Rowe et al., 2009). However, the Graben Segment is also the second most magmatically productive region of the Cascades (Sherrod and Smith, 1990; Hildreth, 2007; Till et al., 2019), and heat flux within the segment is the highest in the arc (Ingebritsen and Mariner, 2010). It is difficult to reconcile the high productivity of the region with an overall low degree of partial melting of the mantle. Furthermore, one would expect that low degree partial melting would result in higher Nb/Ta values, since both are HFSE but Nb is slightly more incompatible (Pearce and Stern, 2006). However, Nb/Ta is the lowest in the Graben and Washington segments.

Increased fertility of an enriched mantle source produces melts with very similar incompatible element-enriched trends as small degree partial melt, making it difficult to discern which process is more responsible (Borg et al., 1997). Experimental melting results indicate that while other major elements behave similarly in both processes (increased Ti, Na, K, Al and decreased Si, Mg), Fe only increases in the case of increased mantle fertility (but is constant at differing melting degrees) and Ca only decreases due to lower degree melting (Klein and Langmuir, 1987; Borg et al., 1997; and references therein). If we compare segment mean compositions of individual primitive magma types, so as to avoid the effect of different melting degrees inferred for each type, we find that each type is higher in FeO in the Washington and Graben Segment than basalts of that type anywhere in the arc (besides the Garibaldi segment), but that CaO is nearly constant. Furthermore, for each individual primitive basalt type, including CABs, which are suggested to be the result of the highest degree melting (Reiners et al., 2000; Rowe et al., 2009), the Washington and Graben Segments have significantly higher Hf, Nb/Zr, Nb/Yb, and steeper REE than any other segment. This suggests that CABs from these segments must have been the result of hydrous melting of a more enriched source than CABs from other segments. Although there is a general consensus that the various primitive basalt types require differing degrees of melting, these trends for a given basalt type may suggest that variable mantle fertility drives the inter-segment differences in incompatible element enrichment rather than just different proportions of low-degree and high-degree basalt types. We discuss this further in the next section.

5.3.4. Mantle fertility beneath the arc-front and rear-arc

As discussed above, the Washington and Graben Segments are more enriched in incompatible trace elements (HFSE and LREE) than all other arc segments except, in some cases, the Garibaldi Segment. However, the Simcoe Volcanic Field, which lies in the rear-arc behind the Washington Segment, is more enriched than anywhere in the arc-front (Fig. 12). The enriched signature of the rear-arc compared to that of the Washington Segment, may indicate westward flow of enriched mantle (Leeman et al., 2005) which is first slightly depleted by rear-arc volcanism before being melted beneath the arc-front. A similar pattern of depletion from the back arc to the arc-front is seen in the Izu-Bonin, Scotia, and Marianas arcs (Pearce and Stern, 2006). This mantle flow is further evidenced by mantle anisotropy observed by various studies, with the fast splitting direction orientated perpendicular to the trench (east-west) (Long et al., 2009; Wannamaker et al., 2014; Long, 2016). Further evidence of the relationship between the enriched rear-arc mantle and the Washington Segment is provided by Mullen et al. (2017) who demonstrate that primitive lavas from Mt Adams deviate from the normal Cascades arc trend in terms of Sr Nd, and Pb isotopes and instead lie on a separate "Adams array" that trends toward a different Simcoe endmember. Slab-induced corner flow could provide the arc with enriched mantle material from the back arc (Long et al., 2012). Alternatively, a slab gap or tear, as suggested by several tomography studies to exist beneath the Washington and Graben segments (Gao and Shen, 2014; Hawley and Allen, 2019) could provide both of the arc-front and back-arc regions with enriched sub-slab mantle material. Either mechanism of upwelling deeper fertile mantle could also help to explain the relatively deep melting signature (low HREE and Y/Zr) of the Graben and Washington Segments (Figs. 11 and 12).

Several authors have discussed the compositional diversity of basalts within the southern portion of the Washington Segment (Cascades Columbia Transect of Leeman et al., 2005), where the arc is widest (45.5–46.5°N), however, our new compilation allows us to better assess east-west changes in the relative proportions of the basalt end-members We observe a distinctive westward decrease in the proportion of IPB-type primitive basalts in this region, consistent with enriched mantle flowing from the back arc and becoming progressively depleted westward (Fig. 13). Furthermore, for each individual primitive basalt type, we observe a distinctive pattern of decreasing La/Nd, Nb/Zr, Nb/Yb eastward from the forearc (122.5°W longitude), then an inflection just west of Mount Adams (121.75°W), and a subsequent eastward increase into the Simcoe rear-arc (121.25–120.75°W) (Supplement S6). This is consistent with lower degree melting in the forearc (where most of the low-degree shoshonite melts are erupted, Fig. 13), then higher degree melting beneath the volumetrically productive Mount St. Helens and Indian Heaven, and then increasing influence of enriched mantle eastward.

A similar, albeit more reduced, eastward enrichment in Nb/Zr is seen from the arc-front of the South Segment to the rear-arc Newberry volcano (Fig. 12), indicating that such mantle flow may be common in the middle portions of the arc (Hawley and Allen, 2019). It is interesting to note that while 66% of Simcoe primitive basalts are IPB-type, they account for only 25% of Newberry primitive basalts. It is possible that continuous depletion of westward-flowing mantle by the High Lava Plains volcanics of eastern

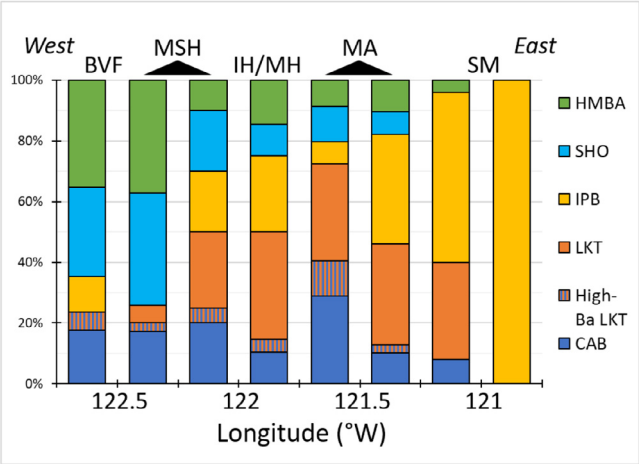

Fig. 13. West to east changes in the proportion of primitive types for the region between 45.5°N and 46.5°N (Cascades Columbia Transect, Leeman et al., 2005). Criteria for primitive classification and color scheme is the same as in Fig. 7 (as defined in Table 3). Total samples from the literature that could be uniquely classified = 300. Positions of Boring Volcanic Field (BVF), Mount St. Helens (MSH), Indian Heaven (IH), Mount Hood (MH), Mount Adams (MA), and Simcoe Volcanic Field (SM) are given for reference. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Oregon (Long et al., 2012; Till et al., 2013) prior to being melted beneath Newberry has led it to have a more depleted signature than the Simcoe rear-arc volcanics to the north.

5.3.5. Role of the overriding lithosphere

The lithosphere beneath the Cascades arc is made up of a patchwork of accreted terrains of differing lithologies and ages, and some our segment boundaries seem to coincide with suggested changes in the underlying basement. In particular, the northern boundary of the Washington Segment and southern boundary of our Graben Segment correspond with the suggested limits of the Columbia Embayment a region of accreted oceanic crust (Trehu et al., 1994), that includes the Late Paleocene-Eocene Siletz Terrain, interpreted to be an oceanic large igneous province (LIP) with a plume-origin (Wells et al., 2014; Phillips et al., 2017). While some authors suggest based on geophysical data that the Siletz Terrane may terminate west of the modern High Cascades boundary (Finn, 1990; McCrory and Wilson, 2013; Bedrosian and Feucht, 2014), others hold that Siletz lithosphere could extend well east of the arc, near the Idaho border (Gao et al., 2011). Our results suggest that the overall IPB-like composition of the Washington and Graben Segments is not likely due to interaction with the accreted Siletz terrain lithosphere as previous authors have suggested (Church et al., 1986; Schmidt et al., 2008). Assimilation of enriched Siletzia crust is unlikely since we would expect such assimilation to cause compositional differences between the primitive and mafic composition of a given segment, yet these bootstrapped means are indistinguishable for both segments. Furthermore, although Siletz basalts (south of 47°N) are slightly more enriched in Nb/Yb and Nb/Zr (Phillips et al., 2017) than the primitive compositions of either Washington or Graben Segments, they are much less enriched than the Simcoe Volcanic Field (Fig. 12), indicating that Siletz mantle cannot be the source of the rear arc enrichment. Additionally, if the enrichment pattern were the result of melting Siletz lithospheric mantle, we would expect IPB-type basalts to be similarly prevalent throughout the east-west transect across the southern Washington Segment (CCT), yet, as mentioned above, we observe a steady eastward increase in proportion of IPB-type basalts (Fig. 13). Thus, we suggest that a westward flow of fertile mantle from the back arc is a more reasonable explanation for enrichment trends of these two segments and the rear arc Simcoe Volcanic Field.

Other authors have discussed the potential role of the lithosphere (crust or mantle) in petrogenesis of primitive basalts elsewhere in the arc, whether as a lithospheric mantle melting source for primitive compositions (Leeman et al., 1990; Harry and Leeman, 1995; Leeman et al., 2005; Mullen et al., 2017) or as a composition-modifying assimilant (Bacon, 1990; Carlson et al., 2018). As discussed above, we suggest that much of the slab signal (fluid and sediment) in the arc, particularly in the South Segment, could be previously-metasomatized lithospheric mantle, and the "lithosphere interaction model" (Mullen et al., 2017) discussed by the aforementioned authors may be a viable process by which the various basalt endmembers are created. However, this does not mean that the observed differences in the age of the crustal lithosphere has a direct effect on the compositions of primitive basalts. Despite a wide range in the ages of the accreted crustal rocks that surround the Cascades (Righter, 2000), the relatively narrow range of isotopic compositions precludes the possibility that basalt diversity is the result of isotopically-heterogeneous accreted lithospheric mantle domains (Mullen et al., 2017). For example, the Glacier Peak and Baker Segments are situated above much older Paleozoic lithosphere (Righter, 2000), however the CABs of these segments are less radiogenic than those of the Washington Segment to the south, suggesting that the age of the crustal lithosphere does not directly correlate with the age of the metasomatized mantle beneath it. Thus, we suggest that although the lithosphere may affect the composition of erupted magmas or even be their source, the *primary drivers* of the

observed *along-arc segmentation* of the Cascades are proportional differences in depth, degree, and type of melting (decompression or flux), and variable enrichments by fertile rear-arc mantle and modern and ancient slab fluids.

6. CONCLUSIONS

The Cascades arc exhibits significant along-strike variability in major and trace element compositions of mafic lavas, and numerous studies have proposed mechanisms such as mantle heterogeneity, regional tectonics, and/or geochemical differences in the overlying lithosphere or subducting slab to be responsible for these compositional differences. Although considerable heterogeneity exists even at a single location, partitioning the arc into compositionally distinct groups allows one to explore the regional-scale causes for such differences. Our study develops a novel approach in which we compile a new dataset of major and trace element analyses of over 2000 mafic samples from published Cascades literature, utilize a Monte Carlo approach with bootstrap resampling to reduce inherent sampling bias, and partition the arc using entirely objective and statistically-based methodology. We propose a new segmentation scheme that maximizes the geochemical differences between six segments: Garibaldi, Baker, Glacier Peak, Washington, Graben, and South Segments. Using the same expanded dataset on both segmentation schemes, our new segments are on average three times more statistically distinct than those of Schmidt et al., 2008, who drew segment boundaries based on more objective methods and using a limited dataset (390 samples). By separating the arc into the most statistically disparate regions and utilizing our new data compilation to calculate unbiased mean compositions of each arc segment, we can better assess the different processes that lead to along-strike geochemical heterogeneity within the Cascades arc.

Our results demonstrate significant intersegment differences in slab fluid/melt input, depth and degree of mantle melting, and mantle fertility. We attribute these differences to along-strike changes in local tectonics, mantle flow, depth of the slab, proximity to slab edges or gaps, and degree of metasomatism of the mantle wedge. While differences in slab age beneath the Garibaldi, Baker and Glacier Peak segments may contribute to geochemical differences in these segments, this is not the case for the southern three segments where slab age is constant. Furthermore, we suggest that incompatible element enrichment trends in the Washington and Graben segments are more likely to be the result of flow of enriched mantle from the back-arc than to be the result of melting IPB-like Siletzia lithospheric mantle. Compositional variability observed at other arcs worldwide may also be the result of along-strike differences in the subduction system which, in turn, affects slab input, mantle fertility, and the degree and depth of mantle melting.

Our results also help to comment on two additional long-standing debates with respect to the Cascades arc. We suggest that the Simcoe, Newberry, and Medicine Lake volcanic fields, which lie east of the main arc axis, have compositions that are statistically distinct from eruptive centers along the main arc axis and should thus be considered rear-arc volcanic centers that require different petrogenesis than magmas from the arc-front. Our study also suggests that Mount Baker and Glacier Peak should not be considered part of the Garibaldi Segment of the arc as many previous studies have suggested (Hildreth, 2007; Schmidt et al., 2008; Mullen and Weis, 2015).

In this study, we have demonstrated the advantage of approaching important petrologic questions, such as along-arc variability, using Big Data and statistically based methodology. With the continual growth of online data repositories, we now have easy access to large datasets that can allow for much more detailed studies of particular regions or at the global scale. However, while these data can be quite powerful, great care must be taken in assessing the quality of data and reducing sampling bias. Statistical approaches, such as those outlined here, can help, and can be used to investigate other compositionally heterogeneous arcs. However, we also suggest that further focus on more complete and evenly distributed sampling of arcs worldwide is warranted. For example, in the Cascades arc, there are several regions that are under-sampled, such as between Mount Jefferson and Three Sisters, south of Three Sisters, and near Lassen Peak and Mount Shasta (Fig. 1). In addition, although the focus of this study is on segmentation of mafic magma compositions– a long debated topic in Cascadia – we include all available data (over 12,000 samples) in the compilation (Supplementary Material S1), and future work could focus on variability of more evolved magmas.

ACKNOWLEDGEMENTS

We would like to thank Brenhin Keller for his advice during the initial stages of writing the MATLAB Bootstrap code. We'd further like to thank Brenhin, Bill Leeman, Rick Conrey, and an anonymous reviewer for suggesting many improvements to the manuscript. The initial idea for this contribution was developed during a class on "Big Data in Petrology" at Oregon State University, and we would also like to thank the GEO 622 students for their help in providing inspiration and advice. Funding for Bradley Pitcher was provided by the GeoPRISMS program of the National Science Foundation [Award number 1144555].

APPENDIX A. SUPPLEMENTARY MATERIAL

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gca.2019.08.035.

REFERENCES

Bacon C. R. (1990) Calc-alkaline, shoshonitic, and primitive tholeiitic lavas from monogenetic volcanoes near Crater Lake, Oregon. *J. Petrol.* **31**, 135–166.

Bacon C. R., Bruggman P. E., Christiansen R. L., Clynne M. A., Donnelly-Nolan J. M. and Hildreth W. (1997) Primitive magmas at five Cascade volcanic fields: melts from hot, heterogeneous sub-arc mantle. 35.

Bedrosian P. A. and Feucht D. W. (2014) Structure and tectonics of the northwestern United States from EarthScope USArray magnetotelluric data. *Earth Planet. Sci. Lett.* **402**, 275–289.

Blackwell D. D., Steele J. L., Frohme M. K., Murphey C. F., Priest G. R. and Black G. L. (1990) Heat flow in the Oregon Cascade Range and its correlation with regional gravity, Curie point depths, and geology. *J. Geophys. Res.* **95**, 19475.

Blakely R. J., Christiansen R. L., Guffanti M., Wells R. E., Donnelly-Nolan J. M., Muffler L. J. P., Clynne M. A. and Smith J. G. (1997) Gravity anomalies, Quaternary vents, and Quaternary faults in the southern Cascade Range, Oregon and California: implications for arc and backarc evolution. *J. Geophys. Res. Solid Earth* **102**, 22513–22527.

Borg L. E., Blichert-Toft J. and Clynne M. A. (2002) Ancient and modern subduction zone contributions to the mantle sources of lavas from the Lassen region of California inferred from Lu-Hf isotopic systematics. *J. Petrol.* **43**, 705–723.

Borg L. E., Clynne M. A. and Btillen T. D. (1997) The variable role of slab-derived fluids in the generation of a suite of primitive calc-alkaline lavas from the southernmost Cascades, California. *Camdian Miner. Gist* **35**, 42–452.

Brenan J. M., Shaw H. F., Ryerson F. J. and Phinney D. L. (1995) Mineral-aqueous fluid partitioning of trace elements at 900°C and 2.0 GPa: constraints on the trace element chemistry of mantle and deep crustal fluids. *Geochim. Cosmochim. Acta* **59**, 3331–3350.

Brocher T. M., Wells R. E., Lamb A. P. and Weaver C. S. (2017) Evidence for distributed clockwise rotation of the crust in the northwestern United States from fault geometries and focal mechanisms. Tectonics, 2016TC004223.

Carlson R. W., Grove T. L. and Donnelly-Nolan J. M. (2018) Origin of primitive tholeiitic and calc-alkaline basalts at Newberry Volcano, Oregon. *Geochem. Geophys. Geosyst.* **19**, 1360–1377.

Carpenter J. and Bithell J. (2000) Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. *Stat. Med.* **19**, 1141–1164.

Carpentier M., Weis D. and Chauvel C. (2014) Fractionation of Sr and Hf isotopes by mineral sorting in Cascadia Basin terrigenous sediments. *Chem. Geol.* **382**, 67–82.

Carr M. J., Feigenson M. D., Patino L. C. and Walker J. A. (2004) Volcanism and geochemistry in Central America: progress and problems. In Inside the subduction factory (ed. J. Eiler). American Geophysical Union. pp. 153–174.

Church S. E., Lehuray A. P., Grant A. R., Delevaux M. H. and Gray J. E. (1986) Lead-isotopic data from sulfide minerals from the Cascade Range, Oregon and Washington. *Geochim. Cosmochim. Acta* **50**, 317–328.

Conrey R. M. (2004) SOTA field trip guide. *Oregon Dep. Geol. Miner. Ind.*, OFR-O-04-0.

Conrey R. M., Sherrod D. R., Hooper P. R. and Swanson D. A. (1997) Diverse primitive magmas in the Cascade arc, Northern Oregon and Southern Washington. *Can. Mineral.* **35**, 367–396.

Conrey R. M., Taylor E. M., Donnelly-Nolan J. M. and Sherrod D. R. (2002) North-central Oregon Cascades: exploring petrologic and tectonic intimacy in a propagating intra-arc rift. *F. Guid. Geol. Process. Cascadia* **36**, 47–90.

Davidson J., Turner S. and Plank T. (2013) Dy/Dy*: variations arising from Mantle sources and petrogenetic processes. *J. Petrol.* **54**, 525–537.

Davis A. S., Clague D. A., Cousens B. L., Keaten R. and Paduan J. B. (2008) Geochemistry of basalt from the North Gorda segment of the Gorda Ridge: evolution toward ultraslow spreading ridge lavas due to decreasing magma supply. *Geochem., Geophys. Geosyst.* **9**, n/a-n/a.

Defant M. J. and Drummond M. S. (1990) Derivation of some modern arc magmas by melting of young subducted lithosphere. *Nature* **347**, 662–665.

Donnelly-Nolan J. M., Grove T. L., Lanphere M. A., Champion D. E. and Ramsey D. W. (2008) Eruptive history and tectonic setting of Medicine Lake Volcano, a large rear-arc volcano in the southern Cascades. *J. Volcanol. Geotherm. Res.* **177**, 313–328.

Eagar K. C., Fouch M. J., James D. E. and Carlson R. W. (2011) Crustal structure beneath the High Lava Plains of eastern Oregon and surrounding regions from receiver function analysis. *J. Geophys. Res.* **116**, B02313.

Elkins Tanton L. T., Grove T. L. and Donnelly-Nolan J. (2001) Hot, shallow mantle melting under the Cascades volcanic arc. Geology 29, 631.

Elliott T. (2003) Tracers of the slab. *Am. Geophys. Union (AGU)*, 23–45.

Finn C. (1990) Geophysical constraints on Washington Convergent margin structure. *J. Geophys. Res.* **95**, 19533.

Gao H., Humphreys E. D., Yao H. and van der Hilst R. D. (2011) Crust and lithosphere structure of the northwestern U.S. with ambient noise tomography: terrane accretion and Cascade arc development. *Earth Planet. Sci. Lett.* **304**, 202–211.

Gao H. and Shen Y. (2014) Upper mantle structure of the Cascades from full-wave ambient noise tomography: evidence for 3D mantle upwelling in the back-arc. *Earth Planet. Sci. Lett.* **390**, 222–233.

Gill J., Michael P., Woodcock J., Dreyer B., Ramos F., Clague D., Kela J., Scott S., Konrad K. and Stakes D. (2016) Spatial and temporal scale of mantle enrichment at the endeavour segment, Juan de Fuca Ridge. *J. Petrol.* **57**, 863–896.

Green N. L. and Harry D. L. (1999) On the relationship between subducted slab age and arc basalt petrogenesis, Cascadia subduction system, North America. *Earth Planet. Sci. Lett.* **171**, 367–381.

Green N. L. and Sinha A. K. (2005) Consequences of varied slab age and thermal structure on enrichment processes in the subarc mantle of the northern Cascadia subduction system. *J. Volcanol. Geotherm. Res.* **140**, 107–132.

Grove T. L., Elkins-Tanton L. T., Parman S. W., Chatterjee N., Müntener O. and Gaetani G. A. (2003) Fractional crystallization and mantle-melting controls on calc-alkaline differentiation trends. *Contrib. Mineral. Petrol.* **145**, 515–533.

Grove T., Parman S., Bowring S., Price R. and Baker M. (2002) The role of an H2O-rich fluid component in the generation of primitive basaltic andesites and andesites from the Mt. Shasta region, N California. *Contrib. Mineral. Petrol.* **142**, 375–396.

Guffanti M. and Weaver C. S. (1988) Distribution of Late Cenozoic volcanic vents in the Cascade range: volcanic arc segmentation and regional tectonic considerations. *J. Geophys. Res.* **93**, 6513.

Harry D. L. and Leeman W. P. (1995) Partial melting of melt metasomatized subcontinental mantle and the magma source potential of the lower lithosphere. *J. Geophys. Res. Solid Earth* **100**, 10255–10269.

Hart W. K., Aronson J. L. and Mertzman S. A. (1984) Areal distribution and age of low-K, high-alumina olivine tholeiite magmatism in the northwestern Great Basin. Geol. Soc. Am. Bull. 95, 186.

Haukoos J. S. and Lewis R. J. (2005) Advanced statistics: bootstrapping confidence intervals for statistics with 'Difficult'; distributions. *Acad. Emerg. Med.* **12**, 360–365.

Hawley W. B. and Allen R. M. (2019) The fragmented death of the Farallon plate. Geophys. Res. Lett., 2019GL083437.

Hildreth W. (2007) Quaternary magmatism in the Cascades: geologic perspectives. *U.S. Geol. Surv. Prof. Pap.* **1744**, 125.

Hildreth W. and Moorbath S. (1988) Crustal contributions to arc magmatism in the Andes of Central Chile. *Contrib. Mineral. Petrol.* **108**, 247–252.

Ingebritsen S. E. and Mariner R. H. (2010) Hydrothermal heat discharge in the Cascade Range, northwestern United States. *J. Volcanol. Geotherm. Res.* **196**, 208–218.

Janiszewski H. A., Gaherty J. B., Abers G. A., Gao H. and Eilon Z. C. (2019) Amphibious surface-wave phase-velocity measurements of the Cascadia subduction zone. *Geophys. J. Int.* **217**, 1929–1948.

Kelemen P. B., Yogodzinski G. M. and Scholl D. W. (2003) Along-strike variation in the Aleutian Island Arc: genesis of high Mg# andesite and implications for continental crust. *Am. Geophys. Union (AGU)*, 223–276.

Keller C. B. and Schoene B. (2012) Statistical geochemistry reveals disruption in secular lithospheric evolution about 2.5 Gyr ago. *Nature* **485**, 490–493.

Kessel R., Schmidt M. W., Ulmer P. and Pettke T. (2005) Trace element signature of subduction-zone fluids, melts and supercritical liquids at 120–180 km depth. *Nature* **437**, 724–727.

Klein E. M. and Langmuir C. H. (1987) Global correlations of ocean ridge basalt chemistry with axial depth and crustal thickness. *J. Geophys. Res.* **92**, 8089.

Labanieh S., Chauvel C., Germa A. and Quidelleur X. (2012) Martinique: a clear case for sediment melting and slab dehydration as a function of distance to the trench. *J. Petrol.* **53**, 2441–2464.

Leeman W. P., Lewis J. F., Evarts R. C., Conrey R. M. and Streck M. J. (2005) Petrologic constraints on the thermal structure of the Cascades arc. *J. Volcanol. Geotherm. Res.* **140**, 67–105.

Leeman W. P., Smith D. R., Hildreth W., Palacz Z. and Rogers N. (1990) Compositional diversity of Late Cenozoic basalts in a transect across the southern Washington Cascades: implications for subduction zone magmatism. *J. Geophys. Res.* **95**, 19561.

Little R. J. A. and Rubin D. B. (2014) *Statistical Analysis with Missing Data*. John Wiley & Sons.

Liu K., Levander A., Zhai Y., Porritt R. W. and Allen R. M. (2012) Asthenospheric flow and lithospheric evolution near the Mendocino Triple Junction. *Earth Planet. Sci. Lett.* **323–324**, 60–71.

Long M. D. (2016) The Cascadia Paradox: Mantle flow and slab fragmentation in the Cascadia subduction system. *J. Geodyn.* **102**, 151–170.

Long M. D., Gao H., Klaus A., Wagner L. S., Fouch M. J., James D. E. and Humphreys E. (2009) Shear wave splitting and the pattern of mantle flow beneath eastern Oregon. *Earth Planet. Sci. Lett.* **288**, 359–369.

Long M. D., Till C. B., Druken K. A., Carlson R. W., Wagner L. S., Fouch M. J., James D. E., Grove T. L., Schmerr N. and Kincaid C. (2012) Mantle dynamics beneath the Pacific Northwest and the generation of voluminous back-arc volcanism. *Geochem., Geophys. Geosyst.* **13**, n/a-n/a.

Matsumoto M. and Nishimura T. (1998) Mersenne Twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Trans. Model. Comput. Simul.* **8**, 3–30.

McCaffrey R., Qamar A. I., King R. W., Wells R., Khazaradze G., Williams C. A., Stevens C. W., Vollick J. J. and Zwick P. C. (2007) Fault locking, block rotation and crustal deformation in the Pacific Northwest. *Geophys. J. Int.* **169**, 1315–1340.

McCrory P. A., Blair J. L., Waldhauser F. and Oppenheimer D. H. (2012) Juan de Fuca slab geometry and its relation to Wadati-Benioff zone seismicity. J. Geophys. Res. Solid Earth 117.

McCrory P. A. and Wilson D. S. (2013) A kinematic model for the formation of the Siletz-Crescent forearc terrane by capture of coherent fragments of the Farallon and Resurrection plates. *Tectonics* **32**, 718–736.

Meng X.-L. and Rubin D. B. (1993) Maximum likelihood estimation via the ECM algorithm: a general framework. *Biometrika* **80**, 267–278.

Mojena R. (1977) Hierarchical grouping methods and stopping rules: an evaluation. *Comput. J.* **20**, 359–363.

Moore N. E. and DeBari S. M. (2012) Mafic magmas from Mount Baker in the northern Cascade arc, Washington: probes into mantle and crustal processes. *Contrib. Mineral. Petrol.* **163**, 521–546.

Morrison G. W. (1980) Characteristics and tectonic setting of the shoshonite rock association. *Lithos* **13**, 97–108.

Moyen J.-F. (2009) High Sr/Y and La/Yb ratios: the meaning of the "adakitic signature". *Lithos* **112**, 556–574.

Mullen E. K. and Weis D. (2015) Evidence for trench-parallel mantle flow in the northern Cascade Arc from basalt geochemistry. *Earth Planet. Sci. Lett.* **414**, 100–107.

Mullen E. K., Weis D., Marsh N. B. and Martindale M. (2017) Primitive arc magma diversity: new geochemical insights in the Cascade Arc. *Chem. Geol.* **448**, 43–70.

Patino L. C., Carr M. J. and Feigenson M. D. (2000) Local and regional variations in Central American arc lavas controlled by variations in subducted sediment input. *Contrib. Mineral. Petrol.* **138**, 265–283.

Pearce J. A. (1982) Trace element characteristics of lavas from destructive plate boundaries. In *Orogenic Andesites and Related Rocks* (ed. R. S. Thorpe). John Wiley and Sons, Chichester, England, pp. 528–548.

Pearce J. A. (1983) Role of the sub-continental lithosphere in magma genesis at active continental margins. Cont. basalts mantle xenoliths.

Pearce J. A. and Peate D. W. (1995) Tectonic implications of the composition of volcanic ARC magmas. *Annu. Rev. Earth Planet. Sci.* **23**, 251–285.

Pearce J. A. and Stern R. J. (2006) Origin of back-arc basin magmas: trace element and isotope perspectives. *Am. Geophys. Union (AGU)*, 63–86.

Pek J., Wong A. C. M. and Wong O. C. Y. (2017) Confidence intervals for the mean of non-normal distribution: transform or not to transform. *Open J. Stat.* **07**, 405–421.

Phillips B. A., Kerr A. C., Mullen E. K. and Weis D. (2017) Oceanic mafic magmatism in the Siletz terrane, NW North America: fragments of an Eocene oceanic plateau? *Lithos* **274–275**, 291–303.

Pitcher B. W., Kent A. J. R., Grunder A. L. and Duncan R. A. (2017) Frequency and volumes of ignimbrite eruptions following the Late Neogene initiation of the Central Oregon High Cascades. *J. Volcanol. Geotherm. Res.* **339**, 1–22.

Porritt R. W., Allen R. M., Boyarko D. C. and Brudzinski M. R. (2011) Investigation of Cascadia segmentation with ambient noise tomography. *Earth Planet. Sci. Lett.* **309**, 67–76.

Reiners P. W., Hammond P. E., McKenna J. M. and Duncan R. A. (2000) Young basalts of the central Washington Cascades, flux melting of the mantle, and trace element signatures of primary arc magmas. *Contrib. Mineral. Petrol.* **138**, 249–264.

Richards J. P. and Kerrich R. (2007) Special paper: adakite-like rocks: their diverse origins and questionable role in metallogenesis. *Econ. Geol.* **102**, 537–576.

Righter K. (2000) A comparison of basaltic volcanism in the Cascades and western Mexico: compositional diversity in continental arcs. *Tectonophysics* **318**, 99–117.

Rowe M. C., Kent A. J. R. and Nielsen R. L. (2009) Subduction influence on oxygen fugacity and trace and volatile elements in basalts across the cascade volcanic arc. *J. Petrol.* **50**, 61–91.

Ruscitto D. M., Wallace P. J., Johnson E. R., Kent A. J. R. and Bindeman I. N. (2010) Volatile contents of mafic magmas from cinder cones in the Central Oregon High Cascades: implications

for magma formation and mantle conditions in a hot arc. *Earth Planet. Sci. Lett.* **298**, 153–161.

Sas M., Debari S. M., Clynne M. A. and Rusk B. G. (2017) Using mineral geochemistry to decipher slab, mantle, and crustal input in the generation of high-Mg andesites and basaltic andesites from the northern Cascade Arc. *Am. Mineral.* **102**, 948–965.

Sun S. and McDonough W. F. (1989) Chemical and isotopic systematics of oceanic basalts: implications for mantle composition and processes. Geol. Soc. London. *Spec. Publ.* **42**, 313–345.

Schmidt M. E., Grunder A. L. and Rowe M. C. (2008) Segmentation of the Cascade Arc as indicated by Sr and Nd isotopic variation among diverse primitive basalts. *Earth Planet. Sci. Lett.* **266**, 166–181.

Shaw A. M., Hilton D. R., Fischer T. P., Walker J. A. and Alvarado G. E. (2003) Contrasting He–C relationships in Nicaragua and Costa Rica: insights into C cycling through subduction zones. *Earth Planet. Sci. Lett.* **214**, 499–513.

Sherrod D. R. and Smith J. G. (1990) Quaternary extrusion rates of the Cascade Range, northwestern United States and southern British Columbia. *J. Geophys. Res. Solid Earth* **95**, 19465–19474.

Sisson T. W. and Layne G. D. (1993) H2O in basalt and basaltic andesite glass inclusions from four subduction-related volcanoes. *Earth Planet. Sci. Lett.* **117**, 619–635.

Stern C. R. and Kilian R. (1996) Role of the subducted slab, mantle wedge and continental crust in the generation of adakites from the Andean Austral Volcanic Zone. *Contrib. Mineral. Petrol.* **123**, 263–281.

Streck M. J. and Leeman W. P. (2018) Petrology of "Mt. Shasta" high-magnesian andesite (HMA): a product of multi-stage crustal assembly. *Am. Mineral.* **103**, 216–240.

Syracuse E. M. and Abers G. A. (2006) Global compilation of variations in slab depth beneath arc volcanoes and implications. *Geochem., Geophys. Geosyst.* **7**, n/a-n/a.

Syracuse E. M., van Keken P. E. and Abers G. A. (2010) The global range of subduction zone thermal models. *Phys. Earth Planet. Inter.* **183**, 73–90.

Taylor J. R. (John R. (1997) An introduction to error analysis : the study of uncertainties in physical measurements. University Science Books.

Thorkelson D. J. and Breitsprecher K. (2005) Partial melting of slab window margins: genesis of adakitic and non-adakitic magmas. *Lithos* **79**, 25–41.

Till C. B., Grove T. L., Carlson R. W., Donnelly-Nolan J. M., Fouch M. J., Wagner L. S. and Hart W. K. (2013) Depths and temperatures of <10.5 Ma mantle melting and the lithosphere-asthenosphere boundary below southern Oregon and northern California. *Geochem. Geophys. Geosyst.* **14**, 864–879.

Till C. B., Kent A. J. R., Abers G. A., Janiszewski H. A., Gaherty J. B. and Pitcher B. W. (2019) The causes of spatiotemporal variations in erupted fluxes and compositions along a volcanic arc. Nat. Commun.

Trehu A. M., Asudeh I., Brocher T. M., Luetgert J. H., Mooney W. D., Nabelek J. L. and Nakamura Y. (1994) Crustal Architecture of the Cascadia Forearc. Science (80-.). 266, 237–243.

Tryon C. A., Kuhn S. L., Slimak L., Logan M. A. V. and Balkan-Atlı N. (2011) Scale in tephrostratigraphic correlation: an example from Turkish Pleistocene archaeological sites. *Quat. Int.* **246**, 124–133.

Turner S. and Foden J. (2001) U, Th and Ra disequilibria, Sr, Nd and Pb isotope and trace element variations in Sunda arc lavas: predominance of a subducted sediment component. *Contrib. Mineral. Petrol.* **142**, 43–57.

Verplanck E. P. and Duncan R. A. (1987) Temporal variations in plate convergence and eruption rates in the Western Cascades, Oregon. *Tectonics* **6**, 197–209.

Walowski K. J., Wallace P. J., Hauri E. H., Wada I. and Clynne M. A. (2015) Slab melting beneath the Cascade Arc driven by dehydration of altered oceanic peridotite. *Nat. Geosci.* **8**, 404–408.

Wannamaker P. E., Evans R. L., Bedrosian P. A., Unsworth M. J., Maris V. and McGary R. S. (2014) Segmentation of plate coupling, fate of subduction fluids, and modes of arc magmatism in Cascadia, inferred from magnetotelluric resistivity. *Geochem. Geophys. Geosyst.* **15**, 4230–4253.

Wells R., Bukry D., Friedman R., Pyle D., Duncan R., Haeussler P. and Wooden J. (2014) Geologic history of Siletzia, a large igneous province in the Oregon and Washington Coast Range: correlation to the geomagnetic polarity time scale and implications for a long-lived Yellowstone hotspot. *Geosphere* **10**, 692–719.

Wells R. E. and McCaffrey R. (2013) Steady rotation of the Cascade arc. *Geology* **41**, 1027–1030.

Wells R. E., Weaver C. S. and Blakely R. J. (1998) Fore-arc migration in Cascadia and its neotectonic significance. *Geology* **26**, 759.

Whitford D. J., Nicholls I. A. and Taylor S. R. (1979) Spatial variations in the geochemistry of quaternary lavas across the Sunda arc in Java and Bali. *Contrib. Mineral. Petrol.* **70**, 341–356.

Wilson D. S. (2002) The Juan de Fuca plate and slab: Isochron structure and Cenozoic plate motions. In The Cascadia Subduction Zone and Related Subduction Systems Seismic Structure, Intraslab Earthquakes and Processes, and Earthquake Hazards. OFR 02-328 (eds. S. Kirby, K. Wang, and S. Dunlop). US Geological Survey. pp. 9–12.

WoldeGabriel G., Hart W. K. and Heiken G. (2005) Innovative tephra studies in the East African Rift System. *Eos, Trans. Am. Geophys. Union* **86**, 255.

Wörner G., Moorbath S., Horn S., Entenmann J., Harmon R. S., Davidson J. P. and Lopez-Escobar L. (1994) Large- and Fine-Scale Geochemical Variations Along the Andean Arc of Northern Chile (17.5°–22°S). In Tectonics of the Southern Central Andes (eds. P. D. K.-J. Reutter, D. E. Scheuber, and D. P. J. Wigger). Springer Berlin Heidelberg. pp. 77–92.

Associate editor: Rosemary Hickey-Vargas