

Peer Review Report

Review Report on From Continent to Ocean: investigating the multi-element and precious metal geochemistry of the Paraná-Etendeka Large Igneous Province using machine learning tools

Original Research, Earth Sci. Syst. Soc.

Reviewer: Kim Cone

Submitted on: 01 Jun 2021

Article DOI: 10.3389/esss.2021.10039

EVALUATION

Q 1 Please summarize the main findings of the study.

The authors use three different MLA techniques on an elemental data set of LIP lavas to ascertain if plume interaction with the rifting SCLM differentially extracts metals over space and time. A comparison is made with the NAIP (and in disagreement). The main findings: 1) geochemical MLA patterns appear to support previously published geochemical findings in this PELIP region with perhaps a margin of S and CN types into a single group, as previously reported; 2) the authors conclude that plume character itself is insufficient to produce patterns of absentee PGEs (the asymmetry); rather, SCLM metal content is key, and the thinning of the SCLM along with plume-induced melting has a strong effect on the appearance of certain metals in the PELIP lavas over space and time.

Q 2 Please highlight the limitations and strengths.

Strengths: the use of MLA techniques in an attempt to group and categorize lavas in the PELIP to support our current understanding of metal-bearing lavas. However, there is little mentioned about the shortcomings and limitations of the methods used. I have questions as to the significance of clustering z-scores with very small sample sizes. Sample sizes for each type were not explicitly shown in the main text, and a reduction in sample size was then revealed later on. The significance of clustering PCA results (how is variance-based clustering more insightful than clustering raw data?) is not discussed. Sample size bias? See reviewer notes.

Q 3 Please comment on the methods, results and data interpretation. If there are any objective errors, or if the conclusions are not supported, you should detail your concerns.

Abstract

Line 22: Define SCLM here, not in line 27.

Line 27: Define “this”—watch the demonstrative pronouns in general, make sure the reader knows exactly to what they refer. Same in line 40.

Line 33: “Onshore lavas on the Etendeka side [Of the province? Of the PELIP?] are...”. For example, this is better stated in line 37 for the Paraná side.

Line 42-43: I’ll expect there is eventually an explanation of what “similarities” are being referred to here. Structural and/or geochemical? Geodynamic?

Introduction

Line 58: You mean, in a geodynamic context?

Line 96: Please reword this “challenging” phrase (awkwardly posed), but I do see what you’re getting at here (something to the effect of “This method allows us to explore ME trends and relationships that typically go undetected using traditional geochemical two-variable plots, etc. ...” . If anything, I’d say that multivariate approaches in this context (geochemistry/petrology) produce results that are otherwise obscured when using only bi-variable analyses.

116: “It” refers to what exactly?

Fig 1a: This is probably the oddest plume geometry I’ve ever seen as a schematic – why is the base of the plume tail so wide? Indicate the SCLM thinning zone?

Fig 1b: The same plume has undergone an aspect ratio change, is there a change in scale for the figure? The thinned SCLM under the spreading Atlantic Ridge is now a light blue and has separate boundaries, this should be marked/indicated/labeled on the figure, since it’s shown as a separate structure.

Line 133: plate movement direction... would be helpful to add this to a current figure/map. This would be perfect to add to Figure 2 (or perhaps even 1b).

Line 145: I’m going to pick on the header here. There is really no significant petrology or petrological descriptions or mineralogy in this section, outside of the general classifications of “basaltic and basaltic-andesitic lavas.” Otherwise, consider referring the reader to SA2 for the short petrological descriptions.

Line 147: “onto” not “on to”

Line 157: There is no prior context for Pitanga and Paranaparema. This should be briefly explained/depicted.

Line 158: What is “Ribeira” here? Again, the term has not been yet defined.

Line 233: (note to self: PELIP, initial n = 134:

Parana (n=102)

Etendeka (n=10)

Rio Grande Rise and wavis Ridge (n = 22) – core #s provided, on Fig. 2.

All sample locations are marked on a map (Fig. 2)–okay

Line 233: Regarding Table 1, I think it’s fair to assume the reader has a basic understanding of stats, and so in the name of transparency, the number of samples per each Leg should be shown for the Deep Sea cores. I have no idea if one Leg has only one sample (Tables 1 and 2 could actually have been merged into one single comprehensive table). Figure 5 would make more sense knowing this.

Materials and Methods

Line 249-258: The point of cutting the larger samples down... you are testing for the degree of homogeneity within the larger sized samples, correct? Are the samples distally located within a single larger rock? Were these smaller sections randomly selected? I see the data in Supplementary A1 but not a written description that would make it explicitly clear (this should be explained in the methods section regardless). Were these geochemical analyses then averaged?

Section 2.3 Machine learning workflow

In general, I’m good with the reader being directed to a previously published work for a more detailed workflow explanation, but some of the details are lost in the process. The

basics of each MLA method are adequately explained, but the authors don't state on what dataset the methods are used (original versus dimensionally-reduced data) until later.

Lines 294-295, 311-312: I think you're trying to get at "global versus local" data structure optimization? PCA is all about capturing global structure (and sensitive to outliers) whereas t-SNE is insensitive to outliers because it really is about optimizing local data clustering (provided that perplexity values are relatively large versus the sample set). The two MLA methods are complementary in this way. Consider explicitly stating this somewhere for those who may be alienated by MLA/data science terms and concepts.

Lines 319-322: To be clear, you are running k-means on the PCA results, not on the original data as they are structured, correct?

Line 342: 116 samples are eventually used for all analyses here, yes? So it's no longer $n = 134$. Why not state this up front earlier in the text? Were the other 18 samples used in any way?

Line 347: As soon as "z-score" is mentioned, standard deviations should be documented somewhere.

(Question for authors: Did you check the distributions of your raw geochemical data, pre-normalization or pre-standardization? Even if PCA is non-parametric...)

Results

Line 360: Please include the number of samples per locality. I'm surprised there is no st.dev. reported here.

Lines 391-393: There is a strong co-trending variable group that define PC1: Zr/Hf/Rb/K₂O, for example, which anti-correlates with CaO along PC1. Are significances of anti-correlative trends addressed in the discussion?

Regarding the PCA biplots. If the scree plot shows most of the variance is tied to PCs 1, 2, and 3, why not plot PC1-2 (which you have), PC2-3, PC1-3? Why even plot PC7-8?

Lines 401 and on, authors' response requested: the loadings here (PC1-2, for example) are reflective of the input from the global data set, and it is reasonable to expect that sample Types (assuming geochemical similarity within a type) lie as dense clusters that control loading geometry/eigenvector orientation and length. In PC1-2, it appears that SG Type 4 strongly contributes to the variance in PC1 for the global dataset through a wide range in values of Zr, Hf, Rb and K₂O. But how is it that these elements are on a single eigenvector? Were these elements paired to reduce redundancy (typically each element would have its own eigenvector)?

Does it make sense to state that Type 4 variability itself is governed by the nearby loading scores and the elements tied to them (which is what you're saying)?

Line 423-424: 5000 iterations because... was convergence already reached or is this an arbitrary value?

Lines 428-436: Since t-SNE alters the original data cluster densities, it would have been helpful to previously provide st.devs. in one of the tables. It helps the reader to see how original cluster density may have been altered through t-SNE (the authors hint at cluster size here and structure; this is why I bring this up).

Lines 437-440: This approach may lose significance (or introduces a lot of bias) when very small sample sizes are employed for some of the locales. This must be discussed in the discussion section.

Line 461 on: Briefly statements on the significance of clustering z-scores (versus original data) and PCA data I assume are somewhere in the discussion. From reading the Methods section earlier, one could not tell if k-means was going to be applied to z-scores or raw data.

Line 467: Remove footnotes from the Results section. I believe this source has already been mentioned in Methods.

Line 475: With respect to "sample locality"... in what context? In the t-SNE figure? Geographically?

Figure 8c. Explain the logic behind assigning k-means groups (8b) to the same locale (8a and 8c), particularly because of the effects of sample size bias. Why not assign the larger sized group to the locale of the smaller sized group?

Line 499: awkward sentence

Lines 508-521: same comment as above for line 401 and on.

Lines 551-552: Discuss this procedure briefly. How robust are the z-score clusters when sample sizes are less than 10 for some locales? Would it matter?

Discussion

Line 542: Are you certain you mean 4B and 4C (or are you referring to figure 3)? Linear trends by themselves (they look more like clusters), or as part of a larger sequence (as part of the SG types)? Again, I'm not sure what figure/s are being referenced.

Line 557: To say that SG Type 4 samples show "affinity" for the strength of the TiO₂ vector is misleading, as they literally and predominantly create and define the TiO₂ loading score itself.

Lines 570-573: But doesn't t-SNE have the tendency to increase the density of sparse clusters and decrease the density of originally dense clusters? Other clusters here are adjacent as well. How is the merging of S and C-N types in to one type supported by PCA?

Lines 594-597: But you have a not-so-insignificant cluster of RGR samples in group 3 as well.

Line 630: content

Lines 636-639: Yes, I see this from the raw data, but how did MLAs provide insight outside of what has been already published/known about the geochemical character in the region?

Line 1491-1492: Plume focus is shown in red?... I see red solid lines (circles) and red dashed lines, so it's not clear. You may want to clarify in the caption.

SA1, Combined-Full: "dupl"? What is the significance of this? I'm trying to find *exactly* which samples you used so that your results can be replicated.

Summ_Stats: Add the number of samples ($n = x$) in the header. It's been challenging to get a grasp of how many samples per type were actually considered/categorized per spreadsheet page, as it's only explicitly mentioned in the main text once up until referring the reader to look at SA1 and SA2.

Q 4 Check List

Is the English language of sufficient quality?

Yes.

Is the quality of the figures and tables satisfactory?

No.

Does the reference list cover the relevant literature adequately and in an unbiased manner?

Yes.

Are the statistical methods valid and correctly applied? (e.g. sample size, choice of test)

No.

If relevant, are the methods sufficiently documented to allow replication studies?

No.

Are the data underlying the study available in either the article, supplement, or deposited in a repository? (Sequence/expression data, protein/molecule characterizations, annotations, and taxonomy data are required to be deposited in public repositories prior to publication)

Yes.

Does the study adhere to ethical standards including ethics committee approval and consent procedure?

Not Applicable.

If relevant, have standard biosecurity and institutional safety procedures been adhered to?

Not Applicable.

Q 5 Please provide your detailed review report to the editor and authors (including any comments on the Q4 Check List):

The biggest issue with this manuscript is that it appears that the conclusions drawn in the discussion can be done without MLA analyses. MLAs here do support the current literature and findings, but these authors suggested (at least once) that the literature only attributes geochemical and metal characteristics of the PELIP lavas to the plume character itself, not to the thinning SCLM. I don't know if this is the case. I have some issues with the PCA bi-plots: typically each variable (here, element) produces a single eigenvector (the lines on the plots). Here, multiple elements form a single eigenvector. It seems like the authors perhaps combined elements that highly correlated, in order to reduce redundancy. If so, that's great, but it needs to be stated. There needs to be a lot more transparency about the number of samples per locale, and how that number dwindled from the Methods section to the Results section. The discussion brings up many

legitimate points, but again, it seems that these points can be brought up without MLAs. I'm not sure I see any truly unique results or conclusions that derive strictly from MLAs. In terms of k-means clustering, the authors don't state why they cluster the results of PCA versus clustering the raw data-- they are choosing to cluster based on variance of geochemical character within each locale and then through clustering this way, decide to move certain samples from certain locales to other locales for their new groups. This doesn't make sense, and I have asked the authors to explain in the event I missed the reason for their grouping logic. The authors refer the reader to a previously published paper for the order of MLAs used and the general logic behind them, but to that paper's **Python code** for details (needless to say, there should be some basic mathematics/equations provided, not just "refer to the Python algorithm"). I suggest a rewrite or rejection.

QUALITY ASSESSMENT

Q 6	Originality	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q 7	Rigor	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q 8	Significance to the field	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q 9	Interest to a general audience	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q 10	Quality of the writing	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q 11	overall quality of the study	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

REVISION LEVEL

Q 12 what is the level of revision required based on your comments:

substantial revisions.