

## Additional comments on ESSD-2019-154 (Vulcan US emissions)

I appreciate that our research community focuses on emissions products. Definitely we need those efforts, globally and nationally! I also like that ESSD plays a helpful role in promoting, certifying and sharing those products.

Unfortunately, for several reasons, this most-recent Vulcan product as submitted and as described fails to meet many requirements and expectations for ESSD. In several comments below I echo and emphasize points made by reviewer #2. As chief editor for ESSD for more than 10 years, I hope I offer useful and well-informed viewpoint.

Note: I read both this paper and the prior (Hestia) publication in ESSD.

Overall, I find very little about data, methods, validation, etc. to give a reader / user confidence. I find the registration requirement at ORNL unacceptable. I repeat, in the strongest terms, the recommendation of reviewer #2: read the guidelines!!! (<https://www.earth-syst-sci-data.net/10/2275/2018/>)

We will attempt to respond to specific comments regarding the data, methods and validation in order to increase user confidence.

Extensive discussion with both ORNL and the editor with whom we worked at ESSD led us to believe that the ORNL DAAC had availability and access rules consistent with those expected and outlined in the guidelines - which we did read (ps. the link provided by the reviewer does not land on a page related to guidelines, worth noting when communicating with authors in the future). We followed:

"Copernicus Publications requests that all ESSD authors deposit their data corresponding to journal articles in reliable (public) data repositories, assign digital object identifiers, and properly cite data sets as individual contributions. Please find your appropriate data repository in the list of data centres supporting the ESSD criteria or consult with ESSD editors."

Many weeks were devoted to discussion on this topic. We thought we had reached agreement on the issue of access. I just email the ORNL DAAC and they do require a sign-in. This is the text of the email they sent in response to my asking about the sign-in procedure:

"Access to all NASA Earth Science data requires a user registration through Earthdata Login. There have been multiple discussions with multiple journals about this in the past, including Kirsten Elger (ESSD Editor in Chief). Apart from requiring that login, there is no restriction on access or use of the data. The information a user provides in the registration process is only used in aggregate to understand the distribution of users."

I've spoken with Kirsten in the past, including at the recent AGU meeting. I suggest the editor discuss the matter with her and one or both of them can reach out to me if my answers and past discussions with Kirsten do not resolve the matter.

See this text on the Earthdata Login page (<https://urs.earthdata.nasa.gov/>).

Why must I register?

The Earthdata Login provides a single mechanism for user registration and profile management for all EOSDIS system components (DAACs, Tools, Services). Your Earthdata login also helps the EOSDIS program better understand the usage of EOSDIS services to improve user experience through customization of tools and improvement of services. EOSDIS data are openly available to all and free of charge except where governed by international agreements.

This is not something where we have control. This is defined at the NASA Earth Sciences Division level and applies to all NASA Earth Science Data."

Please understand my frustration at having spent weeks discussing this topic with the assigned editor at ESSD and the ORNL DAAC upon which we assumed we had agreement only to find that there is not agreement. If a sign-in remains a barrier we will have to withdraw our paper and submit elsewhere. This DAAC is strongly encouraged by our funders as this project is under the Carbon Monitoring System umbrella.

Reader has no idea of sequence or versions. Hestia based on Vulcan? Vulcan builds on lessons learned from Hestia? This version, apparently Vulcan 3.0, improves on which prior version(s)? What improvements? How do the authors confirm improvements? Does the reader need to go back to Nature papers in 2002?

We are not clear on what relevance Hestia, a separate project, has to do with the current manuscript? Hestia is mentioned once in the current manuscript as follows:

"Corrections to location information were made in urban domains associated with the Hestia Project : the Los Angeles Basin, Baltimore, Salt Lake City, and Indianapolis (e.g. Gurney et al., 2018; 2019b.)"

This was only to point out that location information was corrected in the four isolated geographic locations through the Hestia project and the relevant citations made. Hestia is not based on Vulcan – nowhere in the manuscript has any relationships been suggested other than these location corrections. We do not understand why a discussion of Hestia is necessary for this manuscript?

The Nature 2002 paper has nothing to do with Vulcan, Hestia, or inventories at all. It is a paper on atmospheric CO<sub>2</sub> inversions at the global scale, included to offer the reader a foundational paper on inversions and how they use bottom-up gridded emissions information. The reviewer must have a fundamental misunderstanding.

The need to note and or otherwise reference previous versions of Vulcan is difficult: at the time of its publication (2009), standardized data journals such as ESSD were less common and standard data practices were not as common as they are now. Hence, it was published in a journal that did not have the requirements current data journals have. The documentation was available but not in the same way or form we are attempting to do so presently. Hence, there is no standardized pedigree to point to. A solution is to write the Vulcan version 3.0 as standalone and put all documentation into this paper such that every detail is included. Assuming we proceed with ESSD, we will confer with the reviewer or contact editors on this topic.

What external sources? How accessible? How reliable? Near the end the authors write "depending upon support and the availability of data sources described in this study". Will availability prove problematic? For all sources? Specific sources?

We are not entirely clear what the reviewer is referring to with regards to the questions "What external sources? How accessible? How reliable?" We take this as a general query regarding the input data used to generate the Vulcan estimate. Regarding the "external sources" - the data sources are provided in the manuscript – all of which are available through the citations and URLs provided. All input data is identified – there are no other sources.

Regarding the quoted statement - this was included because, as was noted in the manuscript text, the input data to the Vulcan data product is primarily "regulatory" data (all publicly available). The timing of the release of these input data varies with considerable latency. Furthermore, the data structures and data fields have varied in the past and hence, cannot be considered absolutely guaranteed in form indefinitely. The accessibility and reliability are difficult for us to comment on – US government agencies can change data policies or alter collection procedures. Hence, we felt that this caveat was important to include so as the reader is aware that continued temporal extension (beyond the current 2010-2015 time period) or version updates cannot be guaranteed in the future (the current version 3.0 can be, however). The allusion to support also make clear that the magnitude of this work cannot be extended or improved without continued support from funding agencies.

It may be best to remove this statement as it would be difficult to ascertain which data sources have guaranteed availability into the future versus others.

Manuscript needs two tables. First, a clear sequence of prior and related products leading to Vulcan 3. Do not make readers guess or search. Provide reliable up-to-date links. If not open access, make them (all) open access. Second, a clear comprehensive list, in table form, of all sources. Perhaps 50 or more, no matter. Let readers know what you used, what version you started from, with active certified links to all sources. Most users do not want to try to follow every step, but authors must nonetheless provide exact guidance and source information.

We will add the requested information to the manuscript and hope it satisfies the needs of the reader. As with previous comments here, we will consider writing Vulcan 3.0 as a standalone data product in which all information regarding its construction and inputs are presented here avoiding any reference to previous versions.

This reader / editor does not like the "Vulcan'Science'Methods'Documentation,'Version'2.0" It gives no information about date or version. In too many cases it appears to derive from 2002. That .pdf has undergone no review, no critical reading, etc.; it looks like a lab report. Users will not find it useful or reliable. I also read at least partial overlap between manuscript and lab report. If useful, put it all in the manuscript.

We will remove reference to the Vulcan 2.0 documentation. We have added additional relevant details from the earlier documentation to the manuscript. Our aim is for the Vulcan 3.0 to stand alone as a data product without reference to the previous version.

All links are up-to-date. It is hard to establish their reliability as they are data supplied by regulatory/government agencies. They are active and come with no barriers of any kind to their free and open use. They will be double-checked.

Speaking of Hestia, I find only one mention accompanied by a single citation. Why, in Figure 10, do they not show an LA example? No discussion, intercomparison, etc. Meanwhile, Hestia clearly specifies "Hestia-LA data product are supplied by output of the Vulcan Project" (<https://doi.org/10.5194/essd-11-1309-2019>) Do the author intend a series of Hestia-cityname products, apparently isolated from Vulcan manuscripts? Also apparently substantial text overlaps between that document and this? What did our similarity test show?

The Hestia Project is a different effort and does not constitute any inputs to the Vulcan system. Hestia is mentioned once in the manuscript with two (not one) citations. There was no particular reason to show the Los Angeles domain in figure 10. Hestia performs additional space/time conditioning beyond what the Vulcan Project produces. They are numerically consistent at aggregate spatial scales (which varies by sector) hence, a comparison between Vulcan and Hestia is of marginal value. We will pursue this in the future but we do not see why Hestia is relevant to this manuscript or why a comparison between Hestia and Vulcan would be considered relevant.

We cannot find the quoted text in this manuscript so it is difficult to respond to aspects of the question other than to note that Hestia can be thought of as a nested effort within Vulcan – it zooms to finer detail with Vulcan as constraint at the aggregate scale. The Hestia project will continue and may add cities. They will use the Vulcan output and cite the Vulcan version 3.0 paper (wherever it is published) as the input. They will be written separately as they employ methods relevant to Hestia. The overlap in text on the LA paper is merely to provide context within the Hestia papers – sensible given that they derive much from the Vulcan effort. However, the reverse does not occur.

I find efforts to compare with ODIAC unsatisfactory. What does figure 9 tell me, quantitatively and reliably? Nothing. Because authors have not provided confidence trail for development of Vulcan 3, a snapshot comparison to ODIAC proves meaningless. What product works well for what purposes, with what uncertainties, and why? Can we assign differences to night light data in one but not the other? By how much would that impact? Which product needs what improvements? As a data journal, ESSD needs to ensure readers can ask and answer such questions. How does this work contribute to that discussion?

We respectfully disagree. Fossil fuel CO<sub>2</sub> emissions have no reference for validation – ODIAC does not (nor did it present such documentation in the ODIAC ESSD paper) nor do ACES, EDGAR, FFDAS, CDIAC etc. Hence, comparisons offer the only opportunity to better understand the respective data products, something difficult to do in isolation where there is no validation mechanism (yet). As we stated in the manuscript:

“we perform comparison to the ODIAC output over the Vulcan domain in the hope of providing insight into one or both of the emission estimates”

The lower threshold cutoff is evident in both figure 8 and figure 9 and is an important distinction that is driven by the use of nightlights in ODIAC. There are four paragraphs on the comparison with difference metrics. Figure 9 shows that there is a wider range of per gridcell emissions in Vulcan compared to ODIAC with greater magnitude differences at the lower-emitting gridcell end of the range (likely due to the low-end threshold cutoff in nightlights). The differences are relevant as they point to the very different spatial content of a bottom-up versus a downscale or proxy distribution approach, a distinction that the community must consider as they use these data products in various contexts and applications.

I find too many proprietary tools and legend errors. Figure 5: Copyright © National Geographic Society. Figure 6 (where authors have reversed the line colors: red designates months (not years) while purple designates years (not months): Copyright © Esri. Figure 10: ArcMapTM by Esri using the World Imagery basemap layer (Copyright © Esri). Proprietary copyrighted tools and sources are not acceptable in ESSD!

The one legend error in Figure 6 will be corrected.

As for the reference to tools in three of the 10 figures in the manuscript we had iterated with the editor on this topic as well and thought the final drafting was acceptable. Indeed, we initially had not included that copyright attribution on these figures but were asked to include them. The copyright only refers to the “basemap” NOT the data overlaid on the basemap (the Vulcan emissions are the data in these cases). In examining other papers in ESSD, basemaps are found throughout without copyright attribution. Hence, perhaps there is misunderstanding or confusion regarding the requirements?

Other reviewers have pointed out many technical errors. I find the entire manuscript unreliable with key information hidden or unaccessible.

We are responding to the comments made by other reviewers in separate response files.

We do not know how to respond to the general comment of unreliability, hidden/unaccessible information other than to attend to specific comments along these lines and review the manuscript for missing details.

With many months of effort and persistent cooperation by authors and reviewers, ESSD helped those authors bring the EDGAR product into a successful published product. I believe ESSD can and should do the same here, with the first step being to recognize and acknowledge current substantial deficiencies. ESSD = open access. Vulcan = not (yet).

We can only assume the last comment refers to the issues at the ORNL DAAC? We have commented previously on this.