We would like to thank all reviewers for their encouraging comments, which are not listed here for brevity. Here, we focus on comments requiring edits and/or answers. We would also like to emphasize that the reviewers surely spent time to help us with their specific and constructive comments; this is very much appreciated.

Reviewer #1 *(Antony Borel, 06 Aug 2022 15:24)*

Comment #1

The title focuses on dental microwear analysis only while a stone tool dataset is used. I suggest modifying the title to include this aspect of the paper and ensure better visibility for stone tool wear analysts as well.

Response to comment #1

This is very true. We have changed the title to: "Surface texture analysis in Toothfrax and MountainsMap® SSFA module: Different software packages, different results?"

Comment #2

Line 220: you used a polynomial of degree 2 to remove form. Could you please briefly mention why you choose such polynomial? Is the decision made on the basis of a visual inspection of the surfaces or was another means used?

Response to comment #2

This choice was largely arbitrary. Based on a visual inspection of some of the surfaces, a polynomial of degree 2 seemed to be good enough to remove the form. The output from a polynomial of degree 3 was very similar. This is one of the setting that might need more testing. Schulz et al. 2013 (https://doi.org/10.1016/j.wear.2013.01.115) and Arman et al. 2016 (https://doi.org/10.1088/2051-672X/4/2/024007) also used a degree 2 and we generally follow the latter workflow. However, Francisco et al. 2018 (https://doi.org/10.1088/2051-672X/aa9dd3) found that a degree 8 gives the best results.

Digital Surf advises not to go beyond a degree 4 when precision is critical because numerical errors propagate quickly with the higher degrees.

See also response to comment #3.

Comment #3

Line 22 and 223: here also I would suggest explaining a bit more the choice of the values of the parameters, this would help others to make an educated choice of these values with their own surfaces. What does “soft” refer to? What threshold is it exactly? This may be clear for MountainsMap users but are likely to be confusing for others.

Response to comment #3

This is a good idea. Section 2.2.1 has been extended (now lines 198-214):

"Most settings were chosen following Arman et al. (2016)'s workflow.

(1) Remove form with a polynomial of degree 2. The degree 2 was chosen following Schulz et al. (2013) and based on the visual inspection of some of the surfaces. Francisco et al. (2018) found that a degree 8 gave the best results. However, here we follow Schulz et al. (2013)'s and Arman et al. (2016)'s workflows by applying a degree 2, and recommend more testing for future studies.

(2) Level with a least-squares plane by subtraction. According to MountainsMap's reference guide, the least-squares plane "method is recommended for surfaces with random surface texture", which is the case here. Subtraction is recommended for surfaces with low slopes and ensures that the leveling keeps the XY spacing of the points (= digital lateral resolution) constant.

(3) Remove outliers (maximum slope 80°, measurement noise removed). The maximum slope of 80° was chosen in relation to the numerical apertures of the objectives (see below). The setting for the strength of the method is not used when removing outliers based on the maximum slope. The option to remove measurement noise was selected because no other filter was applied for this purpose.

(4) Threshold 0.1-99.9%. Most outliers are likely to be removed by the previous operator, but some points with extreme heights that are likely to be measurement artifacts might remain.

(5) Fill-in non-measured points (smooth shape from neighbors). It seems that MountainsMap and Toothfrax handle non-measured points differently, so we chose to avoid this confounding factor."

Comment #4

Line 263-265: you mention an erroneous calculation of Asfc when Smfc is very high. During my first reading I was expecting a bit more explanation here but I found it later in the discussion. Maybe you can already refer to the discussion here.

Response to comment #4

Changed as recommended (now lines 249-250).

Comment #5

Line 279: R^2 is the coefficient of determination not the regression coefficient except if it has a specific meaning for MountainsMap. Also, based on the details given and what could find in the MountainsMap documentation I do not know what kind of regression model is used to fit the curve of relative area but if nonlinear model is used then R^2 may not be reliable as quality indicator (see for example Spiess, A.-N., Neumeyer, N., 2010. An evaluation of R^2 as an inadequate measure for nonlinear models in pharmacological and biochemical research: a Monte Carlo approach. BMC Pharmacology. 10, 6.). Sometimes software compute R^2 in non-relevant cases, it is why I mention this. I believe DigitalSurf

paid attention to this but please verify it. If you find that nonlinear model is used then I suggest omitting the R^2 in this paper as, as you mentioned, "it is usually not used in dietary or functional interpretations" anyway.
Line 585-586: from this sentence I understand that polynomials of order 2 and 1 are used. Is that correct? If yes, then I have the answer to my previous comment of line 279 about R^2: my comment is not relevant and R^2 can be used here.

Response to comment #5
This is indeed the case that different orders for the regression model are used in MountainsMap and Toothfrax.
We have renamed R^2 to coefficient of determination instead of regression coefficient, as recommended (now line 262).

Comment #6
Line 410: in the "ThreeFactor_filter_strong_contrast_ConfoMap_Toothfrax_epLsar.pdf" graph, the labels of the x axis are overlapping each other, you may consider editing this using smaller font size for example.
Also, the file "treatment_pairs_NewEplsar.pdf" is not named as the other similar files. I think it should be "NewEplsar_filter_weak_treatment_pairs_NewEplsar.pdf" to standardize the names of the files.

Response to comment #6
We have not changed the size of the x-axis labels for the plot in "ThreeFactor_filter_strong_contrast_ConfoMap_Toothfrax_epLsar.pdf" in the research compendium (GitHub/Zenodo) because the plot is shown in Fig. 4b, where it has been edited for a better display.
Regarding the inconsistent file name, the corresponding notebook (rendered to HTML and MD files) has been edited accordingly. The plot is now exported to the file "NewEplsar_filter_weak_treatment_pairs_NewEplsar.pdf" to follow our file naming convention.

Comment #7
Line 442: Fig. 4 may become Fig. 1 to number the figure in the order of their occurrence.

Response to comment #7
We prefer to have the boxplots, which present the raw data, before the results of the Bayesian analyses. So we have reworded to (now lines 405-407):
"[…] *Asfc*, *epLsar*, *HAsfc9* and R^2 are significantly different when calculated with the two software packages (i.e. 95 % HDIs do not include 0; Figs 1-4)."

Comment #8

Line 452: you write “HAsfc81 (sheep dataset) and Smfc follow the trends mentioned above (compare files of the "filterstrong" and "filterweak" runs for the three-factor model in the folder "Pythonanalysis/" on Zenodo)”. I cannot find any plot showing HAsfc81 in the indicated folder and Smfc is only available in “ThreeFactorfilterstrongpriorpredicitiveSmfc.pdf” and “ThreeFactorfilterweakpriorpredicitiveSmfc.pdf”. If these two parameters were not analyzed as mentioned before for surfaces with less than 5% NMP, I think the parentheses should be moved after “Asfc, epLsar, HAsfc9 and R2 are again significantly different when calculated with the two software packages” to avoid confusion.

Response to comment #8

Changed as recommended (now lines 439-442).

Comment #9

Line 534: I suggest editing the part concerning “significance” in table 3. Like for the CI, separating clearly “MountainsMap NewEpLsar”, “MountainsMap epLsar” and “Toothfrax epLsar” would make the table more easily readable.

Response to comment #9

Changed as recommended. Table 2 has been edited too to better match the new layout of Table 3.

Comment #10

Line 698-700: “We therefore recommend re-analyzing raw surface data with MountainsMap (or any of its derivatives) before comparing with published Toothfrax data”. As you identified what are the causes of the differences (in most cases), don’t you think that it would be possible to apply a correction to convert Toothfrax parameters values into MountainsMap parameters values? This would allow using also previously published data for which raw surfaces are not available.

Response to comment #10

It is likely that we have not identified all the causes of the differences. And even for the causes we have identified, there are probably surface-specific effects. Additionally, the iterative calculations make it very difficult to find a simple formula that can be applied to convert Toothfrax values to MountainsMap values. It therefore seems that re-analyzing surfaces is unavoidable.

Comment #11

Line 740: “This research is publication no. XX of the TraCEr laboratory”. This sentence has to be completed (or removed).

Response to comment #11

We waited until now to add the publication number. This is publication number 11 (now lines 666).

Comment #12

This has no implication for the paper as other ways of opening and reproducing the data are working but Binder using the binder badge from GitHub gives an error. I am not really familiar with Binder so I have not searched for the solution yet (I do not know if it comes from my local environment). Here is the errors from the log file in case it would be necessary:

Error installing renv:
==========
Error: ERROR: cannot cd to directory '/home/rstudio/renv/library/R-4.0/x8664-pc-linux-gnu'
Warning messages:
1: In system2(r, args, stdout = TRUE, stderr = TRUE) :
running command "/usr/local/lib/R/bin/R' --vanilla CMD INSTALL -l '/home/rstudio/renv/library/R-4.0/x8664-pc-linux-gnu' '/tmp/RtmpfQFzlM/renv0.14.0.tar.gz' 2>&1' had status 1
2: Failed to find an renv installation: the project will not be loaded.
Error: (converted from warning) packages 'rrtools', 'holepunch' are not available (for R version 4.0.2)
Execution halted
Removing intermediate container 4509d34e8ea1
The command '/bin/sh -c wget https://github.com/tracer-monrepos/SSFAcomparisonPaper/raw/master/DESCRIPTION && R -e "options(repos = list(CRAN = 'http://mran.revolutionanalytics.com/snapshot/2021-01-26/')); devtools::install_deps()"' returned a non-zero code: 1Built image, launching...
Failed to connect to event stream

Response to comment #12

Unfortunately, we could not find the source of the error. We will try again for the next research compendium, but we have preferred to delete the binder badge and dockerfile (and mentions of it in the README and in the main text) for this research compendium. Other options to run the code locally remain valid.

Reviewer #2 *(anonymous reviewer, 25 Aug 2022 07:40)*

Comment #1

Lines 68-69: It would be useful to revise the description of what DMTA is quantifying since surface microasperities are also influenced by tooth-to-tooth contact and corrosion in addition to food and abrasive particles. I have listed several references below that examine these forms of wear using DMTA. It is not necessary for the authors to cite all these papers, but I did want to provide some support for my suggestion.

- Hara, A.T., S.V. Livengood, F. Lippert, G.J. Eckert, and P.S. Ungar. 2016. "Dental surface texture characterization based on erosive tooth wear processes." Journal of Dental Research 95 (5):537-542. doi: 10.1177/0022034516629941.
- Hara, A. T., D. Elkington-Stauss, P. S. Ungar, F. Lippert, G. J. Eckert, and D. T. Zero. 2021. "Three-Dimensional Surface Texture Characterization of In Situ Simulated Erosive Tooth Wear." Journal of Dental Research 100 (11):1236-1242. doi: 10.1177/00220345211005678.
- Krueger, K. L., E. Chwa, A. S. Peterson, J. C. Willman, A. Fok, B. van Heel, Y. Heo, M. Weston, and R. DeLong. 2021. "Technical note: Artificial Resynthesis Technology for the experimental formation of dental microwear textures." American Journal of Physical Anthropology 176 (4):703-712. doi: https://doi.org/10.1002/ajpa.24395.
- Ranjitkar, S, A Turan, C Mann, GA Gully, M Marsman, S Edwards, JA Kaidonis, C Hall, D Lekkas, and P Wetselaar. 2017. "Surface-sensitive microwear texture analysis of attrition and erosion." Journal of Dental Research 96 (3):300-307. doi: https://doi.org/10.1177%2F0022034516680585.

Response to comment #1

These aspects are indeed also important. The references have been added to the introduction (now lines 64-66):

"DMTA quantifies the micro-topography on tooth surfaces resulting from abrasion, attrition and erosion due to the contact with food items and other abrasive particles such as dust or grit during comminution (Hara et al., 2016, 2021; Ranjitkar et al., 2017; Krueger et al., 2021)."

Comment #2

Lines 84-87: While I agree that many researchers have been using only complexity and anisotropy for dietary reconstruction on postcanine teeth, I would like to note that numerous studies examining DMTA in anterior teeth have highlighted Tfv as a particularly useful variable to discriminate between various behavioral regimes. Since the lack of Tfv is still a limitation in the Mountains Map software, the authors may not find this to be a useful critique. However, it does point out that there are essential modifications that still need to occur with Mountains Map before it can totally replace the Toothfrax/Sfrax for certain research purposes. I address this in the CONCLUSIONS as well. Again, below is a representative sample of some of the work I am referring to, but the authors do not need to cite each paper if they choose to address this point.

- Krueger, K. L., and P. S. Ungar. 2009. "Incisor microwear textures of five bioarcheological groups." International Journal of Osteoarchaeology 20 (5):549-560. doi: http://dx.doi.org/10.1002/oa.1093.
- Krueger, K. L. 2015. "Reconstructing diet and behavior in bioarchaeological groups using incisor microwear texture analysis." Journal of Archaeological Science: Reports 1:29-37. doi: http://dx.doi.org/10.1016/j.jasrep.2014.10.002.
- Krueger, K. L., P. S. Ungar, D. Guatelli-Steinberg, J.-J. Hublin, A. Pérez-Pérez, E. Trinkaus, and J. C. Willman. 2017. "Anterior dental microwear textures show habitat-driven variability in Neandertal behavior." Journal of Human Evolution 105:13-23. doi: http://dx.doi.org/10.1016/j.jhevol.2017.01.004.
- Krueger, K. L., J. C. Willman, G. J. Matthews, J.-J. Hublin, and A. Pérez-Pérez. 2019. "Anterior tooth-use behaviors among early modern humans and Neandertals." PLOS ONE 14 (11):e0224573. doi: https://doi.org/10.1371/journal.pone.0224573.

Response to comment #2

Thank you for these references. This is a very good point that we overlooked.
This is now mentioned in the introduction (now lines 82-85):
"However, analyses of incisor microtextures have found that Tfv is an important parameter related to anterior tooth loading and tooth use (Krueger & Ungar, 2010; Krueger, 2015; Delezene, Teaford & Ungar, 2016; Krueger et al., 2017, 2019; see also Caporale & Ungar, 2016 on rodents)."
See also response to comment #4.

Comment #3

Line 524 (typo): I think an “a” should be inserted before “results”.

Response to comment #3

There was indeed a typo. Changed to "…gives results different…" (now lines 467-468).

Comment #4

Section 4.4. Implications: I completely agree—and believe the authors convincingly demonstrate—that the data outputs of each software package are not statistically comparable and should not be mixed. However, I think it would be useful to highlight how many of the general results and their interpretations from Toothfrax and MountainsMap are comparable in a general sense. I think the appropriate analogy is how researchers frame discussions of results from dental microwear analyses using SEM or optical microscopy with the results obtained from DMTA.
I bring this up because of the recommendation to re-analyze raw surface data with MountainsMap. While I agree that re-analysis is essential for future research, the process will likely be slow and the results will resemble a patchwork of the total data that was originally published using Toothfrax/Sfrax for quite some time. Likewise, the absence of Tfv in MountainsMap would require further re-analysis once it is available (or at least

waiting for it to become available) for researchers that rely on this variable (see comments above). This critique does not detract from the main conclusions of the manuscript, but it would perhaps be a useful caveat to mention.

Response to comment #4

This is another good point. We have edited the section 4.4 accordingly:

"While the output produced by the new MountainsMap's SSFA module can be qualitatively compared to the output produced by Toothfrax, our results imply that the outputs cannot be directly combined into a single quantitative analysis. " (now lines 621-623)

"Even though MountainsMap can replace Toothfrax (with the caveats mentioned above), the parameter *Tfv* is still not available in MountainsMap. So, for the time being, the original Sfrax software cannot be replaced to compute *Tfv*. Once MountainsMap's SSFA module incorporates *Tfv*, a comparison like the one presented here will be necessary to check the similarities between the original Sfrax and the future MountainsMap calculations, and potentially, the re-analysis of published data will be also required for this parameter." (now lines 635-640).

Comment #5

The figures and tables are largely understandable without reference to the main body of the article. The minor exception is the use of abbreviations in some cases. While the abbreviations are given in the main text (e.g., SSFA, NMP, and HDI), it may be useful to include the full phrase along with the abbreviation in the legends.

Response to comment #5

Abbreviations are now explained in the table and figure legends.

Comment #6

SUPPLEMENTARY MATERIAL

One small issue with reference to Pedergnana et al discussion (Page 2), there is a typo. Please change "tough" to "though".

Response to comment #6

Thank you for noticing it. The typo has been corrected.

Reviewer #3 *(anonymous reviewer, 01 Sep 2022 08:18)*

Comment #1

Line 218 - 226: In most other publications by the authors they cite an ISO norm or paper indicating why they have used the steps referred to in their templace for the processing of 3D surfaces. Would this not be required here as well?

Response to comment #1

Yes, indeed. This relates to comment #3 from reviewer #1. We have given more details about our workflow (section 2.2.1).

Comment #2

Line 388 - 389: How were priors defined so that they "cover the whole range of data"? How specifically did the authors define their prior probability parameters and how were these parameters ensured to fulfill this "cover the whole range of data" criteria? In line 399 the authors state "The visual inspection helps in assessing whether the prior indeed covers the whole data range", however is there not a more automated/quantifiable way to perform this without having to rely on a "visual inspection"?

Response to comment #2

We have added details to the method to clarify these points (now lines 354-362):
"Priors were chosen so that they cover the whole range of the data. This way, the posterior is mostly influenced by data and not by prior. This was achieved in three steps: (1) by using hierarchical priors, i.e. prior distributions with hyperprior distributions for their parameters (see "s_x" and "$sd_{x.y}$" in the diagrams of the Bayesian models in Supplementary Material); (2) by choosing maximum entropy distributions that are by construction as vague as possible under the given constraints and (3) by suitable parameter choice. For parameter choice we used a version of the empirical Bayes method by using maximum standard deviations from the data as a basis for setting the prior widths. We combined it with educated guesses and found the final parameters iteratively by repeated visual inspection (see below) and adjustment afterwards."

There are ways to quantify the prior influence as the size of a hypothetical sample that would have the same effect on the posterior (see e.g. for details: Morita et al. 2008 https://doi.org/10.1111/j.1541-0420.2007.00888.x; Neuenschwander et al. 2020 https://doi.org/10.1111/biom.13252; Wiesenfarth & Calderazzo 2020 https://doi.org/10.1111/biom.13124). However, as to the best of our knowledge, these methods are not readily available in the common software packages, and it was out of the scope of this project to implement and test them for use here.

Comment #3
Lines 420-424: I think the arviz Python library should also be included in this list of used packages.

Response to comment #3
Indeed, we have added the ArviZ Python library to the list of used packages (now line 392).

Comment #4
Firstly I would suggest formulae be numbered and cited in the text.

Response to comment #4
The formulae in sections 2.2.3, 4.1 and 4.3 have been numbered (also following the PCJ template). These numbers are used for citation in the text, when needed.

Comment #5
Next, I have some concerns with some of the descriptions of these formulae, that should be clarified. It is not clear in lines 570 to 572 that the authors are referring to the Mean Absolute Deviation, therefore I suggest they simply correct the sentence using: "HAsfc is calculated using the Mean Absolute Deviation, instead of the median".

Response to comment #5
We have given more details to explain that, while HAsfc is defined as the MAD, its calculation in Toothfrax uses the mean (now lines 506-513):
"*HAsfc* is normally defined as median absolute deviation (MAD) divided by the median of values (Scott et al., 2006):

(3) $$HAsfc_{definition} = \frac{MAD}{\tilde{x}} = \frac{median(|x_i - \tilde{x}|)}{\tilde{x}}$$

with x_i being the individual *Asfc* values for each subregion, and $\tilde{x}$ being the median of the *Asfc* values of all subregions. However, in Toothfrax, *HAsfc* is calculated using the mean instead of the median:

(4) $$HAsfc_{Toothfrax} = \frac{1}{\tilde{x}} \frac{1}{N-1} \sum |x_i - \tilde{x}| = \frac{mean(|x_i - \tilde{x}|)}{\tilde{x}}$$

HAsfc has been implemented in Mountains Map in the same way as in Toothfrax (eq. (4)), but the true MAD calculation (eq. (3)) is also implemented in MountainsMap under the name *MadHAsfc*."

Comment #6
Next I am interested to know whether MAD has been normalised or not, and if not, why that is? (although I understand if/that this has probably got nothing to do with the authors - if it doesn't, do the authors know whether MAD has been normalised?). A simple reflection on these formulae - theoretically the formula presented in line 568 should be a more reliable measure of this type of data, considering how point clouds (i.e. the type of

data produced by the confocal), are prone to certain statistical errors which would make the median absolute deviation more reliable (Höhle and Höhle, 2009, DOI: 10.1016/j.isprsjprs.2009.02.003). I would therefore think the MountainsMap output to be a better approximation to the real surface characteristics.

Response to comment #6

The current implementation of *MadAsfc* in MountainsMap (see comment #5) is not normalized by the median, but it would be possible because the median is already calculated and available.

Comment #7

In line 651 the authors talk about the calculation of the "mean of all vectors", yet the formula in line 652 does not represent the mean at all - this formula actually describes the sum of all vectors excluding the final value (N - 1). Additionally, I would move the definition of N from line 656 up to line 653, considering how N is used earlier than it is defined. If the authors truly want to convey the mean of all vectors, the formula should be written as (I apologise for writing LaTeX annotation here but do not know of another way of conveying what the formula should be written as);

S_{relL} = \frac{1}{N} \sum_{i = 0}^{N} relL_{i}

Response to comment #7

We have reworded to clarify/correct (now lines 581-585). Note that all vectors are used (from 0 to N-1 = N vectors):

"The sum of all vectors is calculated as:

$$S_{relL} = \sum_{i=0}^{N-1} relL_i \quad (7)$$

with N being the number of angles ($N = \pi/\Delta\alpha$).

Then the mean cartesian coordinates are calculated as the normalized means: …"

Comment #8

Some final reflections: An interesting point that the authors raise in the discussion are the differences between MountainsMap and Toothfrax due to updates in the programming... as I am sure the authors agree, this is a fundamental component of computational science; software must be maintained regularly and updated as required so as to ensure the most accurate and computationally efficient results. Here lies one of my problems with these type of software, however, that they are not open source nor easy to obtain. Finally the authors discuss that raw surfaces must be made available - I completely agree with this, however I unfortunately think this is going to be a hard obstacle to overcome, especially when working with data obtained using confocal microscopy, as most surface scans occupy a lot of memory and require pretty powerful computers in order to process such data... nevertheless, I hope in the future we can find a way to make this sharing of data more feasible.

Response to comment #8

Open-source has many advantages over proprietary software, but maintenance and further developments is probably the main issue with open-source software packages, unfortunately. We do not want to dig into this very important discussion in the paper, though (but see e.g. Coelho & Valente 2017, https://doi.org/10.1145/3106237.3106246). Fact is that there is currently no open-source alternative as complete and easy-to-use for archaeologists and paleontologists as MountainsMap (e.g. Gwyddion). This might change in the future but it will take time.

The amount of data we produce for this type of analysis is not very large and can be uploaded to online repositories such as Zenodo without problems (Zenodo is run by the CERN, which is used to deal with dataset way larger than the ones presented here). However, for sure, online data storage is not without issues, especially considering the climate crisis (see e.g. https://doi.org/10.3390/publications8040054). While these aspects are very important at the societal level, we did not want to open this door in the paper. Nevertheless, from the reproducibility aspect and for the re-analysis of data (which is necessary due to the differences in the software packages), accessibility to the raw data is required and this is what we want to emphasize here.