Peer Community Journal

Section: Archaeology

RESEARCH ARTICLE

Published
2022-12-02

Cite as
Ivan Calandra, Konstantin Bob, Gildas Merceron, François Blateyron, Andreas Hildebrandt, Ellen Schulz-Kornas, Antoine Souron and Daniela E. Winkler (2022) *Surface texture analysis in Toothfrax and MountainsMap® SSFA module: Different software packages, different results?*, Peer Community Journal, 2: e77.

Correspondence
ivan.calandra@rgzm.de

Peer-review
Peer reviewed and recommended by PCI Archaeology, https://doi.org/10.24072/pci.archaeo.100024

Surface texture analysis in Toothfrax and MountainsMap® SSFA module: Different software packages, different results?

Ivan Calandra [1], Konstantin Bob [2], Gildas Merceron [3], François Blateyron [4], Andreas Hildebrandt [2], Ellen Schulz-Kornas [5,6], Antoine Souron [7], and Daniela E. Winkler [6,8,9]

Volume **2** (2022), article e77

https://doi.org/10.24072/pcjournal.204

Abstract

The scale-sensitive fractal analysis (SSFA) of dental microwear textures is traditionally performed using the software Toothfrax. SSFA has been recently integrated to the software MountainsMap® as an optional module. Meanwhile, Toothfrax support has ended. Before switching to the new module, the outputs between the two software packages must be compared for consistency. We have performed such a test using Bayesian modelling on three datasets including dental surfaces of sheep (Merceron, Ramdarshan, et al., 2016) and guinea pigs (Winkler, Schulz-Kornas, Kaiser, Cuyper, et al., 2019) from controlled feeding experiments, as well as surfaces of quartzite and flint flakes used in an actualistic archeological experiment on cleaning procedures (Pedergnana, Calandra, Bob, et al., 2020). We found that the two software packages calculate significantly different values for the SSFA parameters *epLsar*, *Asfc*, *HAsfc9* and R^2, even when the same settings are used. Nevertheless, the treatments (different diets or cleaning procedures) are discriminated similarly within each dataset. While the new software module is as good as the original software to differentiate treatments, our results imply that the outputs from the two software packages are not directly comparable and, as such, cannot be merged. Surface texture analysts should therefore consider re-analyzing published surfaces before integrating them in their studies.

[1]Römisch-germanisches Zentralmuseum, MONREPOS – Neuwied, Germany, [2]Scientific Computing and Bioinformatics, Institute of Computer Science, Johannes Gutenberg University – Mainz, Germany, [3]PALEVOPRIM, UMR CNRS 7262, University of Poitiers – Poitiers, France, [4]Digital Surf – Besançon, France, [5]Department of Cariology, Endodontology and Periodontology, University of Leipzig – Leipzig, Germany, [6]Leibniz Institute for the Analysis of Biodiversity Change, Center for Taxonomy and Morphology, Section Mammalogy and Paleoanthropology, Zoological Museum – Hamburg, Germany, [7]CNRS, MCC, PACEA, UMR 5199, University of Bordeaux – Pessac, France, [8]Department of Natural Environmental Studies, Graduate School of Frontier Sciences, The University of Tokyo – Chiba, Japan, [9]Institute of Geosciences, Applied and Analytical Paleontology, Johannes Gutenberg University – Mainz, Germany

Peer Community Journal is a member of the Centre Mersenne for Open Scientific Publishing
http://www.centre-mersenne.org/

e-ISSN 2804-3871

1. Introduction

The scale-sensitive fractal analysis (SSFA) is a method originally developed for the quantification of wear in industrial and engineering applications (e.g. Brown et al., 1993; Brown, 2000). A surface is rougher at finer scales than at larger ones, and SSFA investigates such changes across scales that are relevant for functional assessments. Since its first applications on primate teeth for dietary reconstructions (Ungar et al., 2003; Scott et al., 2005), SSFA has become an essential tool for biologists, paleontologists, zooarchaeologists and anthropologists as part of the dental microwear texture analysis (DMTA) (Ungar, 2015; DeSantis, 2016; Calandra & Merceron, 2016; Green & Croft, 2018; Schulz-Kornas, Kaiser, et al., 2020). 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). This method is therefore commonly applied to infer diets of extant and fossil vertebrates (Ungar & Evans, 2016; Ungar & Zhou, 2017) and to reconstruct paleo-environments (e.g. Ungar et al., 2012; Merceron, Novello, et al., 2016; Berlioz et al., 2018; Blondel et al., 2018; Ungar et al., 2020). SSFA has also been applied by archeologists to quantify the wear produced on stone and bone tools by different uses (Stemp et al., 2009, 2018; Lesnik, 2011; Stemp & Chung, 2011; Key et al., 2015; Watson & Gleason, 2016; Pedergnana, Calandra, Evans, et al., 2020; Pedergnana, Calandra, Bob, et al., 2020), although its application is less generalized than in biology/paleontology (Calandra, Pedergnana, et al., 2019).

Originally, the SSFA for DMTA has been performed with a custom-made software called Toothfrax, in combination with Sfrax, developed by SurFract Corp (Worcester, MA) (see details in Scott et al., 2006). It computes five parameters: area-scale fractal complexity (*Asfc*, referred to as "complexity"), scale of maximum complexity (*Smc*), length-scale anisotropy of relief (*epLsar*, "anisotropy"), heterogeneity of area-scale fractal complexity (*HAsfc*, "heterogeneity"), and textural fill volume (*Tfv*). Only *Tfv* is calculated in the software Sfrax; the other four parameters are computed with the software Toothfrax. Some recent studies have focused only on complexity and anisotropy because these parameters seem to be the most discriminant ones to reconstruct diets (e.g. Kubo et al., 2017; Schmidt et al., 2019; Arman et al., 2019; Winkler, Schulz-Kornas, Kaiser, & Tütken, 2019; Ackermans et al., 2020; Ungar et al., 2020). 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 et al., 2016; Krueger et al., 2017, 2019; see also Caporale & Ungar, 2016 on rodents).

Many manufacturers of 3D profilometers and confocal microscopes use the MountainsMap® software solution by Digital Surf (Besançon, France) as a software platform for their instruments. They also adjust the modularized MountainsMap® software for their requirements, resulting for example in ConfoMap by Carl Zeiss Microscopy (Jena, Germany), LeicaMap by Leica Microsystems (Wetzlar, Germany), SensoMAP by Sensofar (Barcelona, Spain) and µsoft analysis by Nanofocus (Oberhausen, Germany). This makes MountainsMap® software solution the most used software package to prepare and process the raw 3D surface data before conducting any further quantitative analysis.

The MountainsMap® software allows the calculation of other parameters according to 3D surface texture analysis using standardized ISO parameters and further analyses (STA or 3DST according to Schulz et al., 2010; Calandra et al., 2012). This analysis has been applied to teeth of various vertebrates (e.g. Purnell et al., 2012, 2013, 2017; Gill et al., 2014; Winkler et al., 2016; Calandra et al., 2016; Kubo et al., 2017; Yamada et al., 2018; Schulz-Kornas et al., 2019; Stuhlträger et al., 2019; Winkler, Schulz-Kornas, Kaiser, Cuyper, et al., 2019; Winkler, Schulz-Kornas, Kaiser, & Tütken, 2019; Aiba et al., 2019; Bestwick et al., 2019; Bethune et al., 2019), but also to lithic (e.g. Werner, 2018; Caux et al., 2018; Galland et al., 2019; Macdonald et al., 2019; Ibáñez et al., 2019; Pedergnana, Calandra, Evans, et al., 2020; Pedergnana, Calandra, Bob, et al., 2020; see also Rosso et al., 2017 for an application on ochre) and bone (e.g. Martisius et al., 2018, 2020; Turcotte et al., 2020) surfaces.

Digital Surf implemented the SSFA as new optional module for their MountainsMap® software, first released in 2018 in the MountainsMap® software version 7.4.8676. It is now possible in the MountainsMap® software to calculate different types of DMTA, e.g. SSFA and 3DST, in a single, optimized software with templates for automation.

Three differences between the original Toothfrax software and MountainsMap's SSFA module are noteworthy, though. First, in the SSFA module, the *Smc* parameter has been renamed to *Smfc* (scale of maximum fractal complexity) to differentiate it from the ISO 25178 *Smc* parameter (inverse areal material ratio of the scale-limited surface); we follow this terminology here. Second, the *Tfv* parameter is not available in the SSFA module, even in the latest release of MountainsMap® (v. 9.2.10170), although it will probably be released with a future update. Third, starting with MountainsMap® version 8, a new anisotropy parameter called *NewEpLsar* is available, alongside the original one. This new parameter is meant to correct a mistake in the code of Toothfrax (see section 4.3 below).

The majority of the published data were calculated with the software packages Sfrax and Toothfrax, and this remains true for many recent studies (e.g. Berlioz et al., 2017, 2018; Percher et al., 2018; Ungar & Berger, 2018; Smith & DeSantis, 2018, 2020; Merceron et al., 2018, 2021; Hofman-Kamińska et al., 2018; Tanis et al., 2018; Schmidt et al., 2019; Sewell et al., 2019; Arman et al., 2019; DeSantis et al., 2019, 2020; Stynder et al., 2019; Catz et al., 2020; Ungar et al., 2020; Robinet et al., 2020). Others have already moved to MountainsMap's SSFA module (e.g. Böhm et al., 2019; Winkler, Schulz-Kornas, Kaiser, & Tütken, 2019; Ackermans et al., 2020; Prassack et al., 2020; Schulz-Kornas, Winkler, et al., 2020; Winkler, Tütken, et al., 2020; Winkler, Schulz-Kornas, et al., 2020; Ungar et al., 2021). Sfrax and Toothfrax are not supported anymore; hence, the move to the new SSFA module is inevitable. This raises the question of whether the new module produces results that are identical to the ones from the original software, given identical contexts and configurations. In other words, can the output from the new SSFA module be directly compared to the published output of the Toothfrax software? Should this not be the case, are the results at least close enough, so that the functional interpretations (e.g. animal diet or tool use) remain? And how does the new anisotropy parameter (*NewEpLsar*) perform and compare to the original one?

Four years after the first release of MountainsMap's SSFA module, we believe it is now time to take a step back and address these questions. In this paper, we compare the outputs from the two software packages on three published datasets of experimental sheep teeth (Merceron, Ramdarshan, et al., 2016), guinea pig teeth (Winkler, Schulz-Kornas, Kaiser, Cuyper, et al., 2019) and lithic flakes (Pedergnana, Calandra, Bob, et al., 2020). These datasets cover a range of biological and archeological applications on different materials (enamel, flint and quartzite) and at different sizes (guinea pig and sheep teeth, and lithic flakes). Additionally, the data were acquired using three types of confocal microscopes (programmable array scanning, disc-scanning, and laser-scanning confocal microscopes; Artigas, 2011), with objectives having different magnifications and numerical apertures (100×/NA=0.90, 100×/NA=0.80 and 50×/NA=0.95), and at different pixel sizes (from 0.0852 to 0.16 µm). Therefore, the conclusions should be relevant to many applications in various scientific as well as industrial user communities beyond the ones shown here as examples.

2. Material and Methods

2.1. Material

This section gives brief summaries for the three published datasets analyzed here (Table 1). We refer the reader to the original publications for details. Raw data for the present analysis can be freely accessed on Zenodo (https://doi.org/10.5281/zenodo.7219855; see also section 2.2.3 below).

2.1.1. Sheep dataset

This dataset has been collected and analyzed by Merceron, Ramdarshan, et al. (2016). It contains data on a feeding experiment on 40 ewes (*Ovis aries*) fed four different diets for about 11 weeks: clover, grass, clover + dust, and grass + dust (10 ewes per diet).

The disto-labial enamel band of the protoconid of one of the lower second molars was molded using a polyvinylsiloxane elastomer. The molds were then scanned with a Leica DCM8 confocal profilometer (Leica Microsystems; see Table 1 for details on acquisition settings). Among other pre-processing steps (Table 1), abnormal peaks were automatically erased with a batch algorithm computed on IMAGEJ software based on mathematical morphological tools (this procedure can now be applied in MountainsMap® 8 and its derivatives directly).

Table 1. Overview of the three published datasets re-analyzed. FOV = field of view, N = sample size.

		Sheep	**Guinea pigs**	**Lithics**
Dataset	Reference	Merceron, Ramdarshan, et al. (2016)	Winkler, Schulz-Kornas, Kaiser, Cuyper, et al. (2019)	Pedergnana, Calandra, Bob, et al. (2020)
	N	40 ewes (*Ovis aries*)	18 guinea pigs (*Cavia porcellus*)	8 flakes (flint and quartzite)
	Treatments	Clover	Dry grass	Control
		Clover + dust	Dry lucerne	RubDirt
		Grass	Dry bamboo	BrushNoDirt
		Grass + dust	-	BrushDirt
Acquisition settings	Measuring equipment	Leica DCM8	Nanofocus µsurf Custom	Zeiss LSM800 MAT
	Objective	Leica 100×/0.90	Nikon long-distance 100×/0.80	Zeiss 50×/0.95
	Light source	White LED (550 nm)	Blue LED (470 nm)	Violet laser (405 nm)
	Step size	0.20 µm	0.06 µm	0.25 µm
	FOV	333 × 251 µm	160 × 160 µm	255.56 × 255.56 µm
	Frame size	2584 × 1945 pixels	984 × 984 pixels	3000 × 3000 pixels
	Measured on	Molds	Molds	Originals
Pre-processing	Software	Leica Map 7.0.6863	MountainsMap 7.4.8676	MountainsMap 8.2.9767
	Extracted FOV	200 × 200 µm	60 × 60 µm	50 × 50 µm
	Extracted frame size	1550 × 1550 pixels	369 × 369 pixels	588 × 588 pixels
	Mirroring	Z	Z	-
	Peak removal	batch algorithm for IMAGEJ based on mathematical morphological tools	-	-
	Surfaces saved as	SUR files	NMS files	SUR files
	N scans	40	70	32

2.1.2. Guinea pig dataset

This dataset has been collected and analyzed by Winkler, Schulz-Kornas, Kaiser, Cuyper, et al. (2019). It contains data on a feeding experiment on 18 guinea pigs (*Cavia porcellus*) fed three different diets for three weeks: dry grass, dry lucerne and dry bamboo (6 individuals per diet). The original dataset includes three extra fresh diets that were not included in the present analysis.

The mesial enamel band on the right upper fourth premolars was molded. The molds were scanned using a µsurf Custom confocal disc-scanning microscope (NanoFocus AG; see Table 1 for details on acquisition settings).

For each specimen, four non-overlapping scans were taken. For individuals 2CC5G1 and 2CC5G2 (dry grass group), only three scans could be acquired, as the enamel bands were too thin or showed damage. See Table 1 for details on pre-processing.

2.1.3. Lithic dataset

This dataset has been collected and analyzed by Pedergnana, Calandra, Bob, et al. (2020). It contains data on an actualistic brushing experiment on eight flakes. Flakes were knapped from two raw materials, flint or quartzite (four flakes each), and were subjected to four treatments: application of sediment and cleaning through rubbing with fingers (RubDirt), application of sediment and cleaning through brushing (BrushDirt), only brushing (BrushNoDirt), and control (no sediment and no brushing/rubbing), with one flake of each raw material per treatment.

The samples were scanned at two spots before and after treatment using a coordinate system to relocate the spots (Calandra, Schunk, Rodriguez, et al., 2019), resulting in 32 scans. An LSM800 MAT laser-scanning confocal microscope (Carl Zeiss Microscopy GmbH) was used for the scanning (see Table 1 for details on acquisition settings).

One spot of one sample ("Area 1" of sample FLT3-8) was not correctly acquired before the experiment but this was noticed only after the experiment had been performed; this area was therefore excluded from the subsequent analyses (see section *6.3.2 Exclude FLT3-8_Area1 ("before" wrongly acquired)* in the R script "R_analysis/scripts/SSFA_1_Import" on Zenodo https://doi.org/10.5281/zenodo.7219855). Note that this did not exclude the sample, as the second spot was included. In sum, 32 scans were processed (Table 1) but only 30 scans were included in the statistical analysis (see section 2.2.3).

Unlike Pedergnana, Calandra, Bob, et al. (2020), we restricted the analysis to a smaller, manually chosen area of the surface (Table 1) to limit the number of non-measured points to be filled later (see section 2.2.1).

2.2. Methods

The analysis presented below was performed with Carl Zeiss' ConfoMap as a derivative of Digital Surf's MountainsMap® software. Since the results are not specific to ConfoMap but are valid for any derivative of MountainsMap® featuring the SSFA module, we refer below to "MountainsMap" for simplicity and generality. Note however that the raw 3D surface data and the surface analysis (https://doi.org/10.5281/zenodo.6645445), as well as the statistical analyses (https://doi.org/10.5281/zenodo.7219855 and https://doi.org/10.5281/zenodo.7219884), explicitly mention ConfoMap.

2.2.1. Processing of 3D surfaces

All surfaces have been processed in MountainsMap version 8.2.9767 through a template performing the following steps. 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.

Because the surfaces for the Sheep dataset had already been cleaned of aberrant peaks before (see section 2.1.1), steps 3-4 were not applied on these surfaces.

The result surfaces (studiables after step 5) were saved manually in the uncompressed SUR format (for version 7.2 or older). This is crucial as the new compressed format (default when exporting the resulting studiables through templates) is not interpreted correctly by Toothfrax: *Asfc* value are extremely inflated (> 1000).

Note that in this analysis, as in some recent studies (e.g. Arman et al., 2016; Winkler, Schulz-Kornas, Kaiser, & Tütken, 2019; Ackermans et al., 2020; Pedergnana, Calandra, Evans, et al., 2020; Pedergnana, Calandra, Bob, et al., 2020; Winkler et al., 2021), the SSFA was run on scale-limited surfaces because form and measurement noise were filtered out prior to analysis. While this differs from the original applications of SSFA (Ungar et al., 2003; Scott et al., 2005, 2006), measurement noise can only blur the signal (Arman et al., 2016) and form is not related to the wear processes of interest here (Forbes, 2013; Evans et al., 2014; International Organization for Standardization, 2021).

Removing points with a high slope (> 80°) and filling these non-measured points (NMP) with a smooth shape from neighbors can artificially reduce the degree of fractality of the surfaces. Based on the numerical apertures of the objective used to measure the surfaces (Table 1), slopes > 72° (for NA = 0.95) on flat

surfaces cannot be measured by the instruments; the maximum measurable slopes are even lower for smaller NAs (53° for NA = 0.80 and 64° for NA = 0.90). Even accounting for scattering on rough surfaces, slopes > 80° are very likely to result in aberrant points; keeping these aberrant points would artificially increase the fractality.

Some surfaces had very high ratios of NMP (≥ 20%). While these surfaces were processed, they were excluded from the statistical analysis (see section 2.2.3). For surfaces with less than 20% NMP, it should be noted that the NMP are isolated rather than concentrated (see results on Zenodo). Additionally, we ran the statistical analyses on two sets of surfaces: surfaces with NMP ratio < 5% and surfaces with NMP ratio < 20% (see section 2.2.3).

Altogether, we argue that this processing ensures data quality by removing aberrant points and does not substantially alter the degree of fractality of the surfaces.

The results of the processing for each surface can be freely accessed in both MNT (containing the acquired surface, the workflow and the results) and PDF (exported from the MNT) formats on Zenodo (https://doi.org/10.5281/zenodo.6645445).

2.2.2. SSFA

The SSFA was performed in two software packages: MountainsMap version 8.2.9767 with the SSFA module, and the original Toothfrax software. The latter followed the steps detailed in Scott et al. (2006), except for the maximum scale of the area-scale computation, which was reduced from 7200 to 1200 µm^2. This setting was modified to avoid erroneous calculations of *Asfc* when *Smfc* is very high (> 1200 µm^2; see also section 4.1 below). We did try with a maximum scale of 7200 µm^2 but *Smfc* was never higher than 300 µm^2. This means that limiting the maximum scale to 1200 µm^2 does not change the results in this case. Since the *Tfv* parameter is not available yet in MountainsMap, Sfrax was not used.

The same settings were used in MountainsMap; all other settings were set to the default values. Since the SSFA module is relatively new, we provide a more detailed workflow here. First, a length-scale SSFA was conducted on rows; two parameters were calculated: *epLsar* and *NewEpLsar*, both calculated at 5° intervals and at the 1.8 µm scale of observation by default. Second, an area-scale SSFA was conducted on four corners; five parameters were calculated: R^2, *Asfc*, *Smfc*, *HAsfc9* (3×3 subregions) and *HAsfc81* (9×9 subregions). Note that due to the small size of the analyzed surfaces for the guinea pig and lithic datasets (Table 1), *HAsfc81* was considered only for the sheep dataset and was excluded from the Bayesian analysis (see section 2.2.3). Other *HAsfc* parameters with splittings from 2×2 to 11×11 are also available but were not used. Both analyses were performed with the full data, i.e. the option "draft analysis" was deactivated.

R^2 is the coefficient of determination for the curve of relative area versus scale. It can be seen as a quality indicator (the higher the value, the better), but it is usually not used in dietary or functional interpretations.

The results of the SSFA in MountainsMap for each surface can be freely accessed in both MNT and PDF formats on Zenodo (https://doi.org/10.5281/zenodo.6645445).

2.2.3. Statistical analysis

All raw data, derived data, plots, scripts and reports were incorporated into a research compendium following Marwick et al. (2018), which is freely available on GitHub (https://github.com/tracer-monrepos/SSFAcomparisonPaper) and Zenodo (https://doi.org/10.5281/zenodo.7219855).

Descriptive statistics

The CSV output files from Toothfrax and MountainsMap were first imported, formatted and merged. Descriptive statistics were then calculated and boxplots were plotted; two surfaces from the lithic dataset (section 2.1.3), as well as all surfaces with ≥ 20% NMP (section 2.2.1), were excluded for these steps.

All statistical analyses were performed in the open-source software R version 4.1.2 (R Core Team, 2021) through RStudio version 2021.9.1.372 (RStudio Team, 2021) for Microsoft Windows 10.

The following packages were used: doBy version 4.6.11 (Højsgaard & Halekoh, 2021), ggh4x version 0.2.1.9000 (van den Brand, 2021), ggplot2 version 3.3.5 (Wickham, 2016), openxlsx version 4.2.4 (Schauberger & Walker, 2021), R.utils version 2.11.0 (Soler, 2022), rprojroot version 2.0.2 (Müller, 2020) and tidyverse version 1.3.1 (Wickham et al., 2019). The scripts were written as R markdown (*.Rmd) files

and rendered to HTML and GitHub (i.e. Markdown, *.md) documents using the packages knitr version 1.36 (Xie, 2014, 2015, 2021) and rmarkdown version 2.11 (Xie et al., 2018, 2020; Allaire et al., 2021).

The research compendium was created in R using the packages rrtools version 0.1.5 (Marwick, 2019) and usethis version 2.1.3 (Wickham et al., 2021). The R computing environment (packages and their versions) was incorporated into the research compendium using the package renv version 0.14 (Ushey, 2021).

Bayesian statistics

Overview. In order to test the influence of the different software packages on the surface parameters *Asfc*, *epLsar*, *HAsfc9*, *Smfc*, R^2 and *NewEpLsar*, a Bayesian multi-factor ANOVA was applied on each surface parameter. As mentioned in section 2.2.2, *HAsfc81* is relevant only for the sheep dataset and was therefore excluded from the Bayesian analysis. This method uses models to relate the dependent variables (measurement outcomes, i.e. surface parameters) to the independent variables (factors, i.e. treatments, software, samples; see below) which are assumed to influence the former. It can ultimately be used to compute whether the effects of factors differ significantly. Similar Bayesian statistics have been successfully applied as multilevel and multivariate Bayesian models to explain surface parameter variation measuring micro-topography of bone surfaces (Martisius et al., 2018, 2020).

Compared to the traditional null hypothesis testing procedure, there are several advantages to this approach (Kruschke, 2013). First, this method does not rely on assumptions other than the ones stated below and is therefore more transparent. In particular, the data do not have to be normally distributed. Second, a Bayesian approach allows leveraging prior knowledge and more importantly, the uncertainty of the results can also be estimated by using the full posterior distribution. Finally, the availability of steadily increasing computational power and appropriate software libraries means that the greater complexity of the computation should not present a serious drawback anymore relative to the gain in insight.

Software, i.e. MountainsMap or Toothfrax, is considered here as the first factor, x_1. Treatment (diet and cleaning procedures, e.g. clover, dry bamboo, BrushNoDirt) is the second factor x_2, and sample is the third factor x_3.

Three models were built. Different factor combinations were defined for the respective models. The first model examined the differences due to software, considering the effects of treatment and sample, for each of the surface parameters except *NewEpLsar*. In other words, it looks at the differences due to software for each individual surface within each treatment. It allows us to test whether the two software packages produce statistically identical values. In this model, all three factors (x_1, x_2 and x_3) were considered but not their interactions, as the sample factor (x_3) already explains most of the data and our goal was to test whether the software packages differ conditionally on all treatments and samples. We refer to this model as the "Three-factor model".

The second model was used to inspect the differences in treatment due to software. In this case, individual surfaces were not considered anymore. The goal was to compare the discrimination power between treatments of each software package, for each of the surface parameters except *NewEpLsar*. Thus, the factors software and treatment (x_1 and x_2, respectively) and their interaction were used. We call this model "Two-factor model".

The last model was applied to test whether the *NewEpLsar* parameter (only calculated in MountainsMap) discriminates treatments in the same way as the standard *epLsar* (from both software packages) does. A model using only the treatment factor (x_2) was used. This is the "NewEpLsar model".

All models were run twice: a first time on surfaces with less than 5% NMP ("strong filter"), and a second time on surfaces with less than 20% NMP ("weak filter"; see section 2.2.1).

Details. Before performing the Bayesian multi-factor ANOVA, the values of each surface parameter were transformed into standard scores (z-scores), i.e. the sample mean was subtracted and the result was divided by the sample standard deviation. There are two reasons for this. First, from a theoretical point of view, this enables the use of a single model for every parameter, which in turn improves comparability. Second, from a computational point of view, having all values involved in a narrow numerical range stabilizes the algorithm against numerical errors. Note that while the description of the model below is expressed in standard scores, the estimated model parameters were scaled back to the original numerical ranges for the plots, easing the comparison.

In all models, for every single measured surface parameter, the predicted value μ is related to the factors by a linearized model, which is given by

$$\mu = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + x_1 M_{12} x_2 + x_1 M_{13} x_3 + x_2 M_{23} x_3 \quad (1)$$

where all possible terms up to the second order are included. The terms of the equation (1) can be understood as follows: β_0 is a real number that indicates the overall order of magnitude of the measured values. β_1 is a vector of length 2 that contains the effect strengths of choosing the respective software, while x_1 is a vector that indicates the level of factor 1, i.e. x_1 is [0, 1] when choosing the first level of factor (Toothfrax) and [1, 0] when choosing the second level (MountainsMap). The same applies to β_2 and x_2, but here with 11 different levels for each treatment (4 diets for the sheep dataset, 3 diets for the guinea pig data and 4 cleaning procedures for the lithic dataset), and to β_3 and x_3 with 140 levels (40 analyzed surfaces for the sheep dataset, 70 for the guinea pig data and 30 for the lithic dataset). M_{12}, M_{13} and M_{23} are matrices that indicate the effect strength of the particular combination of two factors. Joint modeling of all three datasets in a single large model improves the estimates of quantities that occur in all datasets (e.g. the effect of the software factor), because there are more samples to calculate the posterior with. However, the effects of the treatments were analyzed within each dataset.

The predicted value μ is related to the observed value y by

$$y \sim t(\mu, \sigma, \nu) \quad (2)$$

where t denotes a Student's t distribution with mean μ, standard deviation σ, and ν degrees of freedom.

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 Information 1); (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.

The posterior distribution is then accessed by sampling using a special variant of Markov Chain Monte Carlo, the Hamiltonian Monte Carlo algorithm (Hoffman & Gelman, 2014) in the implementation by Salvatier et al. (2016). When performing the sampling, the results have to be checked for consistency based on the trace plots and on the energy plots of Hamiltonian Monte Carlo (see notebooks in the folder "Python_analysis/code/" on Zenodo). All models were also visually inspected by providing prior-predictive (see "prior_predictive" plots in the folder "Python_analysis/plots/" on Zenodo) and posterior-predictive (see "prior_posterior_predictive" plots in the folder "Python_analysis/plots/" on Zenodo) plots. These show the distribution of hypothetical data based on the prior and posterior respectively, thus being easy interpretable. The visual inspection helps in assessing whether the prior indeed covers the whole data range, how well the posterior fits the data, and how large the uncertainty in the parameter estimation is.

By selecting the appropriate terms of the general model shown above and providing prior distributions for all remaining model parameters, the model gets adapted to the question at hand. Kruschke (2015, section 15.4.1) provides further insight into model selection for typical analysis questions and was used as guidance. We built three different models: the Three-factor, the Two-factor and the NewEpLsar models (see above). A detailed, graphical description of the models is given in the Supplementary Information 1.

In order to compare the levels of the factors, we computed contrasts, i.e. the probability distribution of the difference in posterior between any two levels of a factor. The contrasts are then plotted (see "contrast" and "treatment_pairs" plots in the folder "Python_analysis/plots/" on Zenodo): the x-axis shows the possible values for the difference and the curve gives the probability (technically: probability density) that each of these values is the "true" value. If 95% of the values are away from 0, i.e. if the 95% high probability density interval (HDI) does not include 0, then the two levels are significantly different. In the three-factor model, the contrast between the two levels of the software factor (β_1 in equation (1)) was computed, excluding the effects of treatment and sample. In the two-factor model, the contrasts between

levels of treatments (β_2 in equation (1)) belonging to the same datasets were computed, for each software individually, taking the interaction terms (M_{12} in equation (1)) into account as well. Lastly, for the NewEpLsar model, the contrasts between levels of treatments (β_2 in equation (1)) belonging to the same datasets were computed.

For the analysis in the Python programming language version 3.8.5 (Van Rossum & Drake, 2010), the following free and open-source software packages were used: ArviZ version 0.9.0 (Kumar et al., 2019), pandas version 1.1.2 (Reback et al., 2020), matplotlib version 3.3.1 (Hunter, 2007), NumPy version 1.19.2 (Harris et al., 2020), seaborn version 0.11.0 (Waskom, 2021), and PyMC3 version 3.9.3 (Salvatier et al., 2016). A Docker image of the analysis is available on Zenodo (https://doi.org/10.5281/zenodo.7219884).

3. Results

The complete research compendium (including both R and Python analyses) is freely available on GitHub (https://github.com/tracer-monrepos/SSFAcomparisonPaper); for long-term accessibility, a release (version 3.1) has been uploaded to Zenodo (https://doi.org/10.5281/zenodo.7219855).

The goal of the present study is not to formally compare the discrimination power among treatments between the Bayesian analysis presented here and the analyses presented in the original publications. Nevertheless, a qualitative comparison is presented in Supplementary Information 2.

The three-factor model shows that, when accounting for variations due to treatments (diet for animals and cleaning procedures for lithics) and samples (individual 3D surfaces), *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).

HAsfc81 was not statistically analyzed because it is only relevant for the sheep datasets, for which larger surfaces were analyzed (see sections 2.2.2 and 2.2.3). Nevertheless, it seems to follow the same trend as *HAsfc9* (Fig. 1).

Smfc could not be analyzed due to a few extremely high values. Graphically, it seems that in general MountainsMap produces values that are higher than Toothfrax. This pattern becomes apparent when some of these extreme values in the sheep and lithic datasets are excluded (see "R_analysis/plots/SSFA_plots-Smfc.pdf" on Zenodo).

When analyzing all surfaces with less than 20% NMP, the results vary only slightly: *Asfc*, *epLsar*, *HAsfc9* and R^2 are again significantly different when calculated with the two software packages (compare files of the "filter_strong" and "filter_weak" runs for the three-factor model in the folder "Python_analysis/" on Zenodo). *HAsfc81* (sheep dataset) and *Smfc* follow the trends mentioned above.

Yet, as shown by the two-factor model (i.e. when samples are not accounted for), these differences did not obscure most of the differences in treatment within each dataset (Fig. 5; see also treatment pair plots in the folder "Python_analysis/plots/statistical_model_two_factors_filter_strong/" on Zenodo). Nine of the 60 pairwise comparisons (15 contrasts for each of the four surface parameters) yielded different results between the two software packages, five of which on R^2 (Table 2). Additionally, there is no pattern in these different contrasts. In other words, none of the software is better at discriminating the treatments than the other, whether considering any single surface parameter or dataset, or taking all parameters into account. In summary, even though the two software packages produce significantly different results at the level of individual surfaces, roughly the same significant differences in treatment were found independently of the software used.

When analyzing all surfaces with less than 20% NMP, even less differences appear between the discriminations of the two software packages (compare files of the "filter_strong" and "filter_weak" runs for the two-factor model in the folder "Python_analysis/ " on Zenodo). This could result from the noise introduced by the lower quality data (NMP ≥ 5%) blurring the signal.

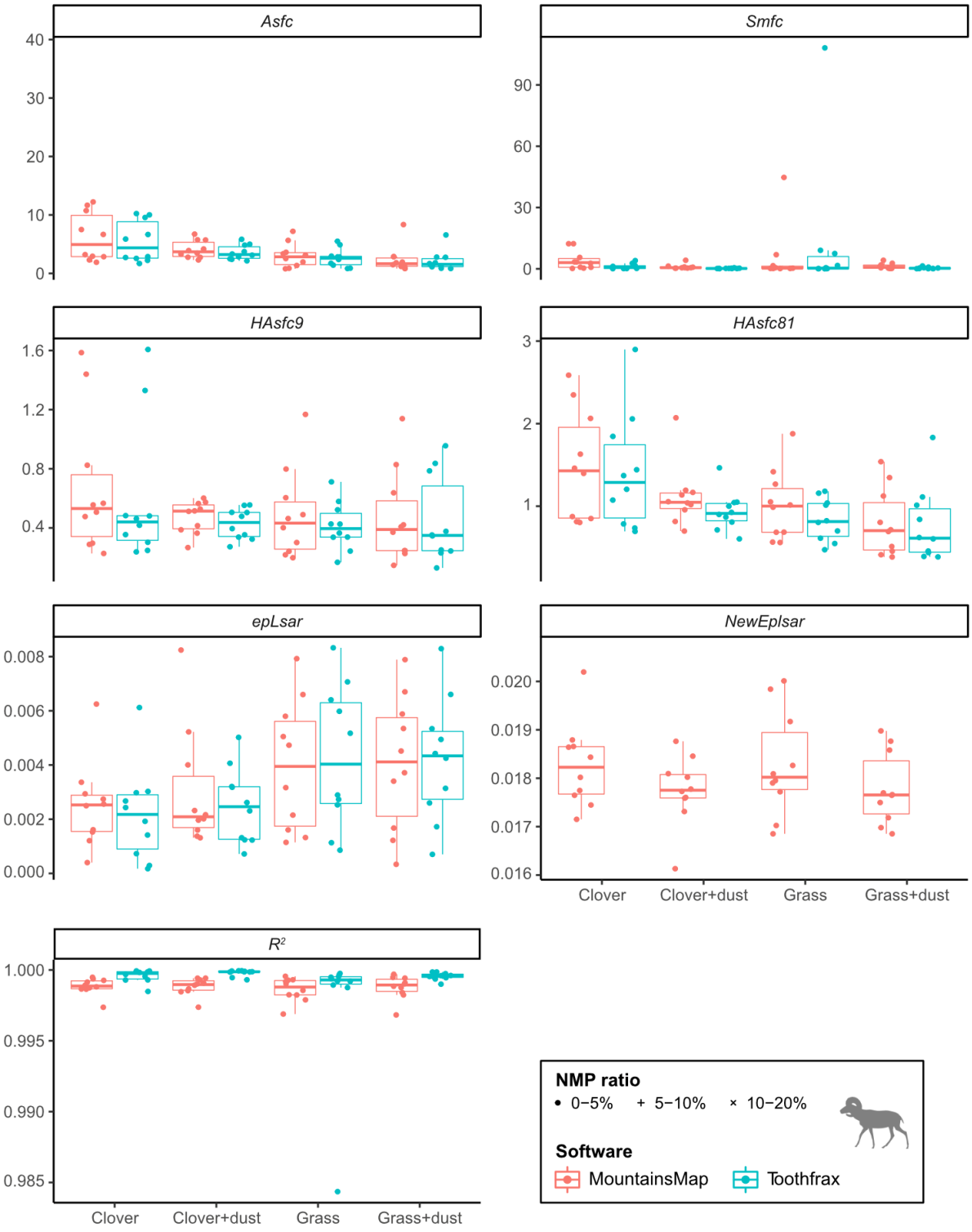


Fig. 1. Boxplots of the Scale-Sensitive Fractal Analysis (SSFA) parameters for the sheep dataset. The boxes represent the interquartile range (IQR), i.e., between the 25th and 75th percentile, with the median shown as a thick horizontal line. The bars extend up to 1.5 IQR on each side of the box. The points are spread horizontally within each group to avoid overlap for readability. NMP = non-measured points. See text for details on surface parameters and treatments. Sheep silhouette from PhyloPic (phylopic.org, Scott Hartman, Public Domain Dedication 1.0 license).

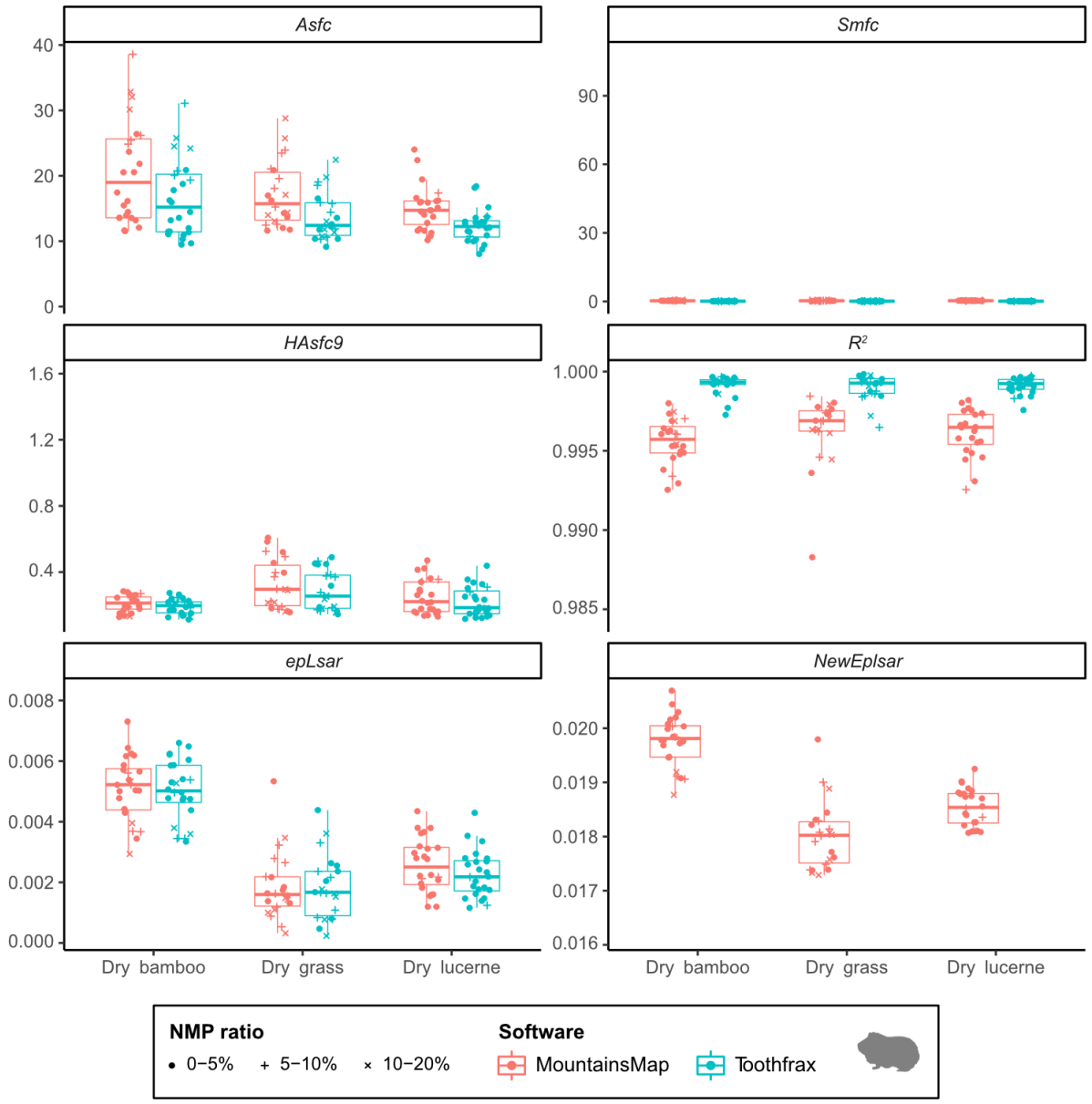


Fig. 2. Boxplots of the SSFA parameters for the guinea pig dataset. See Fig. 1 for details on boxplots and abbreviations. See text for details on surface parameters and treatments. Guinea pig silhouette from Wikimedia Commons (Flappiefh, CC BY 3.0 license).

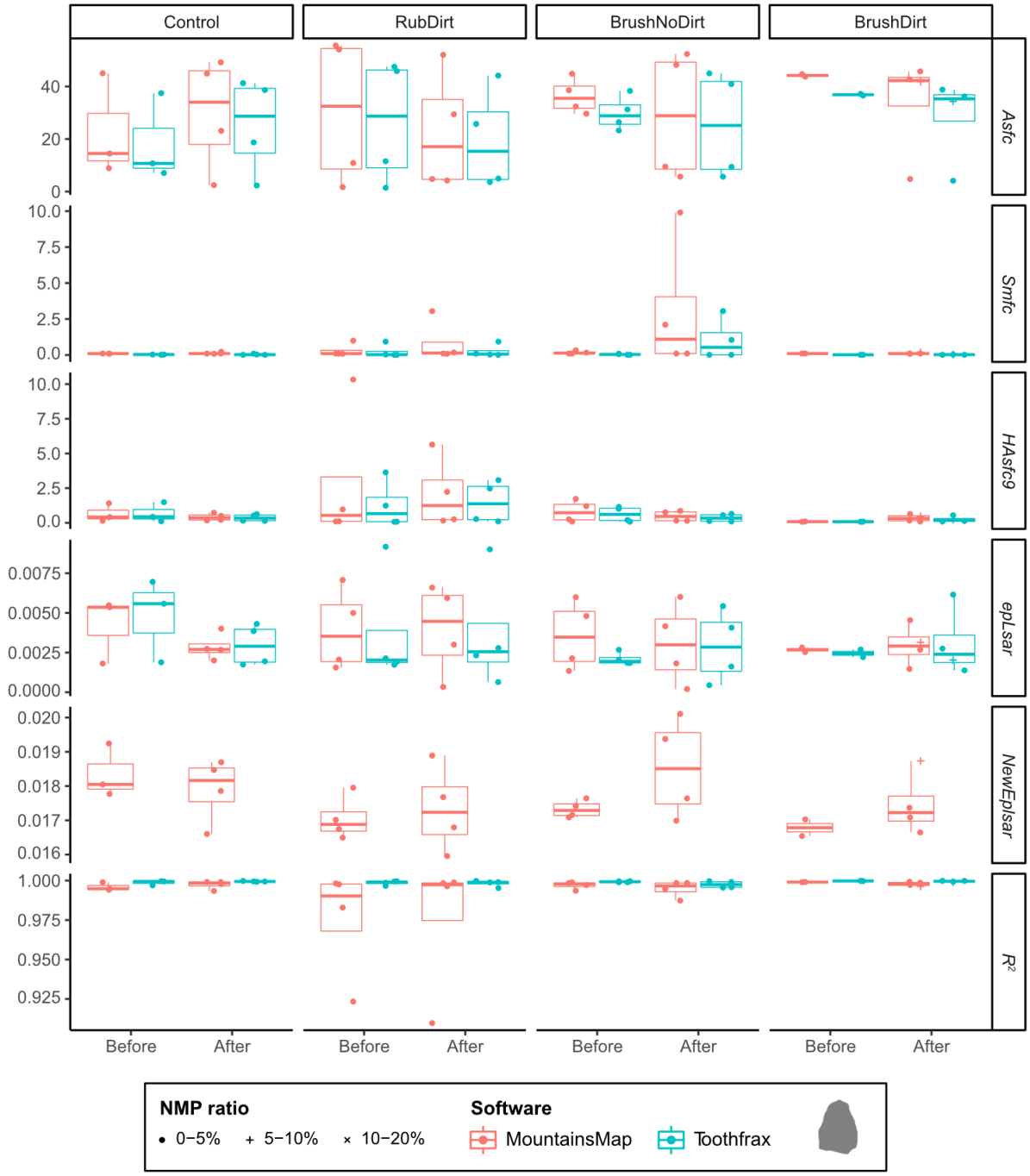


Fig. 3. Boxplots of the SSFA parameters for the lithic dataset. Surfaces with more than 20% NMP were excluded (three surfaces per software package; see "R_analysis/summary_stats/ SSFA_summary-stats.xlsx " on Zenodo). See Fig. 1 for details on boxplots and abbreviations. See text for details on surface parameters and treatments. Silhouette: courtesy from J. Marreiros.

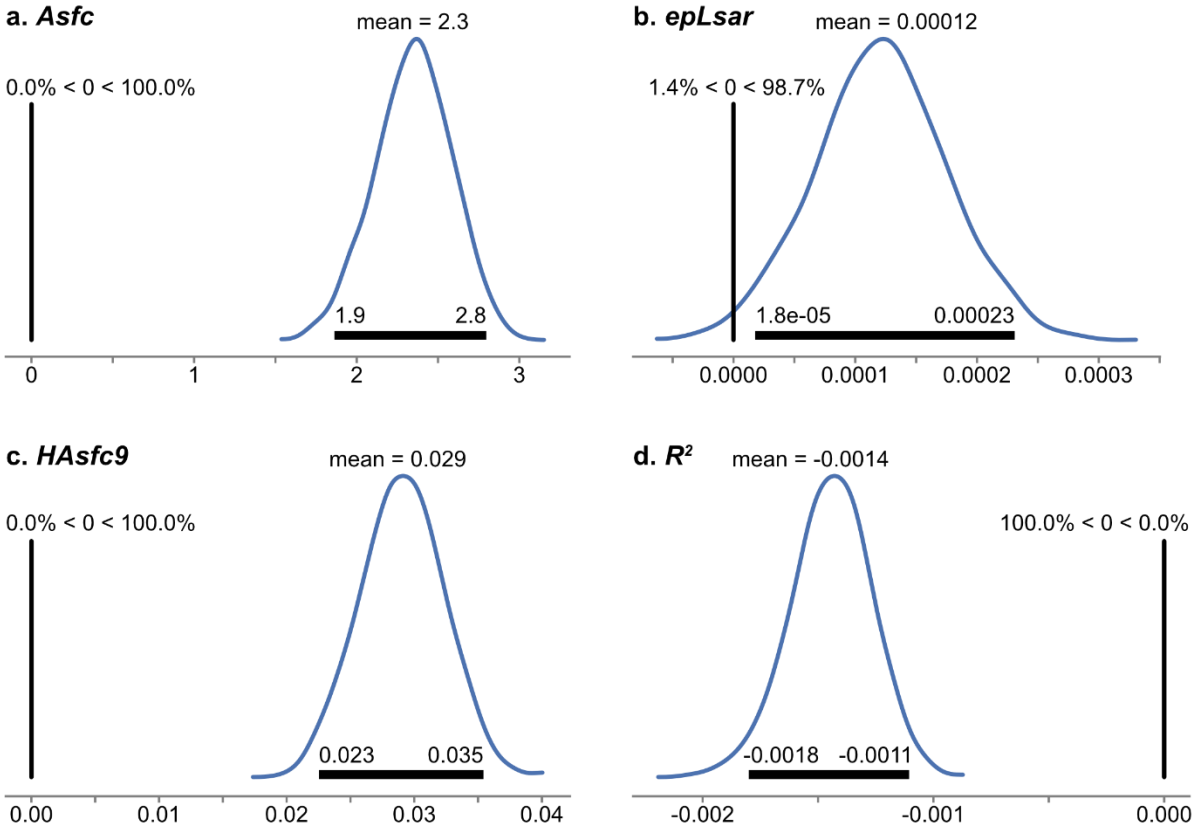


Fig. 4. Contrast plots of MountainsMap *vs.* Toothfrax from the three-factor model for the Scale-Sensitive Fractal Analysis (SSFA) parameters *Asfc* (**a**), *epLsar* (**b**), *HAsfc9* (**c**) and R^2 (**d**) on the surfaces with < 5% non-measured points (NMP). The y-axis shows the probability density and the x-axis represents the difference in parameter between the two software packages. The black horizontal bar in each plot covers the 95% high density interval (HDI), with the associated values for its boundaries (2.5 and 97.5%). The black vertical line marks the 0-effect strength; the values above it indicate the estimated probabilities of the contrast being smaller or greater than 0. See text for details on surface parameters.

Even though the values for *NewEpLsar* (calculated in MountainsMap only) are much higher than the values for *epLsar* (from both software packages; Figs 1-3), they mostly discriminate the treatments in the same way as the *epLsar* values do (Fig. 6; see also treatment pair plots in the folder "Python_analysis/plots/statistical_model_neweplsar_filter_strong/" on Zenodo). There are six differences out of 15 contrasts, though (Table 3). Interestingly, in those six cases, *NewEpLsar* always gives results different from Toothfrax's *epLsar*, although it does not always agree with MountainsMap's *epLsar* (i.e. *NewEpLsar* sometimes discriminates differently from *epLsar* in general).

When analyzing all surfaces with less than 20% NMP, one more difference appears between the discriminations *NewEpLsar* compared to *epLsar* (compare files of the "filter_strong" and "filter_weak" runs for the NewEpLsar model in the folder "Python_analysis/ " on Zenodo). This could result from the noise introduced by the lower quality data (NMP ≥ 5%) "faking" a signal.

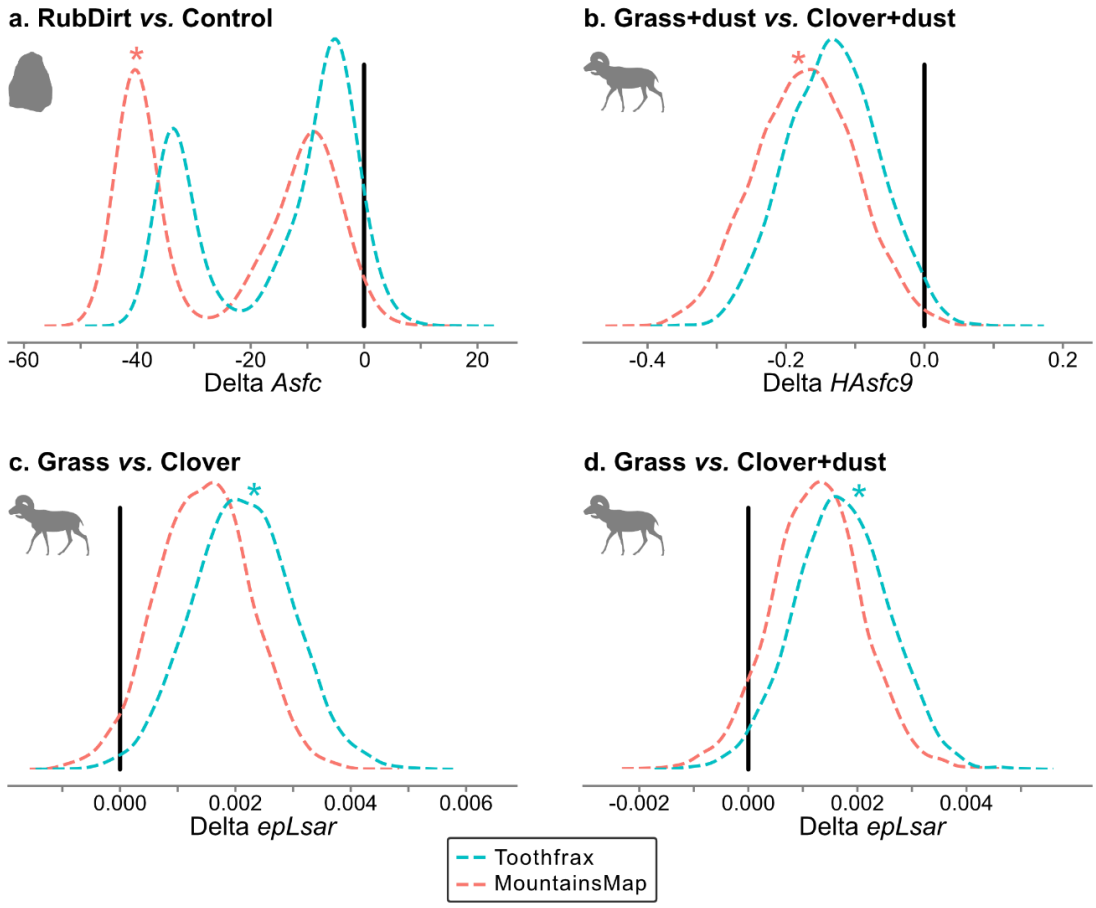


Fig. 5. Contrast plots of treatment pairs for which both software packages disagree on the surfaces with < 5% NMP: *Asfc* (**a**), *HAsfc9* (**b**) and *epLsar* (**c-d**). Stars mark significant differences for the given treatment pair and software. See Table 2 for 95% HDIs and results on R^2, Fig. 4 for details on how to read the plots and for abbreviations, Figs 1&3 for details on silhouettes, and text for details on surface parameters and treatments.

Table 2. Differences in significance for the two software packages. Only treatment pairs and parameters for which the software packages disagree are shown. For all comparisons, see frames [44], [69], [95] and [128] in the notebook "Statistical_Model_TwoFactor_filter_strong" (folder "Python_analysis/code/" on Zenodo). "True" indicates that the given treatment pair is significantly different for the given parameter and software, i.e. its 95% high probability density interval (HDI) does not include 0. See text for details on surface parameters and treatments. Signif. = significance.

	Treatment		**MountainsMap**			**Toothfrax**		
Parameter	**i**	**j**	**Signif.**	**2.5% HDI**	**97.5% HDI**	**Signif.**	**2.5% HDI**	**97.5% HDI**
Asfc	RubDirt	Control	True	-44.051	-2.428	False	-37.081	0.973
HAsfc9	Grass+dust	Clover+dust	True	-0.3111	-0.0322	False	-0.2534	0.0001
epLsar	Grass	Clover	False	-0.00020	0.00289	True	0.00043	0.00376
epLsar	Grass	Clover+dust	False	-0.00034	0.00288	True	0.00006	0.00339
R^2	Dry grass	Dry bamboo	True	0.00154	0.00283	False	-0.00034	0.00045
R^2	Dry lucerne	Dry bamboo	True	0.00055	0.00186	False	-0.00037	0.00018
R^2	Dry lucerne	Dry grass	True	-0.00154	-0.00044	False	-0.00061	0.00020
R^2	RubDirt	BrushDirt	True	-0.00156	-0.00007	False	-0.00112	0.00003
R^2	Grass	Clover+dust	False	-0.00070	0.00053	True	-0.00083	-0.00014

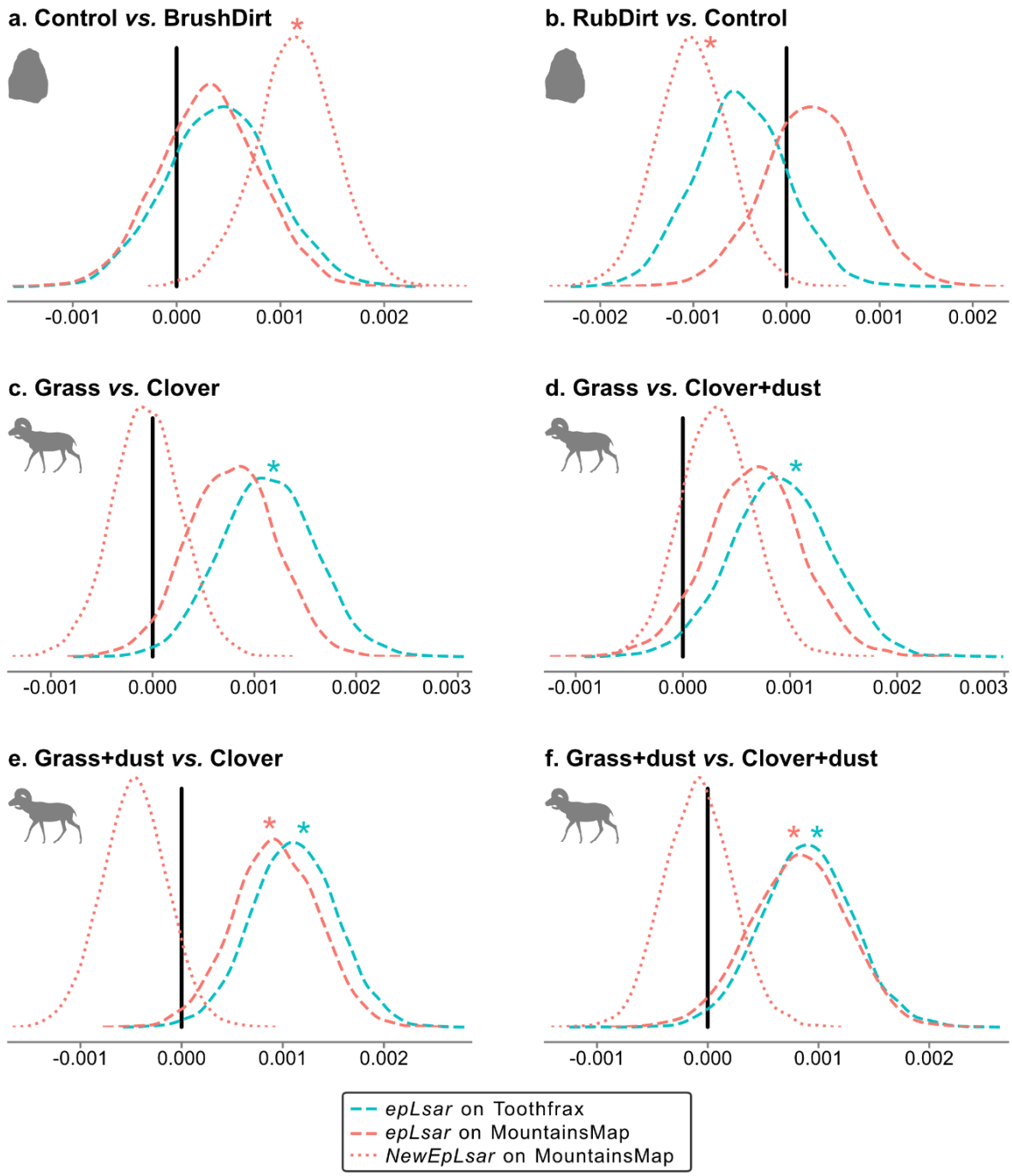


Fig. 6. Contrasts of treatment pairs for which *NewEpLsar* disagrees with *epLsar* from at least one software package on the surfaces with < 5% NMP. See Table 3 for 95% HDIs, Fig. 4 for details on how to read the plots and for abbreviations, Figs 1&3 for details on silhouette, and text for details on surface parameters and treatments.

Table 3. Differences in significance for *NewEpLsar* against *epLsar* from the two software packages. Only treatment pairs and parameters for which *NewEpLsar* disagrees with *epLsar* from at least one software package are shown. For all comparisons, see frame [55] in the notebook "Statistical_Model_NewEplsar_filter_strong" (folder "Python_analysis/code/" on Zenodo). See Table 2 for details on how to read the table and for abbreviations, and text for details on surface parameters and treatments.

Treatment		MountainsMap *NewEpLsar*			MountainsMap *epLsar*			Toothfrax *epLsar*		
i	j	Signif.	2.5% HDI	97.5% HDI	Signif.	2.5% HDI	97.5% HDI	Signif.	2.5% HDI	97.5% HDI
Control	BrushDirt	True	0.00036	0.00185	False	-0.00056	0.00130	False	-0.00059	0.00139
RubDirt	Control	True	-0.00178	-0.00025	False	-0.00070	0.00133	False	-0.00144	0.00047
Grass	Clover	False	-0.00075	0.00059	False	-0.00011	0.00154	True	0.00023	0.00201
Grass	Clover+dust	False	-0.00030	0.00099	False	-0.00018	0.00154	True	0.00003	0.00181
Grass+dust	Clover	False	-0.00112	0.00015	True	0.00013	0.00175	True	0.00031	0.00194
Grass+dust	Clover+dust	False	-0.00067	0.00052	True	0.00001	0.00167	True	0.00012	0.00172

4. Discussion

We compared Toothfrax and MountainsMap's SSFA module by analyzing the same surfaces with both software packages using the same settings; Bayesian multi-factor ANOVAs were then used to assess the significance of the divergences. The surfaces are derived from three datasets of experimental samples (guinea pig and sheep teeth, and flint and quartzite flakes) and the values range over several orders of magnitude (*epLsar* from 0.00017 to 0.00917, *Asfc* from 0.79 to 68.22 and *HAsfc9* from 0.07 to 10.32; Figs 1-3; see also table "/R_analysis/summary_stats/SSFA_summary_stats.xlsx" on Zenodo https://doi.org/10.5281/zenodo.7219855). We therefore assume that it is very likely that the findings will apply to other types of surfaces as well.

4.1. Comparing Toothfrax and MountainsMap outputs

The results of the three-factor model show that the two software packages do produce divergent results. The effect on *Asfc* is very large (Fig. 4a), but the effects on the other parameters are small (Fig. 4b-d). Nevertheless, *Asfc*, *epLsar* and *HAsfc9* are all significantly higher when calculated with MountainsMap (Figs 1-3). Different calculations of *HAsfc* and *epLsar* in the two software packages could explain the differences on these parameters.

HAsfc is normally defined as median absolute deviation (MAD) divided by the median of values (Scott et al., 2006):

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

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:

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

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*.

The spacing of the extracted profiles is different between the implementations of *epLsar* in Toothfrax and MountainsMap. While it should be fixed at 1.8 µm by default, it seems that profiles extracted at different angles can have different spacing in Toothfrax; the spacing is constant in MountainsMap.

R^2 is consistently higher when using Toothfrax, although the mean difference in posterior is very small yet significant (-0.0014; see Figs. 1-3 and 4d). Since this parameter describes the quality of the correlation in the area-scale analysis, it would imply that the change of relative area across scales is more regular in Toothfrax. It could mean that the results of the area-scale analysis (*Asfc* and *HAsfc* parameters) are more robust when using Toothfrax. The difference could also be a result of the higher number of points per

surface used in MountainsMap's SSFA module, especially at the smallest scales. Toothfrax's point spread settings can be used in MountainsMap's SSFA too, but this is not the default (the default settings were used here). Another potential source of difference is that Toothfrax calculates by default a regression of order 2, while MountainsMap uses a regression of order 1.

Although not tested in the Bayesian models due to some very extreme values, it seems that *Smfc* is higher when computed with MountainsMap than with Toothfrax (Figs 1-3; see also "SSFA_plots-Smfc.pdf" in the folder "/R_analysis/plots/" on Zenodo https://doi.org/10.5281/zenodo.7219855). Unlike Toothfrax, MountainsMap has rules to avoid calculating *Smfc* over an inappropriate range of scales (i.e. smallest or largest scales). This could explain the different results. We conclude that the complexity (*Asfc*) and heterogeneity (*HAsfc*) parameters are calculated at different scales with the two software packages using the default settings. We propose that this difference in *Smfc* could actually be the source of the differences in *Asfc* and *HAsfc* (together with differences in the pavement algorithms, see below). It is therefore possible that the differences in complexity and heterogeneity would disappear if they were calculated at the same scale in both software packages, i.e. if both software packages would calculate *Smfc* the same way. In MountainsMap, it is possible to manually adjust the range of scales over which the regression is calculated; it could therefore be tested whether the two software packages produce the same *Asfc* and *HAsfc* values when *Smfc* values are identical.

Another potential source of the differences between the two software packages concerns the pavement algorithm in the area-scale method (used for all parameters but *epLsar* and *NewEpLsar*). In this method, triangles of a given area are used to cover (i.e. pave) the topography. A general description of this pavement is given in different standards (ASME B46.1 and ISO 25178-2); however, this description is conceptual and leaves some freedom for the implementation. Depending on the choices of the developers, the results of the area-scale analysis may be slightly different. In MountainsMap and Toothfrax, the pavement starts in one corner and paves the surface in successive bands by placing a south-west-oriented triangle and then a north-east-oriented triangle, these triangles sharing one side. At the end of each band, there is a zone that cannot be covered by triangles because there is not enough space left. Depending on which corner the pavement starts, the non-paved zone is different. This is why we applied the four-corner method to average the contributions of the four starting points. Furthermore, triangles are placed step by step: a new triangle is added based on the two shared vertices of the preceding triangle and on a third new vertex. At small scales, triangles become more and more distorted, resulting in very elongated triangles at the end of the band. Toothfrax and MountainsMap use different algorithms to deal with these elongated triangles, and they use differing precisions (rounding, floating point…) and error thresholds, resulting in small deviations that propagate.

4.2. Similarities and differences within each dataset

When surfaces are not considered individually anymore but instead grouped within treatments (diets for animals, e.g. clover or dry bamboo, and cleaning procedures for lithics, e.g. BrushNoDirt) in the two-factor model, the values calculated by both software packages usually discriminate the same treatments within each dataset. There are exceptions to this (Fig. 5, Table 2). Nevertheless, these differences concern only nine out of 60 contrasts (15 treatments pairs for four parameters), that is, 15%; when excluding R^2 (not used in functional interpretations), only four out of 60 contrasts (6.67%) are different. It could be that *epLsar* is more discrimant when calculated with Toothfrax, and *Asfc* and *HAsfc* with MountainsMap (Table 2), but this is very speculative with only the four differences.

The effect detected on *Asfc* in the three-factor model is very large (mean difference in posterior = 2.3, Fig. 4a), yet both software packages discriminate the treatments in the exact same way for all but one contrast (Fig. 5a and Table 2). The effects on *epLsar* and *HAsfc9* are much smaller (mean difference in posterior = 0.00012 and 0.029 respectively; Fig. 4b-c), but they still do influence some of the treatment discriminations (Fig. 5b-d). This means that these differences are large enough to influence the functional interpretations of the results in some cases.

4.3. Comparing *epLsar* and *NewEpLsar*

The new *epLsar* parameter (*NewEpLsar*) offered in MountainsMap's SSFA module is calculated differently from the original *epLsar* parameter. Both parameters are calculated through a length-scale analysis, done on all extracted profiles of rotated versions of the surface, by default every 5° at the 1.8 µm-

scale (i.e. the spacing of the points along these profiles) (Scott et al., 2006). The relative-length *relL* is the length of the profile at a given scale divided by the projected length (Brown, 2013). At each angle α, *relL*(α) is calculated with $\alpha = i\Delta\alpha$, with $\Delta\alpha = 5°$ by default and *i* being the index of the rotated version of the surface. Then a polar mean vector is calculated with its norm equal to *relL(i)* and its angle α. From this, cartesian coordinates are calculated for each angle:

(5) $$r_x(i) = relL(i)\cos(i\Delta\alpha)$$
(6) $$r_y(i) = relL(i)\sin(i\Delta\alpha)$$

The sum of all vectors is calculated as:

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

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

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

(8) $$\bar{r_x} = \frac{1}{N}\frac{1}{S_{relL}}\sum_{i=0}^{N-1} r_x(i)$$
(9) $$\bar{r_y} = \frac{1}{N}\frac{1}{S_{relL}}\sum_{i=0}^{N-1} r_y(i)$$

Finally:

(10) $$epLsar = \sqrt{\bar{r_x}^2 + \bar{r_y}^2}$$

In Toothfrax, the calculation apparently uses *cos(2i$\Delta\alpha$)* instead of *cos(i$\Delta\alpha$)* (see equation (5)). The parameter *epLsar* is calculated the same way in MountainsMap to make it comparable to the Toothfrax implementation, but the parameter *NewEplsar* uses the correct angle *i$\Delta\alpha$*.

This difference in calculation likely explains why *NewEpLsar* is approximately one order of magnitude higher than the values of the *epLsar* parameters from both software packages (Figs 1-3). Therefore, *NewEpLsar* data are not comparable to published *epLsar* data. Nevertheless, in most cases, it discriminates the treatments as *epLsar* does (Fig. 6 and Table 3; compare also the plots "Python_analysis/plots/statistical_model_two_factors_filter_strong/treatment_pairs_epLsar.pdf" and "Python_analysis/plots/statistical_model_neweplsar_filter_strong/treatment_pairs_NewEplsar.pdf" on Zenodo https://doi.org/10.5281/zenodo.7219855). It should be noted that the two treatments pairs of the lithic dataset (BrushDirt and RubDirt *vs.* Control) are well discriminated by *NewEpLsar*, but the diets of the sheep datasets that could be differentiated using *epLsar* (grass *vs.* clover/clover+dust and grass+dust *vs.* clover/clover+dust) cannot be separated with *NewEpLsar* (Table 3). It therefore seems that, due to its new algorithm, this new parameter can be either more or less powerful to discriminate the treatments as compared to the original *epLsar* parameter, depending on the surfaces.

4.4. Implications

These results imply that the analyses from both software packages are in general equally powerful to discriminate treatments. Although the results from both software packages look similar (Figs 1-3), our analysis demonstrates that the values from the two software packages cannot be directly compared. Neither of the software packages seems to allow more discrimination among the treatments within any dataset than the other (Tables 2-3): in some cases, Toothfrax better discriminates the treatments, while in others, MountainsMap is better on this aspect. Without calibration against a profile or surface with known (nominal) values for the parameters ("Ground truth" approach, e.g. Todhunter et al., 2017, 2019, 2020), it is not possible to determine which software package is closer to the "true" values. Unfortunately, such a calibration is currently impossible for these parameters.

Most of these differences are likely due to the corrections and/or further developments of the algorithms in MountainsMap. Such adjustments were and are necessary, but they also introduced

differences between the two software packages. Some settings can be adjusted in MountainsMap to make the results closer (but probably still not identical) to those from Toothfrax, although that does not make these results "better". There is therefore a compromise between accuracy/performance of the calculation and fidelity to Toothfrax.

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. This could prove problematic for the future of dental microwear texture analysis (DMTA) if studies using MountainsMap's SSFA module cannot use the extensive published data analyzed with Toothfrax. We therefore recommend re-analyzing raw surface data with MountainsMap (or any of its derivatives) before comparing with published Toothfrax data (see also discussion by Arman et al., 2016). Even though it is potentially computer-time-intensive, it requires little effort from analysts and seems to be currently the best option. Future updates of MountainsMap's SSFA module might tackle (some of) the discrepancies in the computation of the parameters. Nevertheless, the source of some of these discrepancies is due to corrections and developments in the algorithms and, as such, will unlikely be reverted to the original versions. Additionally, a new body of literature using the current MountainsMap's SSFA module is already being produced (see references in the introduction), and re-analysis of data will still be inevitable. It is therefore of utmost importance that original raw surfaces will be accessible for comparative studies and re-analysis in the future (see also Arman et al., 2016).

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.

5. Conclusions

The results of our Bayesian modeling showed that Toothfrax and MountainsMap's SSFA module produce significantly different values for all SSFA parameters investigated (*epLsar*, *Asfc*, *HAsfc9* and R^2); *Smfc* also seems to be different but it could not be statistically tested. This is to be expected from the differences in computation between the two software packages. Even though the discrimination power among the treatments (different diets for experimental animals and different cleaning procedures for lithics) are similar between the two software packages, the present results have important implications. Indeed, the majority of published data about dental microwear texture analysis (DMTA) has been produced by Toothfrax. If the data produced by MountainsMap are not directly comparable, analysts using MountainsMap (or any of its derivatives) cannot simply combine the data from both software packages, but instead need to re-analyze the existing raw surface data with MountainsMap if they intend to include published Toothfrax data into their analyses. This requires some (computer-)time investment. This issue comes on top of an existing lack of standardization in acquisition and processing (e.g. Arman et al., 2016; Calandra, Schunk, Bob, et al., 2019). More problematic though is the fact that very few of the published studies have made the raw surfaces available to the research community. Therefore, we urge all analysts to share their raw surfaces with every published article, either as uncompressed SUR files, or even better for reproducibility as MNT files including the processing steps, as we did here (on Zenodo https://doi.org/10.5281/zenodo.6645445). Ideally, in the future, researchers will even upload the raw surface data directly with each publication. For already published raw data, as well as for future raw data, it might be helpful to first create a central repository to host all these surfaces, their associated data, metadata and paradata, considering a spectrum of open source public copyright licenses in industry and academia (e.g. Creative Commons).

Acknowledgements

Preprint version 4 of this article has been peer-reviewed and recommended by Peer Community In Archaeology (https://doi.org/10.24072/pci.archaeo.100024)

Author contributions

François Blateyron: writing - review & editing
Konstantin Bob: methodology, formal analysis, resources, data curation, writing - original draft, writing - review & editing
Ivan Calandra: conceptualization, methodology, formal analysis, resources, data curation, writing - original draft, writing - review & editing
Andreas Hildebrandt: writing - review & editing
Gildas Merceron: conceptualization, methodology, resources, data curation, writing - review & editing
Ellen Schulz-Kornas: conceptualization, writing - review & editing
Antoine Souron: writing - review & editing
Daniela E. Winkler: conceptualization, methodology, resources, data curation, writing - review & editing

Data, scripts, code, and supplementary information availability

Further Supplementary Information available in open-access on Zenodo

- MountainsMap (ConfoMap) processing results as MNT and PDF files: https://doi.org/10.5281/zenodo.6645445
- Raw data, scripts, raw results and reports of the statistical analyses: https://doi.org/10.5281/zenodo.7219855
- The Docker image of the Python analysis: https://doi.org/10.5281/zenodo.7219884

Conflict of interest disclosure

François Blateyron is working at, and is co-owner of Digital Surf, the editor of the MountainsMap software, used in comparisons described in this paper. However, he was not involved in data acquisition and analysis.

Funding

IC has been supported within the Römisch-Germanisches Zentralmuseum – Leibniz Research Institute for Archeology by German Federal and Rhineland Palatinate funding (Sondertatbestand "Spurenlabor"). This research is publication no. 11 of the TraCEr laboratory. DEW has been funded by the European Research Foundation (ERC Project Vertebrate Herbivory, Co-Grant 681450 to Thomas Tütken) and the Japan Society for the Promotion of Science (JSPS) under a grant-395 in-aid, no. 20F20325.

The funding source had no involvement in study design; in the collection, analysis and interpretation of data; in the writing of the report; nor in the decision to submit the article for publication.

References

Ackermans NL, Winkler DE, Martin LF, Kaiser TM, Clauss M, Hatt J-M (2020) Dust and grit matter: abrasives of different size lead to opposing dental microwear textures in experimentally fed sheep (Ovis aries). Journal of Experimental Biology, 223, jeb220442. https://doi.org/10.1242/jeb.220442

Aiba K, Miura S, Kubo MO (2019) Dental Microwear Texture Analysis in Two Ruminants, Japanese Serow (Capricornis crispus) and Sika Deer (Cervus nippon), from Central Japan. Mammal Study, 44, 183–192. https://doi.org/10.3106/ms2018-0081

Allaire JJ, Xie Y, McPherson J, Luraschi J, Ushey K, Atkins A, Wickham H, Cheng J, Chang W, Iannone R (2021) rmarkdown: Dynamic Documents for R. R package version 2.11.

Arman SD, Prowse TAA, Couzens AMC, Ungar PS, Prideaux GJ (2019) Incorporating intraspecific variation into dental microwear texture analysis. Journal of The Royal Society Interface, 16, 20180957. https://doi.org/10.1098/rsif.2018.0957

Arman SD, Ungar PS, Brown CA, DeSantis LRG, Schmidt C, Prideaux GJ (2016) Minimizing inter-microscope variability in dental microwear texture analysis. Surface Topography: Metrology and Properties, 4, 024007. https://doi.org/10.1088/2051-672X/4/2/024007

Artigas R (2011) Imaging Confocal Microscopy. In: Optical Measurement of Surface Topography (ed Leach R), pp. 237–286. Springer Berlin Heidelberg, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-12012-1_11

Berlioz E, Azorit C, Blondel C, Ruiz MST, Merceron G (2017) Deer in an arid habitat: dental microwear textures track feeding adaptability. Hystrix, the Italian Journal of Mammalogy, 28, 222–230. https://doi.org/10.4404/hystrix-28.2-12048

Berlioz É, Kostopoulos DS, Blondel C, Merceron G (2018) Feeding ecology of Eucladoceros ctenoides as a proxy to track regional environmental variations in Europe during the early Pleistocene. Comptes Rendus Palevol, 17, 320–332. https://doi.org/10.1016/j.crpv.2017.07.002

Bestwick J, Unwin DM, Purnell MA (2019) Dietary differences in archosaur and lepidosaur reptiles revealed by dental microwear textural analysis. Scientific Reports, 9, 11691. https://doi.org/10.1038/s41598-019-48154-9

Bethune E, Kaiser TM, Schulz-Kornas E, Winkler DE (2019) Multiproxy dietary trait reconstruction in Pleistocene Hippopotamidae from the Mediterranean islands. Palaeogeography, Palaeoclimatology, Palaeoecology, 533, 109210. https://doi.org/10.1016/j.palaeo.2019.05.032

Blondel C, Rowan J, Merceron G, Bibi F, Negash E, Barr WA, Boisserie J-R (2018) Feeding ecology of Tragelaphini (Bovidae) from the Shungura Formation, Omo Valley, Ethiopia: Contribution of dental wear analyses. Palaeogeography, Palaeoclimatology, Palaeoecology, 496, 103–120. https://doi.org/10.1016/j.palaeo.2018.01.027

Böhm K, Winkler DE, Kaiser TM, Tütken T (2019) Post-mortem alteration of diet-related enamel surface textures through artificial biostratinomy: A tumbling experiment using mammal teeth. Palaeogeography, Palaeoclimatology, Palaeoecology, 518, 215–231. https://doi.org/10.1016/j.palaeo.2019.01.008

van den Brand T (2021) ggh4x: Hacks for “ggplot2”. R package version 0.2.1.9000.

Brown CA (2000) Issues in Modeling Machined Surface Textures. Machining Science and Technology, 4, 539–546. https://doi.org/10.1080/10940340008945721

Brown CA (2013) Areal Fractal Methods. In: Characterisation of Areal Surface Texture (ed Leach R), pp. 129–153. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-36458-7_6

Brown CA, Charles PD, Johnsen WA, Chesters S (1993) Fractal analysis of topographic data by the patchwork method. Wear, 161, 61–67. https://doi.org/10.1016/0043-1648(93)90453-S

Calandra I, Labonne G, Schulz-Kornas E, Kaiser TM, Montuire S (2016) Tooth wear as a means to quantify intra-specific variations in diet and chewing movements. Scientific Reports, 6, 34037. https://doi.org/10.1038/srep34037

Calandra I, Merceron G (2016) Dental microwear texture analysis in mammalian ecology. Mammal Review, 46, 215–228. https://doi.org/10.1111/mam.12063

Calandra I, Pedergnana A, Gneisinger W, Marreiros J (2019) Why should traceology learn from dental microwear, and vice-versa? Journal of Archaeological Science, 110, 105012. https://doi.org/10.1016/j.jas.2019.105012

Calandra I, Schulz E, Pinnow M, Krohn S, Kaiser TM (2012) Teasing apart the contributions of hard dietary items on 3D dental microtextures in primates. Journal of Human Evolution, 63, 85–98. https://doi.org/10.1016/j.jhevol.2012.05.001

Calandra I, Schunk L, Bob K, Gneisinger W, Pedergnana A, Paixao E, Hildebrandt A, Marreiros J (2019) The effect of numerical aperture on quantitative use-wear studies and its implication on reproducibility. Scientific Reports, 9, 6313. https://doi.org/10.1038/s41598-019-42713-w

Calandra I, Schunk L, Rodriguez A, Gneisinger W, Pedergnana A, Paixao E, Pereira T, Iovita R, Marreiros J (2019) Back to the edge: relative coordinate system for use-wear analysis. Archaeological and Anthropological Sciences, 11, 5937–5948. https://doi.org/10.1007/s12520-019-00801-y

Caporale SS, Ungar PS (2016) Rodent incisor microwear as a proxy for ecological reconstruction. Palaeogeography, Palaeoclimatology, Palaeoecology, 446, 225–233. https://doi.org/10.1016/j.palaeo.2016.01.013

Catz N, Bignon-Lau O, Merceron G (2020) Reindeer feeding ecology and hunting strategies by Magdalenians from Pincevent (Paris Basin, France): New insights from dental microwear textural analyses. International Journal of Osteoarchaeology, 30, 519–528. https://doi.org/10.1002/oa.2879

Caux S, Galland A, Queffelec A, Bordes J-G (2018) Aspects and characterization of chert alteration in an archaeological context: A qualitative to quantitative pilot study. Journal of Archaeological Science: Reports, 20, 210–219. https://doi.org/10.1016/j.jasrep.2018.04.027

Delezene LK, Teaford MF, Ungar PS (2016) Canine and incisor microwear in pitheciids and Ateles reflects documented patterns of tooth use. American Journal of Physical Anthropology, 161, 6–25. https://doi.org/10.1002/ajpa.23002

DeSantis LRG (2016) Dental microwear textures: reconstructing diets of fossil mammals. Surface Topography: Metrology and Properties, 4, 023002. https://doi.org/10.1088/2051-672X/4/2/023002

DeSantis LRG, Crites JM, Feranec RS, Fox-Dobbs K, Farrell AB, Harris JM, Takeuchi GT, Cerling TE (2019) Causes and Consequences of Pleistocene Megafaunal Extinctions as Revealed from Rancho La Brea Mammals. Current Biology, 29, 2488-2495.e2. https://doi.org/10.1016/j.cub.2019.06.059

DeSantis LRG, Sharp AC, Schubert BW, Colbert MW, Wallace SC, Grine FE (2020) Clarifying relationships between cranial form and function in tapirs, with implications for the dietary ecology of early hominins. Scientific Reports, 10, 8809. https://doi.org/10.1038/s41598-020-65586-w

Evans AA, Macdonald DA, Giusca CL, Leach RK (2014) New method development in prehistoric stone tool research: Evaluating use duration and data analysis protocols. Micron, 65, 69–75. https://doi.org/10.1016/j.micron.2014.04.006

Forbes AB (2013) Areal Form Removal. In: Characterisation of Areal Surface Texture (ed Leach R), pp. 107–128. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-36458-7_5

Francisco A, Blondel C, Brunetière N, Ramdarshan A, Merceron G (2018) Enamel surface topography analysis for diet discrimination. A methodology to enhance and select discriminative parameters. Surface Topography: Metrology and Properties, 6, 015002. https://doi.org/10.1088/2051-672X/aa9dd3

Galland A, Queffelec A, Caux S, Bordes J-G (2019) Quantifying lithic surface alterations using confocal microscopy and its relevance for exploring the Châtelperronian at La Roche-à-Pierrot (Saint-Césaire, France). Journal of Archaeological Science, 104, 45–55. https://doi.org/10.1016/j.jas.2019.01.009

Gill PG, Purnell MA, Crumpton N, Brown KR, Gostling NJ, Stampanoni M, Rayfield EJ (2014) Dietary specializations and diversity in feeding ecology of the earliest stem mammals. Nature, 512, 303–305. https://doi.org/10.1038/nature13622

Green JL, Croft DA (2018) Using Dental Mesowear and Microwear for Dietary Inference: A Review of Current Techniques and Applications. In: Methods in Paleoecology: Reconstructing Cenozoic Terrestrial Environments and Ecological Communities Vertebrate Paleobiology and Paleoanthropology. (eds Croft DA, Su DF, Simpson SW), pp. 53–73. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319-94265-0_5

Hara AT, Elkington-Stauss D, Ungar PS, Lippert F, Eckert GJ, Zero DT (2021) Three-Dimensional Surface Texture Characterization of In Situ Simulated Erosive Tooth Wear. Journal of Dental Research, 100, 1236–1242. https://doi.org/10.1177/00220345211005678

Hara AT, Livengood SV, Lippert F, Eckert GJ, Ungar PS (2016) Dental Surface Texture Characterization Based on Erosive Tooth Wear Processes. Journal of Dental Research, 95, 537–542. https://doi.org/10.1177/0022034516629941

Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, Wieser E, Taylor J, Berg S, Smith NJ, Kern R, Picus M, Hoyer S, van Kerkwijk MH, Brett M, Haldane A, del Río JF, Wiebe M, Peterson P, Gérard-Marchant P, Sheppard K, Reddy T, Weckesser W, Abbasi H, Gohlke C, Oliphant TE (2020) Array programming with NumPy. Nature, 585, 357–362. https://doi.org/10.1038/s41586-020-2649-2

Hoffman MD, Gelman A (2014) The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. Journal of Machine Learning Research, 15, 1593–1623.

Hofman-Kamińska E, Merceron G, Bocherens H, Makowiecki D, Piličiauskienė G, Ramdarshan A, Berlioz E, Kowalczyk R (2018) Foraging habitats and niche partitioning of European large herbivores during the Holocene – Insights from 3D dental microwear texture analysis. Palaeogeography, Palaeoclimatology, Palaeoecology, 506, 183–195. https://doi.org/10.1016/j.palaeo.2018.05.050

Højsgaard S, Halekoh U (2021) doBy: Groupwise Statistics, LSmeans, Linear Contrasts, Utilities. R package version 4.6.1.

Hunter JD (2007) Matplotlib: A 2D Graphics Environment. Computing in Science & Engineering, 9, 90–95. https://doi.org/10.1109/MCSE.2007.55

Ibáñez JJ, Lazuen T, González-Urquijo J (2019) Identifying Experimental Tool Use Through Confocal Microscopy. Journal of Archaeological Method and Theory, 26, 1176–1215. https://doi.org/10.1007/s10816-018-9408-9

International Organization for Standardization (2021) ISO 25178-2 – Geometrical product specifications (GPS) – Surface texture: Areal – Part 2: Terms, definitions and surface texture parameters.

Key AJM, Stemp WJ, Morozov M, Proffitt T, de la Torre I (2015) Is Loading a Significantly Influential Factor in the Development of Lithic Microwear? An Experimental Test Using LSCM on Basalt from Olduvai Gorge. Journal of Archaeological Method and Theory, 22, 1193–1214. https://doi.org/10.1007/s10816-014-9224-9

Krueger KL (2015) Reconstructing diet and behavior in bioarchaeological groups using incisor microwear texture analysis. Journal of Archaeological Science: Reports, 1, 29–37. https://doi.org/10.1016/j.jasrep.2014.10.002

Krueger KL, Chwa E, Peterson AS, Willman JC, Fok A, van Heel B, Heo Y, Weston M, DeLong R (2021) Technical note: Artificial Resynthesis Technology for the experimental formation of dental microwear textures. American Journal of Physical Anthropology, 176, 703–712. https://doi.org/10.1002/ajpa.24395

Krueger KL, Ungar PS (2010) Incisor microwear textures of five bioarcheological groups. International Journal of Osteoarchaeology, 20, 549–560. https://doi.org/10.1002/oa.1093

Krueger KL, Ungar PS, Guatelli-Steinberg D, Hublin J-J, Pérez-Pérez A, Trinkaus E, Willman JC (2017) Anterior dental microwear textures show habitat-driven variability in Neandertal behavior. Journal of Human Evolution, 105, 13–23. https://doi.org/10.1016/j.jhevol.2017.01.004

Krueger KL, Willman JC, Matthews GJ, Hublin J-J, Pérez-Pérez A (2019) Anterior tooth-use behaviors among early modern humans and Neandertals. PLOS ONE, 14, e0224573. https://doi.org/10.1371/journal.pone.0224573

Kruschke JK (2013) Bayesian estimation supersedes the t test. Journal of Experimental Psychology: General, 142, 573–603. https://doi.org/10.1037/a0029146

Kruschke JK (2015) Doing Bayesian data analysis: a tutorial with R, JAGS, and Stan. Academic Press, Boston.

Kubo MO, Yamada E, Kubo T, Kohno N (2017) Dental microwear texture analysis of extant sika deer with considerations on inter-microscope variability and surface preparation protocols. Biosurface and Biotribology, 3, 155–165. https://doi.org/10.1016/j.bsbt.2017.11.006

Kumar R, Carroll C, Hartikainen A, Martin O (2019) ArviZ a unified library for exploratory analysis of Bayesian models in Python. Journal of Open Source Software, 4, 1143. https://doi.org/10.21105/joss.01143

Lesnik JJ (2011) Bone tool texture analysis and the role of termites in the diet of South African hominids. PaleoAnthropology, 2011, 268–281. https://doi.org/10.4207/PA.2011.ART57

Macdonald DA, Xie L, Gallo T (2019) Here's the dirt: First applications of confocal microscopy for quantifying microwear on experimental ground stone earth working tools. Journal of Archaeological Science: Reports, 26, 101861. https://doi.org/10.1016/j.jasrep.2019.05.026

Martisius NL, McPherron SP, Schulz-Kornas E, Soressi M, Steele TE (2020) A method for the taphonomic assessment of bone tools using 3D surface texture analysis of bone microtopography. Archaeological and Anthropological Sciences, 12, 251. https://doi.org/10.1007/s12520-020-01195-y

Martisius NL, Sidéra I, Grote MN, Steele TE, McPherron SP, Schulz-Kornas E (2018) Time wears on: Assessing how bone wears using 3D surface texture analysis. PLOS ONE, 13, e0206078. https://doi.org/10.1371/journal.pone.0206078

Marwick B (2019) rrtools: Creates a Reproducible Research Compendium. R package version 0.1.5.

Marwick B, Boettiger C, Mullen L (2018) Packaging Data Analytical Work Reproducibly Using R (and Friends). The American Statistician, 72, 80–88. https://doi.org/10.1080/00031305.2017.1375986

Merceron G, Berlioz E, Vonhof H, Green D, Garel M, Tütken T (2021) Tooth tales told by dental diet proxies: An alpine community of sympatric ruminants as a model to decipher the ecology of fossil fauna. Palaeogeography, Palaeoclimatology, Palaeoecology, 562, 110077. https://doi.org/10.1016/j.palaeo.2020.110077

Merceron G, Colyn M, Geraads D (2018) Browsing and non-browsing extant and extinct giraffids: Evidence from dental microwear textural analysis. Palaeogeography, Palaeoclimatology, Palaeoecology, 505, 128–139. https://doi.org/10.1016/j.palaeo.2018.05.036

Merceron G, Novello A, Scott RS (2016) Palaeontology of the Upper Miocene vertebrate localities of Nikiti (Chalkidiki Peninsula, Macedonia, Greece) – paleoenvironments inferred from phytoliths and dental microwear texture analyses of meso-herbivores. Geobios, 49, 135–146. https://doi.org/10.1016/j.geobios.2016.01.004

Merceron G, Ramdarshan A, Blondel C, Boisserie J-R, Brunetiere N, Francisco A, Gautier D, Milhet X, Novello A, Pret D (2016) Untangling the environmental from the dietary: dust does not matter. Proceedings of the Royal Society B: Biological Sciences, 283, 20161032. https://doi.org/10.1098/rspb.2016.1032

Müller K (2020) rprojroot: Finding Files in Project Subdirectories. R package version 2.0.2.

Pedergnana A, Calandra I, Bob K, Gneisinger W, Paixão E, Schunk L, Hildebrandt A, Marreiros J (2020) Evaluating the microscopic effect of brushing stone tools as a cleaning procedure. Quaternary International, 569–570, 263–276. https://doi.org/10.1016/j.quaint.2020.06.031

Pedergnana A, Calandra I, Evans AA, Bob K, Hildebrandt A, Ollé A (2020) Polish is quantitatively different on quartzite flakes used on different worked materials. PLOS ONE, 15, e0243295. https://doi.org/10.1371/journal.pone.0243295

Percher AM, Merceron G, Akoue GN, Galbany J, Romero A, Charpentier MJ (2018) Dental microwear textural analysis as an analytical tool to depict individual traits and reconstruct the diet of a primate. American Journal of Physical Anthropology, 165, 123–138. https://doi.org/10.1002/ajpa.23337

Prassack KA, DuBois J, Lázničková-Galetová M, Germonpré M, Ungar PS (2020) Dental microwear as a behavioral proxy for distinguishing between canids at the Upper Paleolithic (Gravettian) site of Předmostí, Czech Republic. Journal of Archaeological Science, 115, 105092. https://doi.org/10.1016/j.jas.2020.105092

Purnell MA, Crumpton N, Gill PG, Jones G, Rayfield EJ (2013) Within-guild dietary discrimination from 3-D textural analysis of tooth microwear in insectivorous mammals. Journal of Zoology, 291, 249–257. https://doi.org/10.1111/jzo.12068

Purnell MA, Goodall RH, Thomson S, Matthews CJD (2017) Tooth microwear texture in odontocete whales: variation with tooth characteristics and implications for dietary analysis. Biosurface and Biotribology, 3, 184–195. https://doi.org/10.1016/j.bsbt.2017.11.004

Purnell M, Seehausen O, Galis F (2012) Quantitative three-dimensional microtextural analyses of tooth wear as a tool for dietary discrimination in fishes. Journal of The Royal Society Interface, 9, 2225–2233. https://doi.org/10.1098/rsif.2012.0140

R Core Team (2021) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Version 4.1.2.

Ranjitkar S, Turan A, Mann C, Gully GA, Marsman M, Edwards S, Kaidonis JA, Hall C, Lekkas D, Wetselaar P, Brook AH, Lobbezoo F, Townsend GC (2017) Surface-Sensitive Microwear Texture Analysis of Attrition and Erosion. Journal of Dental Research, 96, 300–307. https://doi.org/10.1177/0022034516680585

Reback J, McKinney W, Jbrockmendel, Bossche JVD, Augspurger T, Cloud P, Gfyoung, Sinhrks, Hawkins S, Klein A, Roeschke M, Tratner J, She C, Terji Petersen, Ayd W, MomIsBestFriend, Garcia M, Schendel J, Hayden A, Jancauskas V, Saxton D, McMaster A, Battiston P, Skipper Seabold, Chris-B1, H-Vetinari, Hoyer S, Kaiqi Dong, Overmeire W, Winkel M (2020) pandas-dev/pandas: Pandas 1.1.2. https://doi.org/10.5281/ZENODO.4019559

Robinet C, Merceron G, Candela AM, Marivaux L (2020) Dental microwear texture analysis and diet in caviomorphs (Rodentia) from the Serra do Mar Atlantic forest (Brazil). Journal of Mammalogy, 101, 386–402. https://doi.org/10.1093/jmammal/gyz194

Rosso DE, d'Errico F, Queffelec A (2017) Patterns of change and continuity in ochre use during the late Middle Stone Age of the Horn of Africa: The Porc-Epic Cave record. PLOS ONE, 12, e0177298. https://doi.org/10.1371/journal.pone.0177298

RStudio Team (2021) RStudio: Integrated Development for R. RStudio, Inc., Boston, MA. Version 1.4.1103.

Salvatier J, Wiecki TV, Fonnesbeck C (2016) Probabilistic programming in Python using PyMC3. PeerJ Computer Science, 2, e55. https://doi.org/10.7717/peerj-cs.55

Schauberger P, Walker A (2021) openxlsx: Read, Write and Edit xlsx Files. R package version 4.2.4.

Schmidt CW, Remy A, Sessen RV, Willman J, Krueger K, Scott R, Mahoney P, Beach J, McKinley J, D'Anastasio R, Chiu L, Buzon M, Gregory JRD, Sheridan S, Eng J, Watson J, Klaus H, Da-Gloria P, Wilson J, Stone A, Sereno P, Droke J, Perash R, Stojanowski C, Herrmann N (2019) Dental microwear texture analysis of Homo sapiens sapiens: Foragers, farmers, and pastoralists. American Journal of Physical Anthropology, 169, 207–226. https://doi.org/10.1002/ajpa.23815

Schulz E, Calandra I, Kaiser TM (2010) Applying tribology to teeth of hoofed mammals. Scanning, 32, 162–182. https://doi.org/10.1002/sca.20181

Schulz E, Calandra I, Kaiser TM (2013) Feeding ecology and chewing mechanics in hoofed mammals: 3D tribology of enamel wear. Wear, 300, 169–179. https://doi.org/10.1016/j.wear.2013.01.115

Schulz-Kornas E, Kaiser TM, Calandra I, Winkler DE (2020) A brief history of quantitative wear analyses with an appeal for a holistic view on dental wear processes. In: Mammalian Teeth - Form and Function (eds Martin T, von Koenigswald W), pp. 44–53. Verlag Dr. Friedrich Pfeil, Munich, Germany.

Schulz-Kornas E, Stuhlträger J, Clauss M, Wittig RM, Kupczik K (2019) Dust affects chewing efficiency and tooth wear in forest dwelling Western chimpanzees (Pan troglodytes verus). American Journal of Physical Anthropology, 169, 66–77. https://doi.org/10.1002/ajpa.23808

Schulz-Kornas E, Winkler DE, Clauss M, Carlsson J, Ackermans NL, Martin LF, Hummel J, Müller DWH, Hatt J-M, Kaiser TM (2020) Everything matters: Molar microwear texture in goats (Capra aegagrus hircus) fed diets of different abrasiveness. Palaeogeography, Palaeoclimatology, Palaeoecology, 552, 109783. https://doi.org/10.1016/j.palaeo.2020.109783

Scott RS, Ungar PS, Bergstrom TS, Brown CA, Childs BE, Teaford MF, Walker A (2006) Dental microwear texture analysis: technical considerations. Journal of Human Evolution, 51, 339–349. https://doi.org/10.1016/j.jhevol.2006.04.006

Scott RS, Ungar PS, Bergstrom TS, Brown CA, Grine FE, Teaford MF, Walker A (2005) Dental microwear texture analysis shows within-species diet variability in fossil hominins. Nature, 436, 693–695. https://doi.org/10.1038/nature03822

Sewell L, Merceron G, Hopley PJ, Zipfel B, Reynolds SC (2019) Using springbok (Antidorcas) dietary proxies to reconstruct inferred palaeovegetational changes over 2 million years in Southern Africa. Journal of Archaeological Science: Reports, 23, 1014–1028. https://doi.org/10.1016/j.jasrep.2018.02.009

Smith GJ, DeSantis LRG (2018) Dietary ecology of Pleistocene mammoths and mastodons as inferred from dental microwear textures. Palaeogeography, Palaeoclimatology, Palaeoecology, 492, 10–25. https://doi.org/10.1016/j.palaeo.2017.11.024

Smith GJ, DeSantis LRG (2020) Extinction of North American Cuvieronius (Mammalia: Proboscidea: Gomphotheriidae) driven by dietary resource competition with sympatric mammoths and mastodons. Paleobiology, 46, 41–57. https://doi.org/10.1017/pab.2020.7

Soler S (2022) cc-licenses: Creative Commons Licenses for GitHub Projects.

Stemp WJ, Childs BE, Vionnet S, Brown CA (2009) Quantification and Discrimination of Lithic Use-Wear: Surface Profile Measurements and Length-Scale Fractal Analysis. Archaeometry, 51, 366–382. https://doi.org/10.1111/j.1475-4754.2008.00404.x

Stemp WJ, Chung S (2011) Discrimination of surface wear on obsidian tools using LSCM and RelA: pilot study results (area-scale analysis of obsidian tool surfaces). Scanning, 33, 279–293. https://doi.org/10.1002/sca.20250

Stemp WJ, Lerner HJ, Kristant EH (2018) Testing Area-Scale Fractal Complexity (Asfc) and Laser Scanning Confocal Microscopy (LSCM) to Document and Discriminate Microwear on Experimental Quartzite Scrapers. Archaeometry, 60, 660–677. https://doi.org/10.1111/arcm.12335

Stuhlträger J, Schulz-Kornas E, Wittig RM, Kupczik K (2019) Ontogenetic Dietary Shifts and Microscopic Tooth Wear in Western Chimpanzees. Frontiers in Ecology and Evolution, 7, 298. https://doi.org/10.3389/fevo.2019.00298

Stynder DD, DeSantis LRG, Donohue SL, Schubert BW, Ungar PS (2019) A Dental Microwear Texture Analysis of the Early Pliocene African Ursid Agriotherium africanum (Mammalia, Carnivora, Ursidae). Journal of Mammalian Evolution, 26, 505–515. https://doi.org/10.1007/s10914-018-9436-y

Tanis BP, DeSantis LRG, Terry RC (2018) Dental microwear textures across cheek teeth in canids: Implications for dietary studies of extant and extinct canids. Palaeogeography, Palaeoclimatology, Palaeoecology, 508, 129–138. https://doi.org/10.1016/j.palaeo.2018.07.028

Todhunter LD, Leach RK, Lawes SDA, Blateyron F (2017) An analysis of type F2 software measurement standards for profile surface texture parameters. Measurement Science and Technology, 28, 065017. https://doi.org/10.1088/1361-6501/aa6924

Todhunter L, Leach R, Lawes S, Harris P, Blateyron F (2019) Mathematical approach to the validation of functional surface texture parameter software. Surface Topography: Metrology and Properties, 7, 015020. https://doi.org/10.1088/2051-672X/ab07ca

Todhunter L, Leach R, Lawes S, Harris P, Blateyron F (2020) Mathematical approach to the validation of field surface texture parameter software. Surface Topography: Metrology and Properties, 8, 015010. https://doi.org/10.1088/2051-672X/ab7367

Turcotte CM, Green DJ, Kupczik K, McFarlin S, Schulz-Kornas E (2020) Elevated activity levels do not influence extrinsic fiber attachment morphology on the surface of muscle-attachment sites. Journal of Anatomy, 236, 827–839. https://doi.org/10.1111/joa.13137

Ungar PS (2015) Mammalian dental function and wear: A review. Biosurface and Biotribology, 1, 25–41. https://doi.org/10.1016/j.bsbt.2014.12.001

Ungar PS, Abella EF, Burgman JHE, Lazagabaster IA, Scott JR, Delezene LK, Manthi FK, Plavcan JM, Ward CV (2020) Dental microwear and Pliocene paleocommunity ecology of bovids, primates, rodents, and suids at Kanapoi. Journal of Human Evolution, 140, 102315. https://doi.org/10.1016/j.jhevol.2017.03.005

Ungar PS, Berger LR (2018) Brief communication: Dental microwear and diet of Homo naledi. American Journal of Physical Anthropology, 166, 228–235. https://doi.org/10.1002/ajpa.23418

Ungar PS, Brown CA, Bergstrom TS, Walker A (2003) Quantification of dental microwear by tandem scanning confocal microscopy and scale-sensitive fractal analyses. Scanning, 25, 185–193. https://doi.org/10.1002/sca.4950250405

Ungar PS, Evans AA (2016) Exposing the past: surface topography and texture of paleontological and archeological remains. Surface Topography: Metrology and Properties, 4, 040302. https://doi.org/10.1088/2051-672X/4/4/040302

Ungar PS, Saylor L, Sokolov AA, Sokolova NA, Gilg O, Montuire S, Royer A (2021) Incisor microwear of Arctic rodents as a proxy for microhabitat preference. Mammalian Biology, 101, 1033–1052. https://doi.org/10.1007/s42991-021-00138-x

Ungar PS, Scott JR, Curran SC, Dunsworth HM, Harcourt-Smith WEH, Lehmann T, Manthi FK, McNulty KP (2012) Early Neogene environments in East Africa: Evidence from dental microwear of tragulids. Palaeogeography, Palaeoclimatology, Palaeoecology, 342–343, 84–96. https://doi.org/10.1016/j.palaeo.2012.05.005

Ungar PS, Zhou Z (2017) Dental biotribology: Wearing away the boundary between biology and engineering. Biosurface and Biotribology, 3, 115–118. https://doi.org/10.1016/j.bsbt.2017.11.002

Ushey K (2021) renv: Project Environments. R package version 0.14.

Van Rossum G, Drake FL (2010) The Python language reference. Python Software Foundation, Hampton, NH.

Waskom ML (2021) seaborn: statistical data visualization. Journal of Open Source Software, 6, 3021. https://doi.org/10.21105/joss.03021

Watson AS, Gleason MA (2016) A comparative assessment of texture analysis techniques applied to bone tool use-wear. Surface Topography: Metrology and Properties, 4, 024002. https://doi.org/10.1088/2051-672X/4/2/024002

Werner JJ (2018) An experimental investigation of the effects of post-depositional damage on current quantitative use-wear methods. Journal of Archaeological Science: Reports, 17, 597–604. https://doi.org/10.1016/j.jasrep.2017.12.008

Wickham H (2016) ggplot2: Elegant Graphics for Data Analysis. Springer, New York, USA.

Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019) Welcome to the Tidyverse. Journal of Open Source Software, 4, 1686. https://doi.org/10.21105/joss.01686

Wickham H, Bryan J, Barrett M (2021) usethis: Automate Package and Project Setup. R package version 2.1.3.

Winkler DE, Andrianasolo TH, Andriamandimbiarisoa L, Ganzhorn JU, Rakotondranary SJ, Kaiser TM, Schulz-Kornas E (2016) Tooth wear patterns in black rats (Rattus rattus) of Madagascar differ more in relation

to human impact than to differences in natural habitats. Ecology and Evolution, 6, 2205–2215. https://doi.org/10.1002/ece3.2048

Winkler DE, Clauss M, Rölle M, Schulz-Kornas E, Codron D, Kaiser TM, Tütken T (2021) Dental microwear texture gradients in guinea pigs reveal that material properties of the diet affect chewing behaviour. Journal of Experimental Biology, 224, jeb242446. https://doi.org/10.1242/jeb.242446

Winkler DE, Schulz-Kornas E, Kaiser TM, Codron D, Leichliter J, Hummel J, Martin LF, Clauss M, Tütken T (2020) The turnover of dental microwear texture: Testing the “last supper” effect in small mammals in a controlled feeding experiment. Palaeogeography, Palaeoclimatology, Palaeoecology, 557, 109930. https://doi.org/10.1016/j.palaeo.2020.109930

Winkler DE, Schulz-Kornas E, Kaiser TM, Cuyper AD, Clauss M, Tütken T (2019) Forage silica and water content control dental surface texture in guinea pigs and provide implications for dietary reconstruction. Proceedings of the National Academy of Sciences, 116, 1325–1330. https://doi.org/10.1073/pnas.1814081116

Winkler DE, Schulz-Kornas E, Kaiser TM, Tütken T (2019) Dental microwear texture reflects dietary tendencies in extant Lepidosauria despite their limited use of oral food processing. Proceedings of the Royal Society B: Biological Sciences, 286, 20190544. https://doi.org/10.1098/rspb.2019.0544

Winkler DE, Tütken T, Schulz-Kornas E, Kaiser TM, Müller J, Leichliter J, Weber K, Hatt J-M, Clauss M (2020) Shape, size, and quantity of ingested external abrasives influence dental microwear texture formation in guinea pigs. Proceedings of the National Academy of Sciences, 117, 22264–22273. https://doi.org/10.1073/pnas.2008149117

Xie Y (2014) knitr: A Comprehensive Tool for Reproducible Research in R. In: Implementing Reproducible Computational Research (eds Stodden V, Leisch F, Peng RD), p. . Chapman and Hall/CRC, Boca Raton, Florida.

Xie Y (2015) Dynamic Documents with R and knitr. Chapman and Hall/CRC, Boca Raton, Florida.

Xie Y (2021) knitr: A General-Purpose Package for Dynamic Report Generation in R. R package version 1.36.

Xie Y, Allaire JJ, Grolemund R (2018) R Markdown: The Definitive Guide. Chapman and Hall/CRC.

Xie Y, Dervieux C, Riederer E (2020) R Markdown Cookbook. Chapman and Hall/CRC.

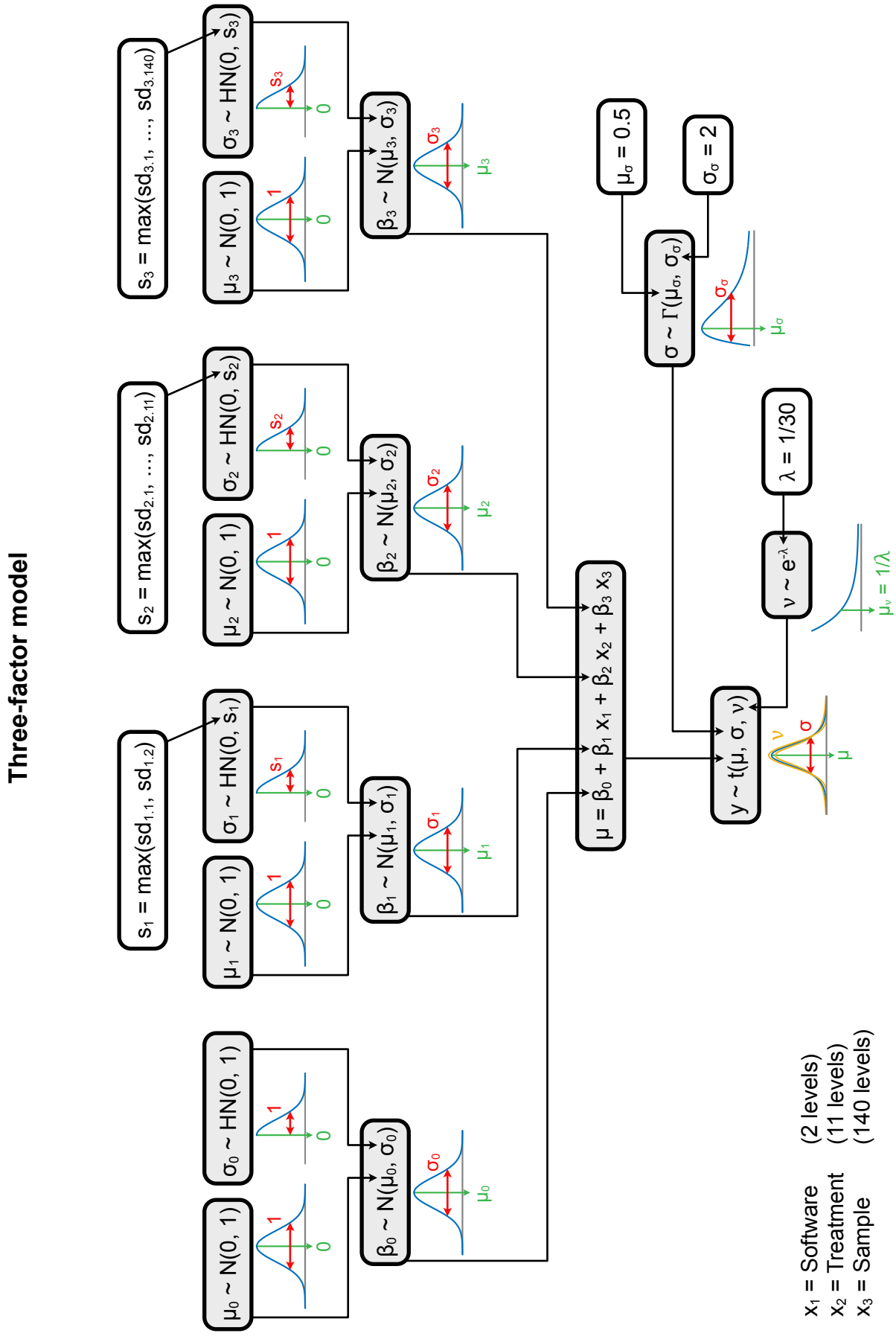
Yamada E, Kubo MO, Kubo T, Kohno N (2018) Three-dimensional tooth surface texture analysis on stall-fed and wild boars (Sus scrofa). PLOS ONE, 13, e0204719. https://doi.org/10.1371/journal.pone.0204719

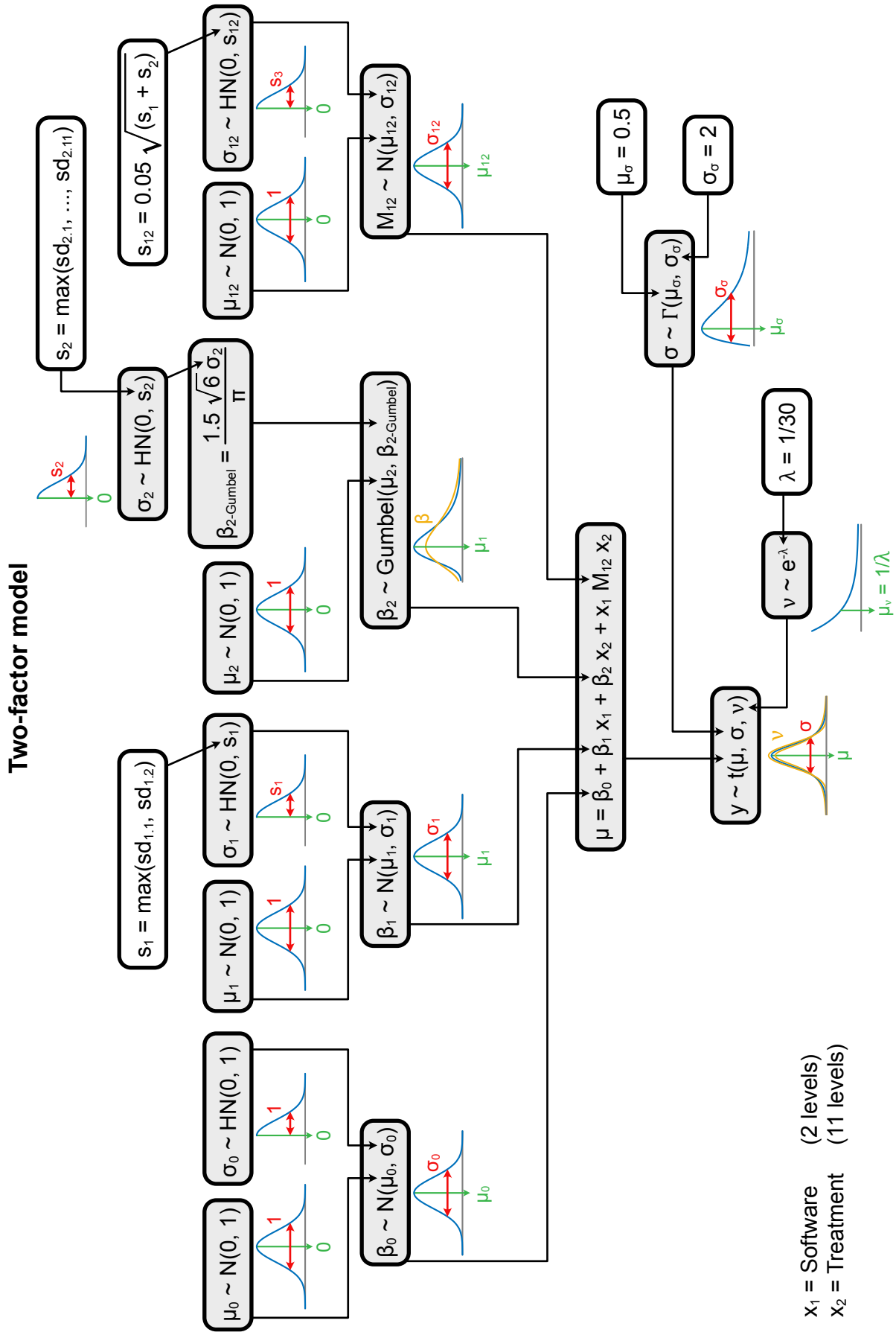
Supplementary Information 1

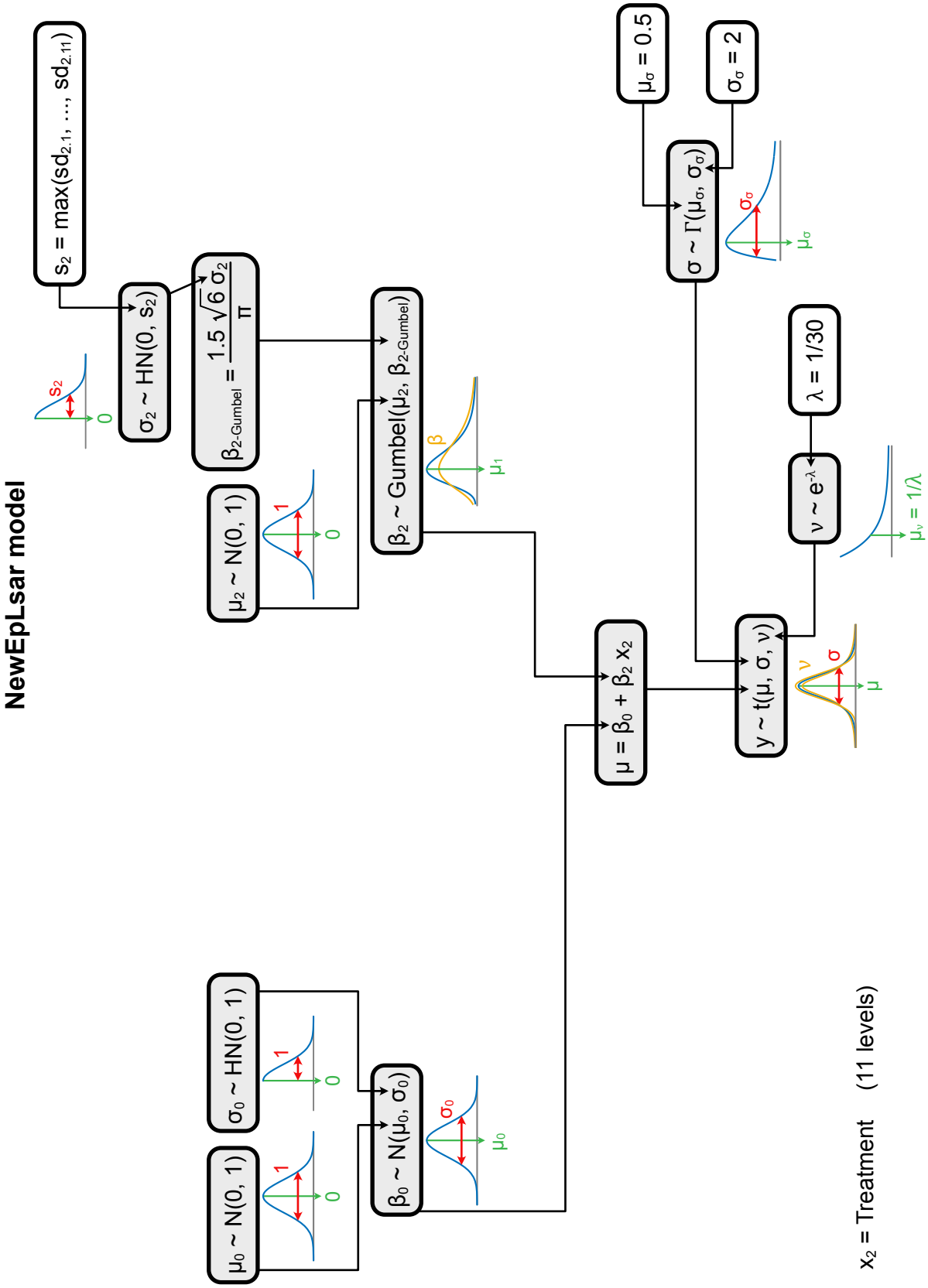
Diagrams of the Bayesian models used

Diagrams of the Bayesian models used (three-factor model, two-factor model and NewEpLsar model), similar to Kruschke's (2015) diagrams.

Γ = Gamma distribution, Gumbel = Gumbel distribution, HN = half-normal distribution, N = normal distribution, sdx.y = standard deviations when varying on factor only and sx = maxima of the sdx.y, t = Student's t distribution. The symbol "~" should be read as "is distributed as". Grey boxes enclose random variables, while white boxes are for fixed variables.


Three-factor model
$s_3 = \max(sd_{3.1}, \ldots, sd_{3.140})$
$\sigma_3 \sim HN(0, s_3)$
$\mu_3 \sim N(0, 1)$
$\beta_3 \sim N(\mu_3, \sigma_3)$
$s_2 = \max(sd_{2.1}, \ldots, sd_{2.11})$
$\sigma_2 \sim HN(0, s_2)$
$\mu_2 \sim N(0, 1)$
$\beta_2 \sim N(\mu_2, \sigma_2)$
$s_1 = \max(sd_{1.1}, sd_{1.2})$
$\sigma_1 \sim HN(0, s_1)$
$\mu_1 \sim N(0, 1)$
$\beta_1 \sim N(\mu_1, \sigma_1)$
$\sigma_0 \sim HN(0, 1)$
$\mu_0 \sim N(0, 1)$
$\beta_0 \sim N(\mu_0, \sigma_0)$
$\mu = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$
$\mu_\sigma = 0.5$
$\sigma_\sigma = 2$
$\sigma \sim \Gamma(\mu_\sigma, \sigma_\sigma)$
$\lambda = 1/30$
$\nu \sim e^{-\lambda}$
$\mu_\nu = 1/\lambda$
$y \sim t(\mu, \sigma, \nu)$
x_1 = Software (2 levels)
x_2 = Treatment (11 levels)
x_3 = Sample (140 levels)


Two-factor model
$s_2 = \max(sd_{2.1}, \ldots, sd_{2.11})$
$s_{12} = 0.05\sqrt{(s_1 + s_2)}$
$\sigma_2 \sim HN(0, s_2)$
$\beta_{2\text{-Gumbel}} = \frac{1.5\sqrt{6}\,\sigma_2}{\pi}$
$\sigma_{12} \sim HN(0, s_{12})$
$\mu_{12} \sim N(0, 1)$
$M_{12} \sim N(\mu_{12}, \sigma_{12})$
$\mu_2 \sim N(0, 1)$
$\beta_2 \sim Gumbel(\mu_2, \beta_{2\text{-Gumbel}})$
$s_1 = \max(sd_{1.1}, sd_{1.2})$
$\sigma_1 \sim HN(0, s_1)$
$\mu_1 \sim N(0, 1)$
$\beta_1 \sim N(\mu_1, \sigma_1)$
$\sigma_0 \sim HN(0, 1)$
$\mu_0 \sim N(0, 1)$
$\beta_0 \sim N(\mu_0, \sigma_0)$
$\mu = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + x_1 M_{12} x_2$
$\mu_\sigma = 0.5$
$\sigma_\sigma = 2$
$\sigma \sim \Gamma(\mu_\sigma, \sigma_\sigma)$
$\lambda = 1/30$
$\nu \sim e^{-\lambda}$
$\mu_\nu = 1/\lambda$
$y \sim t(\mu, \sigma, \nu)$
x_1 = Software (2 levels)
x_2 = Treatment (11 levels)


NewEpLsar model
$s_2 = \max(sd_{2.1}, \ldots, sd_{2.11})$
$\sigma_2 \sim HN(0, s_2)$
$\beta_{2\text{-Gumbel}} = \frac{1.5\sqrt{6}\,\sigma_2}{\pi}$
$\mu_2 \sim N(0, 1)$
$\beta_2 \sim Gumbel(\mu_2, \beta_{2\text{-Gumbel}})$
$\mu_\sigma = 0.5$
$\sigma_\sigma = 2$
$\sigma \sim \Gamma(\mu_\sigma, \sigma_\sigma)$
$\lambda = 1/30$
$\nu \sim e^{-\lambda}$
$\mu_\nu = 1/\lambda$
$\sigma_0 \sim HN(0, 1)$
$\mu_0 \sim N(0, 1)$
$\beta_0 \sim N(\mu_0, \sigma_0)$
$\mu = \beta_0 + \beta_2 x_2$
$y \sim t(\mu, \sigma, \nu)$
x_2 = Treatment (11 levels)

Supplementary Information 2

Comparison of the results from the present Bayesian analysis with those from the analyses in the original publications

The goal of the present study is not to formally compare the discrimination power among treatments between the Bayesian analysis presented here and the analyses presented in the original publications. It is nevertheless interesting to have a look at how both compare.

Sheep dataset

Merceron, Ramdarshan, et al. (2016) used two-way factorial ANOVAs (with diet and dust load as factors) coupled with post hoc tests and jackknife resampling procedures. They found that "only diet is significant" and that "the ewes fed on clover fodders have higher complexity (Asfc) and heterogeneity (HAsfc [81]) and lower anisotropy (epLsar) than the ewes fed on grass fodders" (p. 3). We have found significant differences between grass+dust and clover+dust for *HAsfc9* (MountainsMap only), *Asfc* and *epLsar* (both software packages). Additionally, grass and clover / clover+dust (Toothfrax only), and grass+dust and clover (both software packages) differ significantly on *epLsar* (see treatment pair plots in the folder "Python_analysis/plots/statistical_model_two_factors_filter_strong" and the notebook "Statistical_Model_TwoFactor_filter_strong" in the folder "Python_analysis/code/" on Zenodo https://doi.org/10.5281/zenodo.7219855). In summary, almost all differences we found are between grass / grass+dust and clover / clover+dust. While not all pairwise comparisons were significant, we found that overall ewes fed on clover fodders have higher *Asfc* and *HAsfc9* and lower *epLsar* than ewes fed on grass fodders (Fig. 1), just like Merceron, Ramdarshan, et al. (2016) did.

The differences between Merceron, Ramdarshan, et al. (2016) analysis and ours is that they used a two-factor ANOVA whereas we used a one-factor (Bayesian) ANOVA: the two grass and the two clover treatments are pooled together when testing for diet in the two-way ANOVA, but they are treated individually in the one-way ANOVA. The groups are ranked in this order: grass+dust > grass > clover+dust > clover for *epLsar*, and in the reverse order for *Asfc* and *HAsfc81* (Fig. 1). Thus, when pooling grass+dust with clover+dust and grass with clover for testing dust load in the two-way ANOVA, the values in each group are averaged. Differences are therefore expected to be weaker and can disappear, as it seems to happen here. Additionally, it seems that the values of the most extreme treatment (grass+dust) are large enough to pull the average of grass+dust and grass far enough from the other group (clover+dust and clover) in the test for diet in the two-way ANOVA. This could explain why, most of the time, only the most extreme treatment (grass+dust) is significantly different to the others in our one-way analysis.

Guinea pig dataset

Winkler, Schulz-Kornas, Kaiser, Cuyper, et al. (2019) did not apply SSFA but rather 3D surface texture analysis. However, SSFA's *epLsar* is roughly the inverse of the ISO 25178 *Str* and of the texture isotropy *IsT* parameters, and SSFA's *Asfc* is similar to the ISO 25178 *Sdr* parameter. Winkler, Schulz-Kornas, Kaiser, Cuyper, et al. (2019) applied a robust heteroscedastic Welch–Yuen omnibus test coupled with robust heteroscedastic pairwise tests. They found that guinea pigs fed dry lucerne have texture with significantly lower complexity (*Sdr*) than those fed dry bamboo. While we observe the same trend on *Asfc* (Fig. 2), the difference was not significant in our analysis (see treatment pair plots in the folder "Python_analysis/plots/statistical_model_two_factors_filter_strong" and the notebook "Statistical_Model_TwoFactor_filter_strong" in the folder "Python_analysis/code/" on Zenodo https://doi.org/10.5281/zenodo.7219855). We found that guinea pigs fed dry bamboo have significantly higher anisotropy (*epLsar*) values than those fed dry lucerne and dry grass. There is a trend of decreasing *epLsar* from dry bamboo to dry lucerne and to dry grass, but the latter two groups are not significantly different from each other. Winkler, Schulz-Kornas, Kaiser, Cuyper, et al. (2019) found the expected reverse pattern using the isotropy parameter *IsT*. With *Str*, they however found only significant differences between the dry lucerne and the dry bamboo groups; it seems that this parameter is less sensitive for these surfaces than *epLsar* or *IsT*.

Lithic dataset

Pedergnana, Calandra, Bob, et al. (2020) also applied a Bayesian multi-factor ANOVA, but using treatment and type of raw material (flint or quartzite) as factors. The model used was in general very similar, although different priors were used. There is a major difference, though: Pedergnana, Calandra, Bob, et al. (2020) ran the model on the difference in parameters between the surfaces before and after cleaning (delta values), while we ignored this factor. Additionally, the only contrasts they considered were those of the cleaning procedures (RubDirt, BrushNoDirt and BrushDirt) versus controls. They found that BrushNoDirt and BrushDirt are significantly different from controls for *epLsar*, and that BrushNoDirt significantly differs from controls for *HAsfc9*. In the present analysis, we have found significant differences only for *Asfc* between RubDirt and BrushDirt and between RubDirt and BrushNoDirt (both software packages), as well as between RubDirt and Control (MountainsMap only) (see Fig. 3 and treatment pair plots in the folder "Python_analysis/plots/statistical_model_two_factors_filter_strong" and the notebook "Statistical_Model_TwoFactor_filter_strong" in the folder "Python_analysis/code/" on Zenodo https://doi.org/10.5281/zenodo.7219855). These discrepancies are likely due to the different factors used in the models, to the use of delta values in the original study, as well as to the use of smaller surfaces with less NMPs in the present study (see section 2.1.3).