HAL
open science

Enamel surface topography analysis for diet discrimination. A methodology to enhance and select discriminative parameters

Arthur Francisco, Cécile Blondel, Noël Brunetière, Anusha Ramdarshan, Gildas Merceron

To cite this version:

Arthur Francisco, Cécile Blondel, Noël Brunetière, Anusha Ramdarshan, Gildas Merceron. Enamel surface topography analysis for diet discrimination. A methodology to enhance and select discriminative parameters. Surface Topography: Metrology and Properties, 2018, 6 (1), pp.015002. 10.1088/2051-672X/aa9dd3 . hal-03042216

HAL Id: hal-03042216

https://hal.science/hal-03042216v1

Submitted on 6 Dec 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Enamel surface topography analysis for diet discrimination.

A methodology to enhance and select discriminative parameters.

Arthur Francisco[1*], Cécile Blondel[2], Noël Brunetière[1], Anusha Ramdarshan[2], Gildas Merceron[2*]

[1] Institut Pprime, CNRS, Université de Poitiers, ISAE-ENSMA, F-86962 Futuroscope Chasseneuil, France

[2] Institut de Paléoprimatologie et Paléontologie Humaine: Evolution et Paléoenvironnements UMR 7262,

CNRS and Université de Poitiers, 86073 Poitiers Cedex 9, France

[*]corresponding authors: arthur.francisco@univ-poitiers.fr and gildas.merceron@univ-poitiers.fr

Abstract

Tooth wear and, more specifically, dental microwear texture is a dietary proxy that has been used for years in vertebrate paleoecology and ecology. DMTA, Dental Microwear Texture Analysis, relies on a few parameters related to the surface complexity, anisotropy and heterogeneity of the enamel facets at the micrometric scale. Working with few but physically meaningful parameters helps in comparing published results and in defining levels for classification purposes. Other dental microwear approaches are based on ISO parameters and coupled with statistical tests to find the more relevant ones.

The present study roughly utilizes most of the aforementioned parameters in their more or less modified form. But more than parameters, we here propose a new approach: instead of a single parameter characterizing the whole surface, we sample the surface and thus generate 9 derived parameters in order to broaden the parameter set.

The identification of the most discriminative parameters is performed with an automated procedure which is an extended and refined version of the workflows encountered in some studies. The procedure in its initial form includes the most common tools, like the ANOVA and the correlation analysis, along with the required mathematical tests. The discrimination results show that a simplified form of the procedure is able to more efficiently identify the desired number of discriminative parameters. Also highlighted are some trends like the relevance of working with both height and spatial parameters, as well as the potential benefits of dimensionless surfaces. On a set of 45 surfaces issued from 45 specimens of three modern ruminants with differences in feeding preferences (grazing, leaf-browsing and fruit-eating), it is clearly shown that the level of wear discrimination is improved with the new methodology compared to the other ones.

Keywords

Surface statistics, dental microwear, surface texture, surface anisotropy

Nomenclature and abbreviations

Surfaces

PS2, *PS8*	least square polynomial surfaces of degree two, and degree eight
S1	primary extracted surface
SA, SB, SC	cleaned *S1* surface, *SA* minus its *PS2*, *SA* minus its *PS8*

Parameters (in order of appearance in text)

*fst_, lst_, mea_, med_, ent_*	05 percentile, 95 percentile, mean, median, value on the entire surface respectively.
*max_, min_, std_, MAX_, MIN_*	ten highest and lowest value mean, standard deviation, maximum, minimum
Sa, Sp,	arithmetic mean of the absolute of the heights, highest height
Sq, Sv, Sz	height standard deviation, absolute of the smallest height, amplitude of the heights
Sku, Ssk	surface kurtosis, skewness
$f_{ACF}(tx,ty)$, *Rmax, Rmin*	autocorrelation function, f_{ACF} ellipse major(minor) axis
s, Sal, Str	height of the f_{ACF} ellipsis, *Rmin*, *Rmin/Rmax*
Sk, Smr1, Smr2, Spk, Svk	core height, end(beginning) of the hill area, hill(dale) area equivalent
Asfc	Area-Scale Fractal analysis Complexity parameter
$epLsar_{1.8}$	Exact Proportion Length–Scale Anisotropy of Relief, calculated with a scale observation of 1.8 µm
Sbc	surface box counting dimension
Sar, Sm, Smd	surface relative area, height mean, height median
Sres, Ssa, Ssb	residual of Abbott-Firestone tangent fit, Abbott-Firestone tangent fit limits
Sh	percentage of nearly-horizontal surface
Stp	ratio amplitude from 0.49 *Sz* to 0.51 *Sz*

Species

AB, AA, CS	*Alcelaphus buselaphus*, *Alces alces*, *Cephalophus silvicultor*

Statistics

α, β	risk of kind I and II respectively.
λ	Box-Cox exponent
H_0, H_1	null, alternative hypotheses
k, p, n	number of groups, number of parameters, number of individuals
LSD, HSD	Fisher's *post hoc* Least Significant Difference, Tukey's *post hoc* Honest Significant Difference
Ses	skewness standard error

1. Introduction

Dental Microwear Texture Analysis (DMTA) is a mature set of techniques with which to draw conclusions on the dietary habits of extinct species of mammals and other vertebrates. Knowledge of extant species is growing. The correlations between microwear and diet have begun to constitute a growing database, making it possible to accurately infer the diets of extinct species. Such inferences also mean inferring the ecological niche partitioning and the composition of past vegetation. This constitutes one of the final goals of DMTA .

Ungar [1] proposes a detailed history of DMTA from the guesswork at the outset all the way to the well-established and quasi-standardized methodology that uses confocal microscopy in conjunction with Scale-Sensitive Fractal Analysis, SSFA. In order to highlight the key steps converging with the aim of the present paper, only a few works from the wealth of related literature are referenced.

The actual efficient methodology of DMTA was made possible by two pioneering works. The first concerns the features observed on a tooth surface and their correlation with the diet while the second concerns 3D feature characteristics.

Walker *et al*. [2] showed a strong correlation between dental microwear and dietary habits for two species of African herbivorous mammals, the hyraxes. Tooth surface images were obtained by Scanning Electron Microscopy (SEM), making it possible to identify micrometric patterns such as microscratches, pits and polished areas. It was thus concluded that the browsing species of hyrax is associated to the presence of pits on tooth enamel facets whereas the grazing species exhibit more scratches. The latter were supposed to be due to the presence of phytoliths, hard silica particles in the grass. The presence of dust and grit on foods may also contribute, as also confirmed by Lucas *et al.* [3]. However, Merceron *et al.* [4] showed that dusts on plants do not overwhelm the dietary signals detected through DMTA. In order to quantitatively compare different species, Teaford and Walker [5] proposed common sense parameters: the surface feature number, length and width. The use of SEM coupled with a digitizer is where quantitative dental microwear analysis begins. Using the same methodology, things became clearer when Solounias *et al.* [6] proposed a ruminant dietary spectrum composed of three diet categories. These were called “browsers”, “grazers” and “intermediates”, following the ecophysiological dietary adaptations proposed by Hofmann [7]. The browsers and the grazers represented the two poles, as previously identified by Walker *et al.* [2]. It therefore became obvious that the main difficulty facing the analysis would lie in the discrimination of the intermediate range.

During this period of widespread dental microwear analysis use, Boyde and Fortelius [8] used a new technique for 3D data acquisition: scanning confocal microscopy. Having a digitized 3D surface would henceforth allow automated pattern quantification. However, a numeric tool for multiscale analysis was missing. This was first contributed by

Brown and Savary [9], with the profile fractal characteristic investigation of industrial surfaces. The fractal nature of a profile spans a certain range of scales, limited both by the acquisition device and the profile length. This work was extended to 3D surfaces by Brown *et al.* [10], which lead in some part to the well-known *Asfc* complexity parameter.

As the latter developments were not yet combined, researchers developed the SEM methodology as well as a faster and less expensive one based on stereomicroscopy, *e.g.* [11,12]. The pairing of scanning confocal microscopy with SSFA appeared with Ungar *et al.* [13] who proposed the use of Length-Scale Fractal Analysis to catch surface anisotropy, with high values for the grazer pole, along with Area-Scale Fractal Analysis for the complexity, with high values for the browser pole. Some years later, this approach would become a standard in DMTA. With Scott *et al*. [14], the advantages of this methodology were made clear. Firstly, the parameters of *Asfc* complexity, and *epLsar*(1.8µm) anisotropy proved to be efficient. Secondly, the approach no longer suffered from the low repeatability and observer biases of other techniques, whatever the quality of the protocols they had developed. The repeatability/reproducibility issue has since been quantified by different authors such as Grine *et al.* [15] and later by Mihlbachler *et al.* [16] who noted that the observer variation in the interpretation and measurement of microwear features was potentially one of the more problematic aspects of DMTA. Finally, Scott *et al*. [17] enhanced the technique of white-light confocal microscopy coupled with SSFA. The analysis utilized the two parameters related to complexity and anisotropy, as well as three more related to heterogeneity of complexity, scale of maximal complexity and textural fill volume. In current literature on DMTA, it is perhaps the most widely used methodology. Ten years on and it has become accurate to speak of a "mature technique".

In the same period, Kaiser and Brinkmann [18] proposed to study the opportunity for standardized roughness parameters as defined by the ISO 4287:1997 « Geometrical Product Specifications – Surface texture: Profile method » [19] and the ISO 4287:1984, unfortunately no longer available as it was canceled in 1998. One of the most interesting parameters, from a discriminative perspective, appeared to be the « bearing ratio », the ISO designation of which would now be « material length of profile at level *c* ». Although the study was based on surface profiles, the attempt to build a bridge between dental and industrial microwear was a valuable project with even more promising results. This kind of approach is worth pursuing as it could bring two different scientific fields closer. ISO 25178 « Geometrical Product Specifications – Surface texture: Areal » [20] has been available since 2007. Schulz *et al.* [21] "updated" Kaiser's work with 3D parameters instead of 2D parameters, using confocal microscopy. In addition, the common SSFA parameters were also used and, in so doing, the field of investigation of microwear was enlarged. Later, other researchers have also used the ISO 25178 parameters with successful discrimination [22–26], with the aim to enhance the parameter dataset and to extract better discriminative parameters, as in the present study. Among the different microscope devices, white-

light confocal microscopy coupled with SSFA has become more widely used by researchers, who have found it to be a reliable methodology with meaningful results, *e.g.* [27,28]. However, Scott [28] clearly showed with box-whisker plots that dental microwear analysis can still be improved. If SSFA is a valuable tool for discriminating between different marked diets, it is still difficult to discriminate between intermediate diets when using small samples. Therefore, the greater the number of relevant parameters, the greater the chances are to identify accurate ones, able to capture subtle differences in diets. This final stage does not yet seem to have been reached and therefore this work aims to investigate how new parameters could be automatically selected from a large set. To our knowledge, the statistical tests are the only means for the automatic selection of parameters that are able to discriminate different groups, related to different diets. It must be recalled that some researchers have also extended the usual dental microwear analysis parameter set, along with statistical tests to eliminate the irrelevant parameters [21–26] but the work proposed here brings some novelties. The latter are here numbered with a hash. The statistical tests are built on mathematical assumptions, about the parameter distributions, that need to be fulfilled. As some of the assumptions are very strict, workarounds are also proposed, #1. Different kinds of parameters are proposed here, inspired by SSFA or industrial roughness habits and others which are more exotic. Some of them are built to be highly correlated (having a linear relationship with each other), other parameters are built to be random: the set of parameters needs to be as general as possible.

In order to get the most from a given parameter, each numeric surface to be studied is sub-sampled. The parameter is then not only calculated for the entire surface but also for every single subsurface. Thus, different statistics are taken into account, subsequently enriching the set of parameters, #2. Because of the high amount of generated parameters, we propose here a procedure that selects the best discriminative parameters. Such a procedure has then to be automatic with as few arbitrary choices as possible. The core of the first procedure is the ANalysis Of VAriance (ANOVA), which is the most utilized discriminative tool. It is embedded in a global treatment chain that aims to be the most efficient as possible. Although most of the treatments are widely used [21–26], the procedure remains original because it enhances the mathematical treatments and it adds pragmatism with workarounds. Indeed, a statistical test remains a theoretical tool that can be overly strict.

2. Material and methods

2.1. Surface preparation

The dental facets of modern ungulates, including the African grazing hartebeest antelope (*Alcelaphus buselaphus*), the European leaf-browsing moose (*Alces alces*) and the African fruit-browsing duiker (*Cephalophus silvicultor*) constitute the set of surfaces on which the parameters are tested, see also [29,30]. Solounias and Semprebon [11] also tested these

species. No excessive overlap in the parameters between browsers and grazers is expected, unlike the duikers, known to span almost the entire dietary spectrum. Molds are made using a polyvinylsiloxane elastomer (Regular Body President, ref 6015 - ISO 4823, medium consistency, polyvinylsiloxane addition type; Coltene Whaledent). This product is known to be the most efficient one to replicate a given surface at fine scales [31,32].

The tooth surfaces are scanned using the Leica DCM8 optical profiler (Leica Microsystems) with a 100X magnification lens (Numerical Aperture = 0.9; Working Distance = 0.9 mm, Lateral Resolution up to 140 nm, Vertical Resolution up to 2 nm), a confocal profilometer using white light confocal technology. The sample dimension is 280×200 µm, which yields 2168×1555pt images. The primary extracted surface, *S1*, may exhibit artifacts, sometimes due to matter residuals trapped on the real surface and sometimes due to the device itself. A simple procedure is used to clean *S1* of abnormal peaks, yielding an outlier-free surface *SA* (see electronic supplementary material – ESM – for details, §1.2).

It is considered that the geometry of the tooth is nearly a second order polynomial *PS2* within the sample limits. Thus the relief due to wear is better visualized on a surface *SB* from which its polynomial surface *PS2* has been subtracted, Fig.1.a). However, the surface may have spalls with superimposed scratches. This is why a higher order polynomial *PS8* should also be subtracted; the authors have found that an 8th order is a convenient degree for that purpose (Fig.1.a)). It should be noted here that subtracting polynomial surfaces is of common use even if the degrees are usually below 8. The polynomial surfaces are determined using highly classical least square approximations. Thus, for each sample, two surfaces *SB* and *SC* are analyzed; they are the result of the filtered surface *SA* minus the polynomial surfaces *PS2* and *PS8* respectively. *SA* is not used for analyzes because the dental macro geometry hides the relief due to wear for most of the parameters, *e.g.* the height statistics.

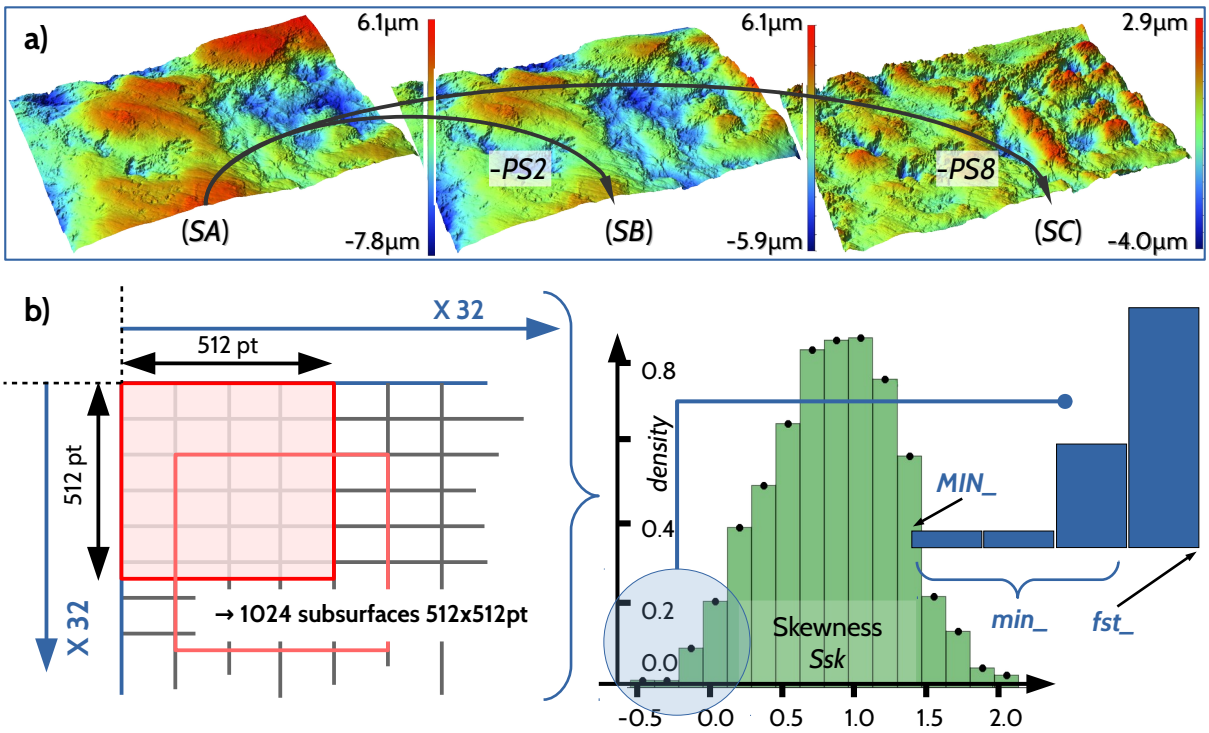

Figure 1: Surface preparation and sampling. a) The outlier-free surface *SA* becomes *SB* after its *PS2* (2nd order polynomial surface) removal, and *SC* after its *PS8* (8th order polynomial surface) removal. b) Each *SB* and *SC* surface is uniformly sampled according a 32×32 grid; statistics are provided on the parameters that are calculated for each 512×512 subsurface. As an example the figure illustrates the skewness parameter distribution resulting from the 1024 subsurfaces and three statistics related to the low skewness values.

2.2. Surface sampling

To take into account the surface heterogeneity, parameters are calculated in different parts of the surface. These are related to surface heights, wavelengths, etc. A sampling procedure is carried out so that 32 surfaces of 512×512pt are extracted along the two directions, as illustrated on Fig.1.b). Thus, with a total of 1024 results, it becomes possible to compile statistics such as: the mean (prefix *mea_*), the median (*med_*), the 5th percentile (*fst_*), the 95th percentile (*lst_*), the minimum (*MIN_*), the maximum (*MAX_*), the standard deviation (*std_*), the ten lowest values mean (*min_*) and the ten highest values mean (*max_*), making a total of nine statistics per parameter. Fig.1.b) illustrates how additional parameters are built, starting from *Ssk*. 1024 *Ssk* values are computed making it possible to define *MIN_Ssk*, *min_Ssk* and *fst_Ssk*, as shown in the zoomed part.

The choice of overlapping the samples is justified firstly by the interest in capturing the extent of the studied parameter variability. This is followed by the need to generate strong correlations in order to test the robustness of the parameter selection process. It is true that the borders of the surfaces are less represented than the mid-zone, but this will be studied in a future work. For the sample size, as analyzed by Ramdarshan *et al.* [33], sampling using 50×50µm surfaces allows for diet discrimination. An intensive use of Fast Fourier Transforms is required for wavelength analysis, hence

choosing powers of two dimensions (512×512pt = 66×66µm) speeds up the process. As a result, a satisfactory trade-off is found between the sample dimension and the computing time.

A seen above, for each parameter nine statistics are calculated. The entire surface parameters (prefix *ent_*) is also added to the set of parameter statistics, which now represents a total of ten values for *SB* and ten values for *SC*, for each parameter.

2.3. Surface parameters

The international standard ISO 25178 deals with the specification and the measurement of 3D surface textures. ISO 25178-2 specifies terms, definitions and parameters for the determination of surface texture by areal methods. Variations in dietary habits between the three species used as models here are reflected in their dental microwear textures. Among the parameters able to characterize the differences in these textures, some fall within the scope of the ISO 25178.

2.3.1. Height parameters

- The arithmetical mean height, *Sa*, is the arithmetic mean of the absolute of the heights
- The maximum peak height, *Sp*, is the highest
- The root mean square height, *Sq*, is the height standard deviation
- The maximum pit height, *Sv*, is the absolute of the smallest height
- The maximum height, *Sz*, is the amplitude of the heights.

These parameters are expected to be higher for marked scratches and splinter surfaces. Conversely, after a long and soft abrasion, the parameters may exhibit smaller ranges. The latter process occurs with tiny hard particles, and/or soft food: it lowers the peaks and does not create pits.

- The skewness, *Ssk*, quantifies the asymmetry of the height distribution. When the top of a surface has been worn out, *Ssk* deviates significantly from 0 (towards negative values)
- The kurtosis, *Sku*, measures how tailed a distribution is. If the surface is polished, many more heights close to the top fall within the same range, while the valleys remain unchanged. As a consequence, *Sku* deviates positively from its “central” value, *i.e.* three for a Gaussian distribution.

Because of their mathematical form, *Ssk* and *Sku* should be efficient indicators for soft abrasive diets, with higher absolute value range.

2.3.2. Spatial parameters

The autocorrelation function $f_{ACF}(tx,ty)$ is calculated to assess the surface anisotropy; it is provided by most image processing software. The reader is referred to the ESM §1.3 for details about the f_{ACF} related parameters.

- The autocorrelation length, *Sal*, is the horizontal distance of the $f_{ACF}(tx,ty)$ which decays the fastest to a specified value *s* between 0 and 1, default value *s*=0.2 (ISO 25178-3).

Because some surfaces exhibit long wavelengths, the default value is a bit low. In addition, to increase its robustness *Sal* is redefined as the average of *Sal* for *s*=0.3, 0.4 and 0.5

- The intersection of the autocorrelation function with the plane *z=s* is close to an ellipsis, of which the minor and major axes are respectively *Rmin=Sal* and *Rmax*.
- The ratio *Str=Rmin*/*Rmax* is a parameter for which low values characterize strong anisotropy.

Along with high *Sal* values, low *Str* values occur with large scratches, which are long wavelength patterns; a directional pattern decreases *Str* and the longer wavelengths, the greater *Sal* and *Rmax*.

Rmax appears to be at least as discriminant as the conventional *epLsar*1.8µm, as shown on Fig.2.a) Furthermore, a much better classification, Fig.2.b), can be attained with the pair of parameters (*Rmax*, *Ssk*) than with the traditional pairing (*epLsar*, *Asfc*). As the two parameters are markers for the anisotropy and with the autocorrelation being a widespread function, *epLsar*1.8µm was not retained in the present work.

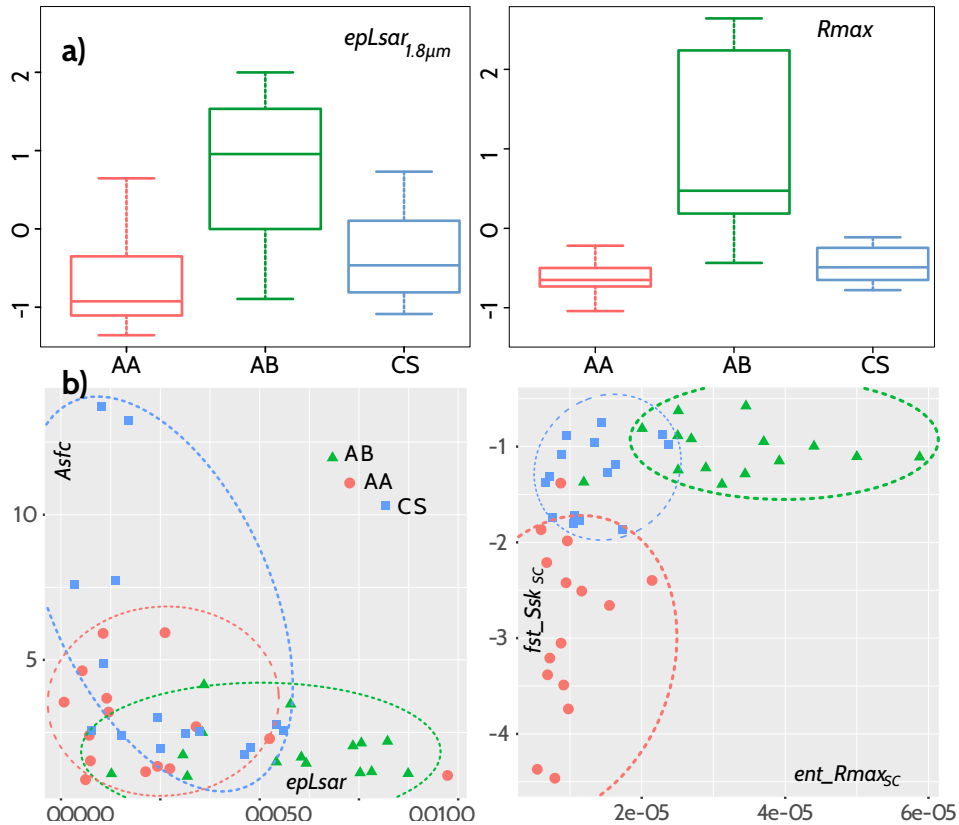

Figure 2: The present methodology clearly enhances the classical SSFA diet discrimination. a) Box plots of the parameters *epLsar*$_{1.8\mu m}$ and *Rmax*. b) Biplots (*epLsar*, *Asfc*) and (*Rmax*, *Ssk*) showing better performance of the latter.

2.3.3. Functions and related parameters

The "areal material ratio function of the scale-limited surface", also known as the Abbott-Firestone curve, describes how the matter is vertically spread, ESM Figs.S3.a)-b).

- The "areal material ratio" is defined at two specific heights: the end of the hill area, parameter *Smr1*, and the beginning of the dale area, parameter *Smr2*, ESM Fig.S3.a).
- *Spk* and *Svk* are parameters determined so that *A1* and *A2* areas are equivalent to the hill area and the dale area respectively. ESM Fig.S3.b). The equivalence makes the parameters less sensitive to surface outliers.
- The parameter *Sk* is the core height, ESM Fig.S3.b)

The way that surface relief influences the areal material ratio parameters is somewhat fuzzier. However, it can be said that asymmetric height distributions should yield different values for *Smr1* and *Smr2* and that polished surfaces should have low *Smr1* and *Spk*.

2.3.4. Fractal methods, relative area analysis

For an ideal fractal surface, self-similarity occurs at each scale of observation: the function log(relative area)=f(log(element area)) is linear, [20].

- The area-scale fractal analysis complexity parameter (*Asfc*) was introduced by Brown [10] and later defined as a thousand times minus the line slope, ESM Fig.S3.c)

At best, *Asfc* discriminates between dental facets that have been subjected to crushing actions on hard food items from others. Soft diets are more likely to be tracked by low *Asfc* values.

2.3.5. Fractal methods, box-counting dimension

- The surface box counting dimension (*Sbc*) is defined as the curve slope absolute value of the function log(*Ns*)/log(*s*) where *Ns* is the number of square boxes of size *s* needed to cover the structure, ESM Fig.S3.d)

2.3.6. Additional parameters

- The height mean, *Sm*, the height median, *Smd*, the surface relative area, *Sar* are added to the ISO 25178 parameters.

Because of the high exponent in their formulas, *Ssk* and *Sku* are sensitive to important heights. If outliers remain despite the filtering procedure, *Ssk* and *Sku* values can be erroneous. A means to overcome this drawback is to capture most of the Abbott-Firestone curve with a well-suited function. As shown by Francisco and Brunetiere [34] the function

$\tan_{a,b}(x)=\tan(-\frac{\pi}{2}(1-a)+\frac{\pi}{2}(2-a-b)x)$ fits material ratio curves well. Furthermore the $\tan_{a,b}$ function only depends on two parameters *a* and *b*, which are strongly linked to *Ssk* and *Sku*, ESM Fig.S3.e)

- *Ssa=a* and *Ssb=b* are therefore chosen as surface parameters. The same goes for *Sres*, defined as the least square residual after the $\tan_{a,b}$ fit.
- *Sh* is the percentage of nearly-horizontal surface, ESM Fig.S3.f)
- *Stp* is defined as the amplitude of the heights comprised between two values, close to the material core middle, set at 0.49*Sz* and 0.51*Sz* respectively, ESM Fig.S3.g). It is influenced by the presence of flat zones.

Ssa and *Ssb* are closely related to the material ratio curve: *Ssa* increases with deep pits and *Ssb* increases with high peaks. As an example, if there exist a few deep pits, *Ssa* is close to 1 and if the surface exhibits a large plateau, *Ssb* may be between 0 and 0.5.

A total of 24 parameters has been presented. The eight parameters related to material distribution (*Ssa*, *Ssb*, *Sres*, *Sk*, *Smr1*, *Smr2*, *Spk*, and *Svk*) are not used in the sampling procedure, but only for the entire surface, for computation speed reasons. Thus, 16 parameters are used for sampling on both *SB* and *SC*, *i.e.* the cleaned initial surface minus its 2nd and 8th degree polynomial respectively, which yields 16×10×2=320 parameters, because of ten statistics for each parameter. Eight parameters are only calculated on the whole surfaces *SB* and *SC*, which accounts for 16 additional calculations. Thus, the study involves a total of 336 parameters.

3. Results

3.1. The conservative ANOVA-based procedure

3.1.1. The pre-ANOVA tests

The ANOVA general principles are recalled in ESM §2.1-2 along with a discussion on the statistical power of the present analysis. Because of the criterion built to decide whether there are at least two different groups, some a priori assumptions must be verified:

1. the data are independent within and across groups,
2. parameter values are assumed to be normally distributed within each group,
3. the groups have the same variances,
4. the absence of "severe" outliers.

The risk in violating the ANOVA *F*-test assumptions is to inflate the Type I error rate – pinpointing differences where there are none. As a consequence the null hypothesis can be spuriously rejected. As pointed out by Osborne [35] p324, a great number of studies dealing with the effects of the assumption violation have been carried out and have shown that the *F*-test is robust for some but not all assumption violation situations. The following conclusions are widely admitted:

1. The independence assumption is critical, but is considered to be true here because the samples come from different individuals.
2. The non-normality of the parameter distributions is not considered to be problematical unless the distributions are highly skewed or the sample sizes are different.

   Bulmer [36] p63 defines as "highly skewed", distributions which skewness statistic $|Ssk| \geqslant 1$. However some statisticians define as $|Ssk| \geqslant 2$ a rule of thumb to consider the distribution too skewed. Another way of defining the normal limits as proposed by others is to have a skewness statistic below twice the skewness standard error *Ses*: $Ses = \sqrt{\frac{6}{n}}$ [37] p166, or, in its unbiased expression $Ses = \sqrt{\frac{6n(n-1)}{(n+1)(n-2)(n+3)}}$ [38] p76. The latter criterion is more rigorous because it comes from a two-tailed test for a null skewness at the 5% significance level. Hence, for n=15, *Ssk* must not exceed 2x0.58=1.16 for the unbiased formula. This simple criterion is quite convenient for an automated procedure but the risk is that multimodal distributions cannot be detected.
3. With regards to variance homogeneity, or homoscedasticity, different rules can be found. If the sizes of each sample are equal, ANOVA is mostly robust when it comes to the violation of the homogeneity of the variances when the ratio from largest sample variance to the smallest is no more than 3, or 2 for the standard error [39] p518. However, it is important to note that simulations of such a situation, described in [40] p131, show that the 5% confidence level is doubled.

4. The *F*-test is not resistant to extreme outliers.

To increase the parameter eligibility to the ANOVA assumptions, the parameters are systematically Box-Cox transformed #3 – see ESM §2.3 for details of the transformation. It is worth noting that systematically log-transforming the data can decrease the data normality: the transformations are sometimes unnecessary and other times, power transformations yield better results. If no transformation is required, the Box-Cox function leaves the data unchanged. From a starting set of 336 parameters, of which only 90 untransformed passed the assumption tests, 150 parameters were successfully transformed with the Box-Cox procedure. The Box-Cox transformation is therefore carried out prior to any statistical test on normality or variance homogeneity. If a parameter natively has normality/homoscedasticity properties, the Box-Cox transformation leaves it unchanged.

Although many statisticians advise checking graphically for normality with *Q*-*Q* plots and variance homogeneity with box plots, the only way for a procedure to be automated is to rely on statistical tests. The authors are aware of the risk in erroneously eliminating a parameter (Type II error – not seeing true differences) or in keeping a non-significant parameter (Type I error), specifically here since the data sample size is 15. It is, however, considered the price to pay for an automated procedure. The elements that motivate the choice of the following tests are presented in ESM §2.4. The normality is checked with Shapiro-Wilk's test (SW). In the case of rejection, the normality rule of thumb is used. If the parameter passes the SW test, its homoscedasticity is checked with Bartlett's test but if the parameter is nearly normal, Levene's test is used instead. If a normal parameter fails Bartlett's test, the homoscedasticity rule of thumb is used. The standard ANOVA *F*-test is then carried out on homoscedastic parameters, whereas Welch's modified ANOVA *F*-test is carried out on nearly homoscedastic parameters.

3.1.2. The post-ANOVA tests

If outlying parameter values are identified, according to the criteria presented in the ESM §2.4 #4, it is risky to automatically consider them erroneous because, for a given surface, very few outlying values occur across the numerous parameters. The 45 surfaces were checked (the species they belong to, the sizes, their qualitative appearance, etc.) but a given parameter can yield an "abnormal" value on a surface. This does not mean that the whole surface is "abnormal", unless multiple parameters qualify it as such. The fact remains that their influence in the discriminating process can lead to a spurious group mean determination. Therefore, depending on requirements, the authors propose the removal of outlier values from the dataset during this latter state.

About 43% of the initial parameters, or 146 out of 336, are removed because they do not meet the normality/homoscedasticity assumptions. About 21% of the remaining parameters, or 40 out of 190, are found to be non-discriminative after the ANOVA *F*-test. Therefore, starting from a set of 336 parameters, the subsequent analysis is

performed on 45% of them, or n=150 (details in ESM Fig.S5). It is worth noting that because of Type I errors ("false positives"), fewer parameters may be really helpful for the group identification.

Scientists use to consider highly correlated parameters as being redundant because they can be expressed as a function of other, less correlated ones. Consequently, a correlation procedure was added, #5,to identify the highest parameter correlations. The parameter correlation matrix is therefore calculated, the mean absolute correlation of each parameter is determined and the parameter with the largest value is removed. The process is repeated until no mean absolute correlation exceeds a given cutoff; at this point, the number of parameters that exceed the cutoff is known. This elimination procedure is performed for cutoffs from 0.1 to 1.0 by steps of 0.02, Fig. 3. Because of the curve shape and the fact that the cutoffs are uniformly spaced, the abscissa points are sparsely spread at the beginning of the curve, then accumulate at the end. Fig.3 details how many parameters are excluded above a 95% cutoff.

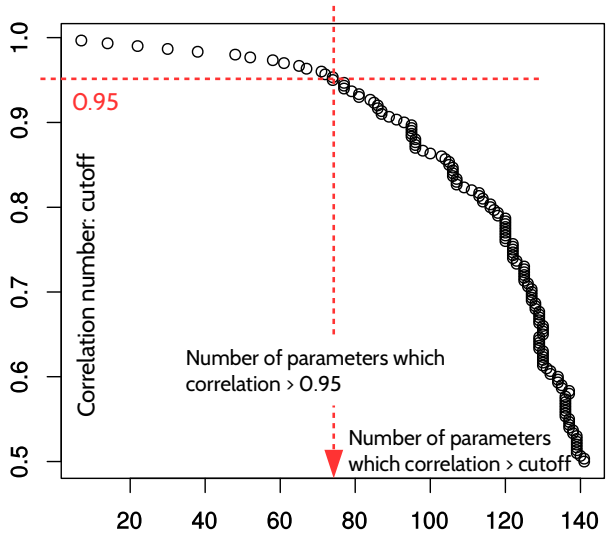

Figure 3: Number of parameters with a correlation above the cutoff value 95%

There is no universal upper limit for the cutoff value in order to maximize discrimination process efficiency. Visually, it appears that below the 95% value too many parameters are involved; hence 0.95 is the chosen cutoff. About 50% of the parameters are thus removed, the final set length being 74.

As seen above, the ANOVA F-test has put aside the parameters that made no significant difference between the three groups and now there are no more nearly collinear parameters. The next question to answer is which group means are different, because the ANOVA response is simply that there is likely to be a difference between the groups. A '*post hoc*' pairwise analysis is performed for that purpose. Cohen [41] p403, details which pairwise comparison procedure to choose, depending on the dataset type. For three groups of equal size for which variances are not too different, the straightforward analysis is the so-called Fisher's LSD test (LSD: Least Significant Difference). It is known to be the

less conservative test, seeing differences where it should not. Among the most popular of the more conservative tests, there is Tukey's HSD test (HSD: Honest Significant Difference), which appears to be "overly conservative". As there are only three groups for pairwise comparisons, the LSD test is used in the present study (ESM §2.5 for details).

Because there are three groups, the authors propose that nine parameters be first chosen among the smallest *p*-values (three parameters for each group). It should be then possible to identify trends in the selected parameters regarding the physical quantities, the surfaces (*SB* or *SC*) and the better statistics (*min_*, *fst_*, *med_*, etc.). In a second step, only three parameters are kept for classification purposes.

When calculating the *p*-values, it appears that, because of lower values for the AA-AB pair, the group that is the most difficult to identify among the 45 data is the third one, CS Table 1). This is not surprising because CS patterns range across the typical patterns of both grazers and browsers. Therefore, the three parameters yielding the lowest LSD *p*-values for the AB-CS pair are the first to be retained. The same follows for the AA-CS pair and then for the AB-AA pair. The global theoretical scheme explaining why low *p*-values are selected is illustrated in Fig.4. The nine 'best' parameters are listed in Table 1 with their pairwise test *p*-values.

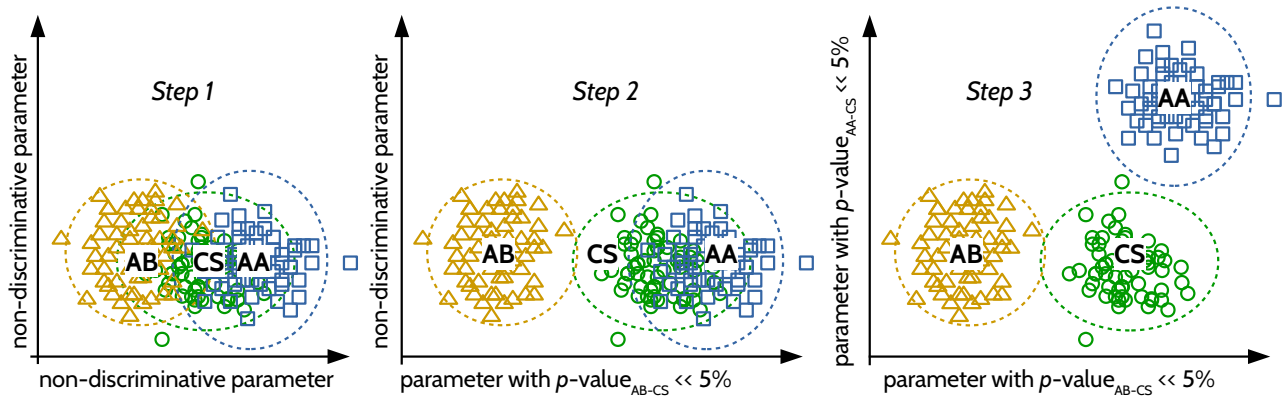

Figure 4: How low *post hoc p*-values drive three hypothetical groups (AA, AB, CS) back along an axis. As CS is the more problematic group, the first chosen parameter is the one with the lowest AB-CS *p*-value. The second one is the parameter with the lowest AA-CS *p*-value. To complete the Top3 set, another parameter with the lowest AB-AA *p*-value is added (not represented here)

parameter	physical meaning	type	AB-AA	AB-CS	AA-CS
ent_Rmax$_{SC}$	*acf* ellipsis major axis value calculated on the entire *SC* surface	spatial	3.76E-11	3.81E-08	3.79E-02
fst_Rmax$_{SC}$	*acf* ellipsis major axis 5-percentile value of the sampled *SC* surface	spatial	8.27E-12	4.41E-08	1.05E-02
fst_Str$_{SC}$	*acf* ellipsis axis ratio 5-percentile value of the sampled *SC* surface	spatial	1.40E-08	4.71E-08	~~7.15E-01~~
fst_Ssk$_{SC}$	skewness 5-percentile value of the sampled *SC* surface	height	2.38E-12	~~6.16E-02~~	9.73E-10
min_Sp$_{SC}$	highest height ten lowest value mean of the sampled *SC* surface	height	3.43E-06	~~2.56E-01~~	7.63E-08
ent_Ssk$_{SC}$	skewness value calculated on the entire *SC* surface	height	5.82E-09	~~3.37E-01~~	1.42E-07
med_Sku$_{SC}$	kurtosis median value of the sampled *SC* surface	height	8.41E-10	2.03E-03	4.06E-05
mea_Rmax$_{SC}$	*acf* ellipsis major axis mean value of the sampled *SC* surface	spatial	9.37E-10	6.28E-07	~~5.44E-02~~
min_Rmax$_{SC}$	*acf* ellipsis major axis ten lowest value mean of the sampled *SC* surface	spatial	1.02E-09	5.41E-06	1.26E-02

Table 1 – *Post hoc* results, the nine lowest *p*-values, choosing first AB-CS, then AA-CS, then AB-AA. A crossed *p*-value (above 5%) means non-significant differences

It is worth noting that:

- there are as many height parameters as spatial parameters,
- height parameters are well suited for the group AA discrimination (low AB-AA and AA-CS *p*-values),
- spatial parameters are well suited for the group AB discrimination (low AB-AA and AB-CS *p*-values),
- only two parameters out of nine belong to the whole surface parameters,
- the parameters are all built on the SC surface, the outlier-free surface minus its eighth order polynomial surface,
- the outliers (parameter values considered as abnormal), that have been previously identified, are not present in the Top9 parameter set.

This way of choosing the parameters can appear rather uncommon and subject to discussion. However, other ways were explored:

- calculating the maximum *p*-value in a row and choosing the lowest values,

- averaging the *p*-values of (AB-AA, AB-CS) for the group AB, (AB-AA, AA-CS) for the group AA and (AB-CS, AA-CS) for the group CS
- etc.

However, none of these fared any better than the proposed method.

The last step in reducing the number of parameters consists in keeping only the group parameters that correlate the least to the two others (Table 2) #6. The goal therefore is to have a single representative parameter for each group, which yields a Top3 parameter set.

parameter	**physical meaning**	**type**	**AB-AA**	**AB-CS**	**AA-CS**
fst_Rmax_{SC}	*acf* ellipsis major axis 5-percentile value of the sampled *SC* surface	spatial	8.27E-12	4.41E-08	1.05E-02
min_Sp_{SC}	highest height lowest value of the sampled *SC* surface	height	3.43E-06	~~2.56E-01~~	7.63E-08
mea_Rmax_{SC}	*acf* ellipsis major axis mean value of the sampled *SC* surface	spatial	9.37E-10	6.28E-07	~~5.44E-02~~

Table 2 – The three less correlated parameters that best identify the groups. A crossed *p*-value (above 5%) means non-significant differences

An overview of the full ANOVA procedure is illustrated in Fig.5.

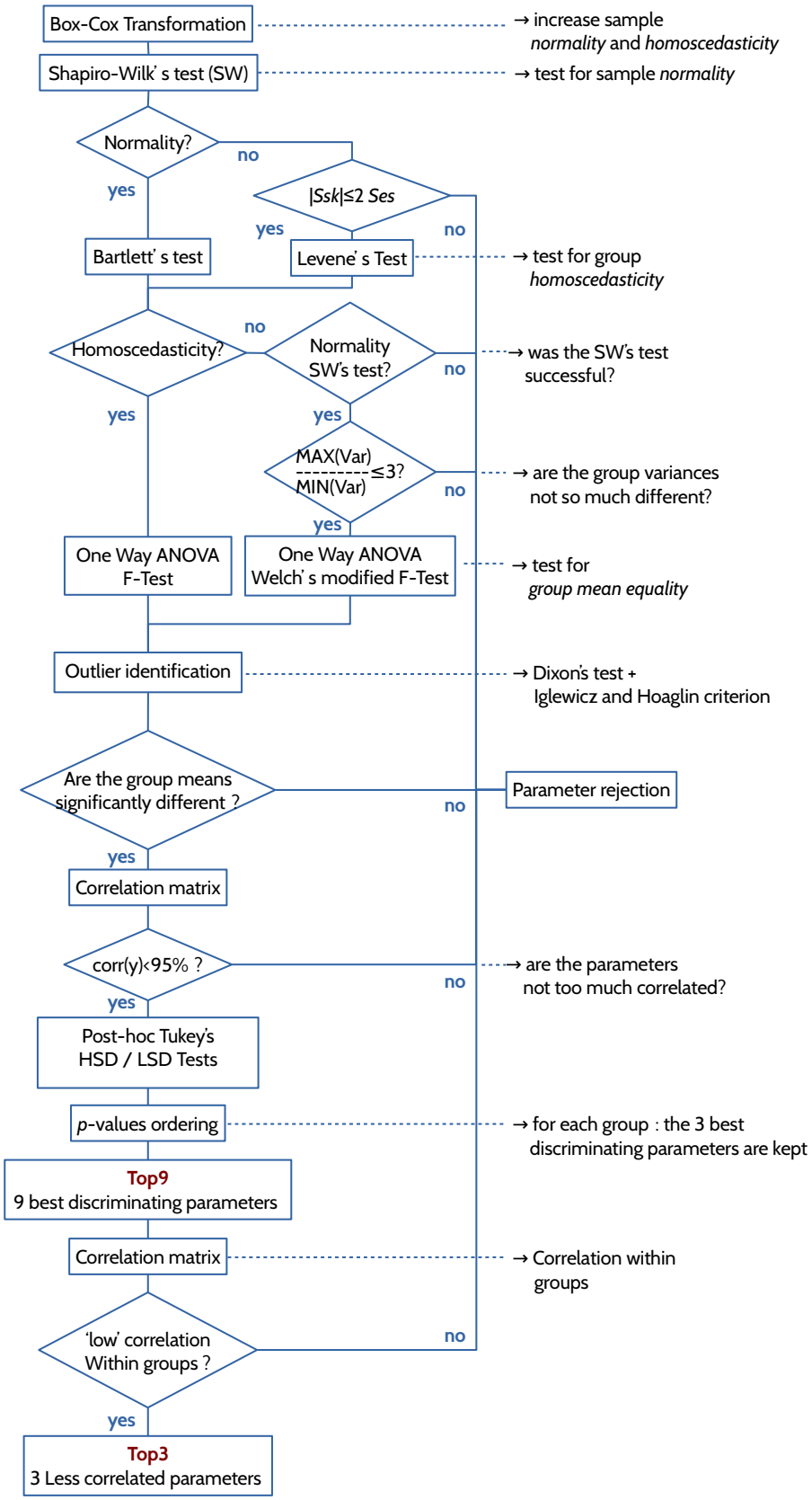

Figure 5: Global ANOVA approach. The parameters are systematically Box-Cox transformed. Then the classical assumptions, related to ANOVA, are verified (normality and homoscedasticity). After the ANOVA *F*-test, the outliers are marked. The most correlated parameters are rejected (cutoff=95%) before performing *post hoc* tests. For each group, the three parameters with the lowest *p*-values are put together in a Top9 set. Finally the three parameters that are the less correlated inside the groups compose the final Top3 set.

The ANOVA procedure results validate the following points, regarding the selected parameters:

1. the discriminative set should include both height and spatial parameters,
2. sampling the surfaces brings separability power to the parameters,
3. removing high degree polynomials from the primary surface also increases separability.

The major drawback of the ANOVA approach lies in its highly probabilistic nature. The assumptions inherent to the model have to be carefully verified. As a result, the statistic tests accumulate as well as the subsequent risks, so that interesting parameters may be erroneously disqualified. Moreover, if the overall complexity and the open question of the *p*-value ordering are considered, the ANOVA has to be performed carefully.

3.2. Results of the conservative ANOVA procedure

The three species of ruminants can be represented on biplots, each biplot involving two parameters. The Top3 set (fst_Rmax_{SC}, min_Sp_{SC}, mea_Rmax_{SC}) needs three different views, Fig.6.a-c) The biplot b) (mea_Rmax_{SC}, fst_Rmax_{SC}) exhibits too much overlapping to be useful. Both parameters are built on the physical quantity $Rmax_{SC}$ and that is why a clear correlation appears. Both have difficulties in separating the groups AA and CS, with *p*-values 0.0105 and 0.0544. The biplot c) (mea_Rmax_{SC}, min_Sp_{SC}) is less overlapped but it is still difficult to distinguish the different categories. The one that best scatters the groups is a), the (min_Sp_{SC}, fst_Rmax_{SC}) biplot.

The separation is not perfect, but it can be said that:

- AB group has the highest anisotropy, which is a long known result. Within the 1024 subsurfaces of an AB surface, the smallest *Rmax* values are higher than those of AA and CS groups. It simply means that anisotropy is encountered everywhere on an AB surface.
- Conversely, AA surfaces are somewhat isotropic. AA is also characterized by low peak height, which is not surprising because soft diets tend to make the surfaces smoother.
- As expected, the CS surfaces are characterized by gouges and scratches, which lead to an important relief. In addition, if the surface features are not scattered uniformly, there may be a non-negligible variability during the surface acquisition. This explains why there are high peaks and why CS surfaces span over the *Rmax* axis.

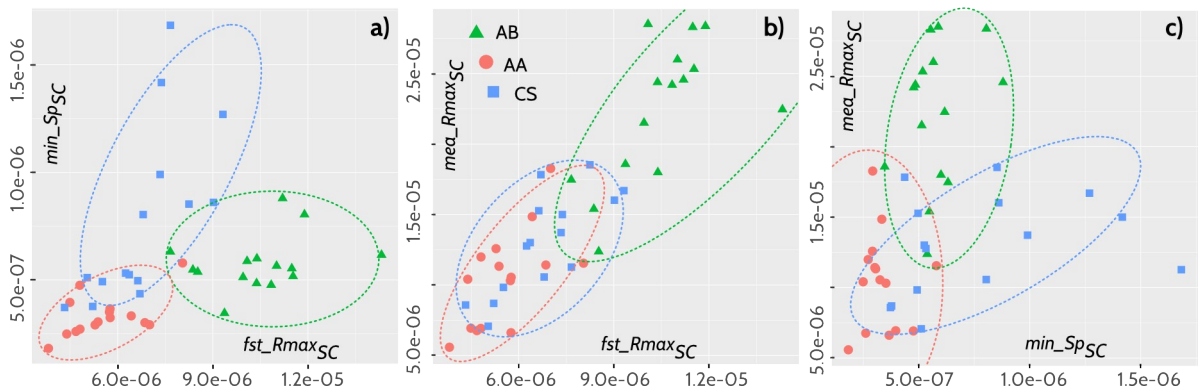

Figure 6: a)-b)-c) Biplots with the conservative ANOVA Top3 selection. As expected the CS group meets some AB and AA characteristics, which explains the important ellipsis overlapping. The ellipses are qualitative representations of the groups.

3.3. The simplified procedure

As previously defined, the ANOVA procedure has been built according to some mathematical rules, dealing with the ANOVA assumptions. It uses a commonsense tool for dimensionality reduction, *i.e.* the removal of highly correlated parameters. The benefits of such precautions have been investigated (see ESM §3). Firstly, the correlation steps were omitted and the results compared, then the relevance of the ANOVA itself was assessed.

The results (ESM §3) show that the high correlation removal is not relevant because removing highly correlated parameters has turned into dropping relevant parameters. Hundreds of parameters can be processed in a few seconds to determine their *p*-values. The correlation steps have therefore been dropped.

The parameter Box-Cox transformation has also been suppressed to verify its usefulness. The resulting group separation is not quite as pronounced because the parameters that were chosen through the procedure are slightly less discriminative. Finally, removing the ANOVA *F*-test has no effect on classification efficiency. The ANOVA itself has been shown to be redundant with *post hoc* tests like HSD and, to ease discriminative parameter identification, the parameters should be systematically Box-Cox transformed. The discriminating procedure reduces to Fig.7.

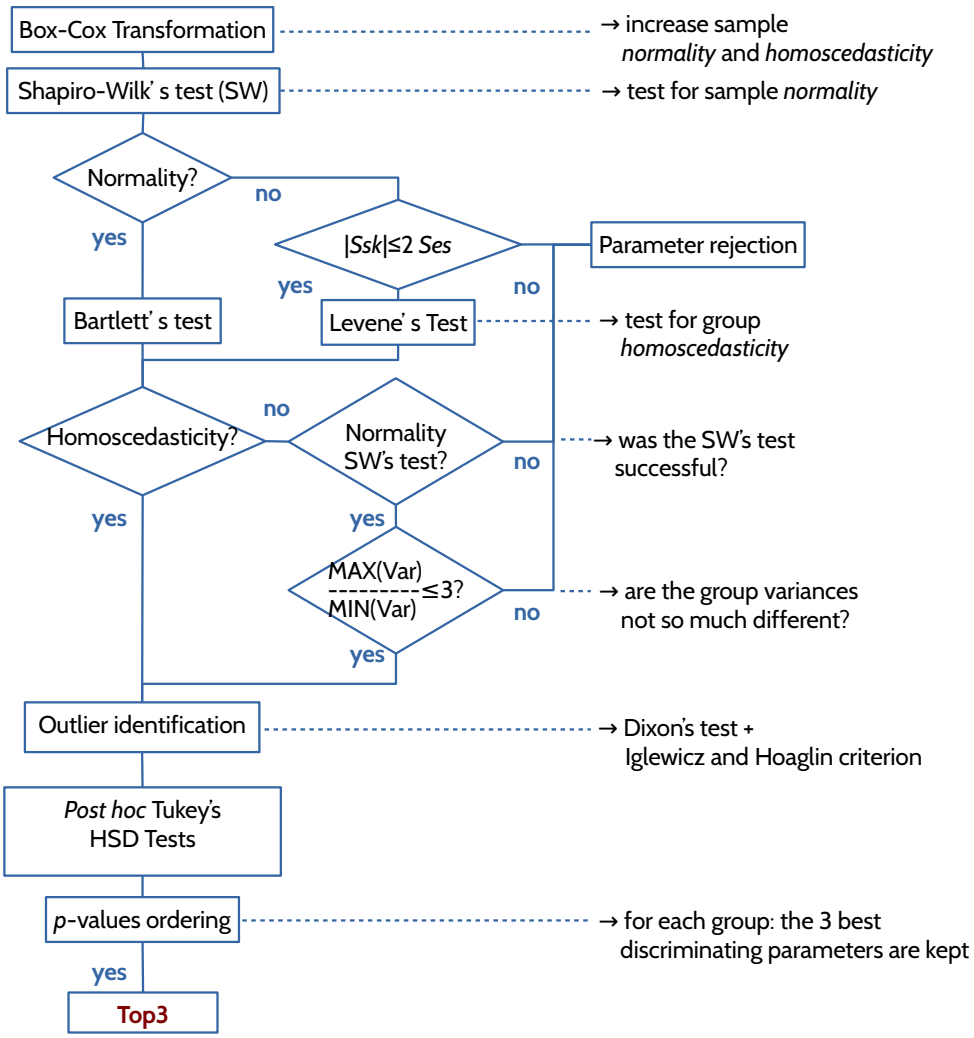

Figure 7: Simplified discriminating procedure. The correlation steps have been removed, as well as the ANOVA *F*-test.

3.4. The use of dimensionless surfaces

In numerous scientific areas, researchers choose to work on dimensionless parameters but the reasons invoked, such as common scales, parameter grouping, numerical errors and more, will not be discussed here. What can motivate this choice here is the opportunity to define "standard" thresholds for the parameters: if two different AA individuals are to be compared, for instance, the degree of global wear can be very different. The hope is that the features (pits, grooves, etc.) are the same, thus scaling heights with the whole surface height standard deviation may lead to more widely applicable thresholds, this being regardless of the degree of wear. The simplest way to make parameters dimensionless is to work with height-dimensionless surfaces *SB* and *SC*, with respect to their height standard deviation (*SA*: outlier-cleaned surface, *SB*: *SA* minus its 2nd order polynomial, *SC*: *SA* minus its 8th order polynomial). As concerns the spatial parameters, whether a characteristic length is defined for non-dimensioning the parameters *Rmax* and *Sal*, whether only the dimensionless parameter *Str* is used. The easiest but efficient solution is the second option.

Using the reduced procedure previously detailed, Table 3 and Fig.8.b) show that the same efficiency is reached as for non-normalized surfaces. For a "more" dimensionless set of parameters, ent_Rmax_{SC} is replaced by fst_Str_{SC} (the lowest *p*-valued *Str*), for which *p*-values are higher (1.41×10^{-08}, 4.71×10^{-08}, 7.16×10^{-01}), however the group separation remains effective, Fig.8.c) in comparison with Fig.8.b)

parameter	physical meaning	type	AB-AA	AB-CS	AA-CS
fst_Ssk_{SC}	skewness 5-percentile value of the sampled *SC* surface	height	2.39E-12	~~6.16E-02~~	9.74E-10
ent_Rmax_{SC}	*acf* ellipsis major axis value calculated on the entire *SC* surface	spatial	3.77E-11	3.81E-08	3.80E-02
fst_Rmax_{SC}	*acf* ellipsis major axis 5-percentile value of the sampled *SC* surface	spatial	8.28E-12	4.41E-08	1.06E-02

Table 3 – Top3 parameter set for normalized surfaces, with the simplified procedure. A crossed *p*-value (above 5%) means non-significant differences

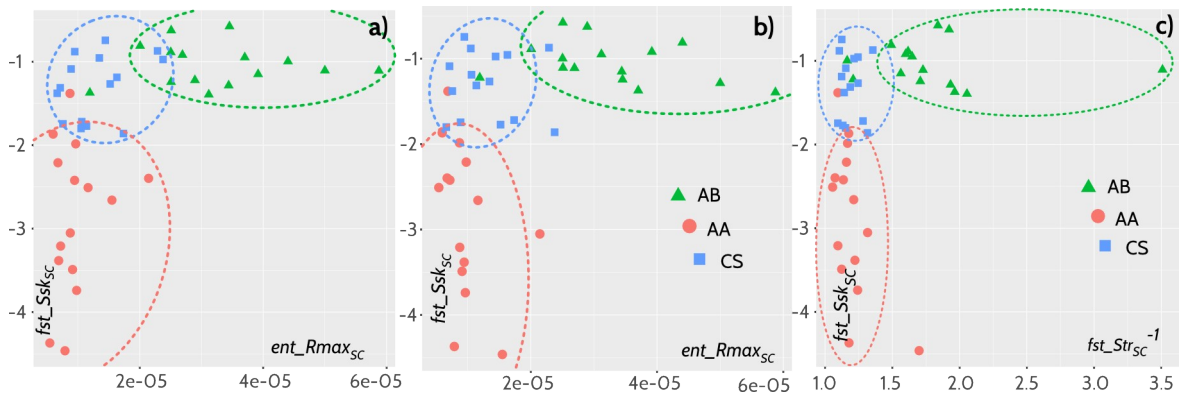

Figure 8: Biplot comparison. a) Biplot of the best pair of parameters obtained without the last correlation step. Inside each group, the best *p*-valued parameter is kept. The figure exhibits less scattered data, compared to Fig.6.a) Preferring the parameters chosen on *p*-values rather than correlation improves the discrimination. b) Biplot with the two best parameters on dimensionless surfaces. As ent_Rmax_{SC} and fst_Ssk_{SC} are not impacted by surface normalization, the quality of the discrimination remains qualitatively the same. c) Choosing the dimensionless parameter fst_Str_{SC} instead of ent_Rmax_{SC} ensures the discrimination will be as accurate as b). Moreover, it gives the chance to compare the present results with others, for which the wear rates could be different. The ellipses are qualitative representations of the groups.

4. Concluding remarks

Together with a more conservative identification of the very dental facets on which dental microwear analysis should be run [33], the present study has made important improvements to enhance dietary group discrimination. These concern the use of extended parameters thanks to two derived surfaces (*SB* and *SC*) coupled with surface sampling. A reflexion about the reproducibility of the method and the results should however take place. Indeed some aspects of the present work involve rather long coding developments and, the scope of the conclusions is worth being discussed.

4.1. About the parameters

The autocorrelation function is an efficient means for detecting anisotropy but it is very sensitive to long wavelengths. The higher order statistics (*Ssk, Sku*) are efficient in detecting height asymmetry and flatness but they are also sensitive to long wavelengths. Therefore, subtracting a high order polynomial from the *SA* surfaces turns these tools into better indicators of typical diets. As concerns the degree of the polynomial, several tries have been made on the surfaces and the best degree was found to be eight. However, no structured parametric study has been yet carried out to validate this conclusion.

It has clearly appeared that the best combination of discriminative parameters includes both a spatial and a height parameter. In future work, a refined version of *Asfc* should be defined to discriminate the CS group, as the fractal nature of this parameter makes it independent from other height and spatial parameters and the literature has shown its ability to isolate typical CS signals.

4.2. About the surface sampling

It was expected that surface sampling should be a key feature in the parameter selection procedure because, intuitively, a single surface scratch cannot be proof of global anisotropy. In the same manner, a global skewed surface is not proof of a soft diet but rather of an "everywhere skewed" surface. Even if the initial assumptions should be somewhat tempered because ent_Rmax_{SC} has proved to be a fine anisotropy indicator calculated on the whole surface, most of the discriminative parameters (low *p*-values) come from surface sampling. Future work should investigate the best way to partition the surface: number of samples, randomly or not, with or without overlapping, etc.

4.3. About the correlation steps and the ANOVA

The conservative ANOVA procedure has been built to be in compliance with the ANOVA mathematical prerequisites and the general idea that the Top3 parameter set should be low correlated. The last correlation step is removing the more within-group correlated parameters. This suppresses high discriminative power parameters and therefore must not be used.

The first correlation step, removing >95% of correlated parameters, substantially reduces the parameter set. Although it is more readily encountered with larger parameter sets and that it is rather introduced prior a neural network use, it can prevent the Top9 from containing the same kind of parameters. In the present study, this step was revealed to be useless because the best parameter couple was the same as before.

The *post hoc* test yields the same results with or without the ANOVA, which is in accordance with some authors in

literature. In the present case the ANOVA *F*-test is therefore useless. However, the tests that check the required assumptions should be kept, *i.e.* normality, variance homogeneity and outlier presence. So, the parameter transformations to increase the aforementioned required properties are useful. They increase the subsequent test power in detecting group mean differences. The widely used and systematic log-transformation can easily be replaced by more efficient transformations, such as Box-Cox.

4.4. About surface normalization

Working with dimensionless parameters offers the opportunity to compare very different individuals from the perspective of a surface dimension, species, rate of wear, etc. The task is easily ensured by normalizing the surfaces, which makes the height parameters dimensionless and by choosing *Str* for anisotropy discrimination. As a result, quality remains the same, the surface normalization should then be systematically tried for discrimination and classification purposes.

4.5. About reproducibility

Behind "reproducibility" the researchers usually expect that:

1. A published work, presenting a method that yields better results than previous ones, could be "reproduced".

   In the present case, unless all sources are made freely available (fortran, python and R) and documented, it represents a very long task. So, the authors have planned in a medium/long term to release the codes (after the necessary cleaning and documentation) as well as "reference" surfaces. However, subtracting a 2- or 8-degree polynomial surface and sampling the resulting surface remain affordable tasks to be performed along with the usual parameters.

2. A published method could eventually be extended to different materials.

   Attempts have been made to discriminate non-herbivorous species with the Top3 selected parameters, but it lead to rather bad results, and other parameters revealed to be more relevant. However it should not mask the main purpose here, which is sharing with the dental microwear analysis researchers the benefits of the parameter set extension associated to a thorough, but not overly strict, procedure. If the materials are related to herbivores – for which grazer/browser signals are dominant – we make the assumption that very few kinds of parameter are needed among the most common height and spatial ones, in conjunction with surface sampling and long/middle wavelength removal. This topic will be addressed in a future work.

Acknowledgments

The Project TRIDENT was supported by the French National Research Agency, ANR (ANR-13-JSV7-0008-01; PI: G.M.).

The authors wish to express their gratefulness to the referees for having reviewed the present work thoroughly, despite its complexity and length, making the whole paper clearer and more readable.

References

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

2. Walker A, Hoeck HN, Perez L. 1978 Microwear of mammalian teeth as an indicator of diet. *Science* **201**, 908–910. (doi:10.1126/science.684415)

3. Lucas PW *et al.* 2013 Mechanisms and causes of wear in tooth enamel: implications for hominin diets. *J. R. Soc. Interface* **10**, 20120923. (doi:10.1098/rsif.2012.0923)

4. Merceron G *et al.* 2016 Untangling the environmental from the dietary: dust does not matter. *Proc. Biol. Sci.* **283**. (doi:10.1098/rspb.2016.1032)

5. Teaford MF, Walker A. 1984 Quantitative differences in dental microwear between primate species with different diets and a comment on the presumed diet of Sivapithecus. *Am. J. Phys. Anthropol.* **64**, 191–200. (doi:10.1002/ajpa.1330640213)

6. Solonias N, Teaford M, Walker A. 1988 Interpreting the Diet of Extinct Ruminants: The Case of a Non-Browsing Giraffid. *Paleobiology* **14**, 287–300.

7. Hofmann RR. In press. Evolutionary steps of ecophysiological adaptation and diversification of ruminants: a comparative view of their digestive system. *Oecologia* **78**, 443–457. (doi:10.1007/BF00378733)

8. Boyde A, Fortelius M. 1991 New confocal LM method for studying local relative microrelief with special reference to wear studies. *Scanning* **13**, 429–430. (doi:10.1002/sca.4950130608)

9. Brown CA, Savary G. 1991 Describing ground surface texture using contact profilometry and fractal analysis. *Wear* **141**, 211–226. (doi:10.1016/0043-1648(91)90269-Z)

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

11. Solounias N, Semprebon G. 2002 Advances in the Reconstruction of Ungulate Ecomorphology with Application to Early Fossil Equids. *Am. Mus. Novit.* , 1–49. (doi:10.1206/0003-0082(2002)366<0001:AITROU>2.0.CO;2)

12. Merceron G, Viriot L, Blondel C. 2004 Tooth microwear pattern in roe deer (Capreolus capreolus L.) from Chizé (Western France) and relation to food composition. *Small Rumin. Res.* **53**, 125–132. (doi:10.1016/j.smallrumres.2003.10.002)

13. Ungar PS, Brown CA, Bergstrom TS, Walkers A. 2003 Quantification of dental microwear by tandem scanning confocal microscopy and scale-sensitive fractal analyses. *Scanning* **25**, 185–193.

14. 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. (doi:10.1038/nature03822)

15. Grine FE, Ungar PS, Teaford MF. 2002 Error rates in dental microwear quantification using scanning electron microscopy. *Scanning* **24**, 144–153. (doi:10.1002/sca.4950240307)

16. Mihlbachler MC, Beatty BL, Caldera-Siu A, Chan D, Lee R. 2012 Error rates and observer bias in dental microwear analysis using light microscopy. *Palaeontol. Electron.* **15**, 12A.

17. Scott RS, Ungar PS, Bergstrom TS, Brown CA, Childs BE, Teaford MF, Walker A. 2006 Dental microwear texture analysis: technical considerations. *J. Hum. Evol.* **51**, 339–349. (doi:10.1016/j.jhevol.2006.04.006)

18. Kaiser TM, Brinkmann G. 2006 Measuring dental wear equilibriums—the use of industrial surface texture parameters to infer the diets of fossil mammals. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **239**, 221–240.

(doi:10.1016/j.palaeo.2006.01.013)

19. In press. ISO 4287:1997 - Geometrical Product Specifications (GPS) -- Surface texture: Profile method -- Terms, definitions and surface texture parameters. *ISO*. See http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=10132 (accessed on 8 December 2016).

20. In press. ISO 25178-2:2012 - Geometrical product specifications (GPS) -- Surface texture: Areal -- Part 2: Terms, definitions and surface texture parameters. *ISO*. See http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=42785 (accessed on 8 December 2016).

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

22. Purnell M, Seehausen O, Galis F. 2012 Quantitative three-dimensional microtextural analyses of tooth wear as a tool for dietary discrimination in fishes. *J. R. Soc. Interface* **9**, 2225–2233. (doi:10.1098/rsif.2012.0140)

23. 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. *J. Zool.* **291**, 249–257. (doi:10.1111/jzo.12068)

24. 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. (doi:10.1038/nature13622)

25. Purnell MA, Darras LPG. 2016 3D tooth microwear texture analysis in fishes as a test of dietary hypotheses of durophagy. *Surf. Topogr. Metrol. Prop.* **4**, 014006. (doi:10.1088/2051-672X/4/1/014006)

26. 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. *Sci. Rep.* **6**, srep34037. (doi:10.1038/srep34037)

27. Ungar PS, Merceron G, Scott RS. 2007 Dental Microwear Texture Analysis of Varswater Bovids and Early Pliocene Paleoenvironments of Langebaanweg, Western Cape Province, South Africa. *J. Mamm. Evol.* **14**, 163–181. (doi:10.1007/s10914-007-9050-x)

28. Scott JR. 2012 Dental microwear texture analysis of extant African Bovidae. *mammalia* **76**, 157–174. (doi:10.1515/mammalia-2011-0083)

29. Gagnon M, Chew AE. 2000 Dietary preferences in extant African Bovidae. *J. Mammal.* **81**, 490–511.

30. Franzmann AW. 1981 *Alces alces*. *Mamm. Species* , 1–7.

31. Galbany J, Estebaranz F, Martínez LM, Romero A, De Juan J, Turbón D, Pérez-Pérez A. 2006 Comparative analysis of dental enamel polyvinylsiloxane impression and polyurethane casting methods for SEM research. *Microsc. Res. Tech.* **69**, 246–252. (doi:10.1002/jemt.20296)

32. Goodall RH, Darras LP, Purnell MA. 2015 Accuracy and Precision of Silicon Based Impression Media for Quantitative Areal Texture Analysis. *Sci. Rep.* **5**, srep10800. (doi:10.1038/srep10800)

33. Ramdarshan A, Blondel C, Gautier D, Surault J, Merceron G. 2017 Overcoming sampling issues in dental tribology: insights from an experimentation on sheep. *Paleontol. Electron.*

34. Francisco A, Brunetière N. 2016 A hybrid method for fast and efficient rough surface generation. *Proc. Inst. Mech. Eng. Part J J. Eng. Tribol.* **230**, 747–768. (doi:10.1177/1350650115612116)

35. Osborne JW. 2007 *Best Practices in Quantitative Methods*. Thousand Oaks, Calif: SAGE Publications, Inc.

36. Bulmer MG. 1979 *Principles of Statistics*. 0002–Revised edition edn. New York: Dover Publications.

37. Adams KA, Lawrence EK. 2014 *Research Methods, Statistics, and Applications*. SAGE Publications.

38. Cramer D. 1998 *Fundamental Statistics for Social Research: Step-by-step Calculations and Computer Techniques Using SPSS for Windows*. Psychology Press.

39. Ramachandran KM, Tsokos CP. 2009 *Mathematical Statistics with Applications*. Academic Press.

40. Ramsey FL, Schafer DW. 2002 *The Statistical Sleuth: A Course in Methods of Data Analysis*. Duxbury/Thomson Learning.

41. Cohen BH. 2014 *Explaining Psychological Statistics*. 4th Edition. Hoboken, New Jersey: John Wiley & Sons.