Supplementary material

Gathering and Analyzing Surface Parameters for Diet Identification Purposes.

Arthur Francisco [1,*], Noël Brunetière [1] and Gildas Merceron [2]

[1] Institut Pprime, CNRS, Université de Poitiers, ISAE-ENSMA, F-86962 Futuroscope Chasseneuil, France; arthur.francisco@univ-poitiers.fr

[2] PALEVOPRIM UMR 7262, CNRS and Université de Poitiers, 86073 Poitiers Cedex 9, France; gildas.merceron@univ-poitiers.fr

[*] Correspondence: arthur.francisco@univ-poitiers.fr; Tel.: +33-5-45-25-19-79

Outline

1. Preparation

1.1. The Autocorrelation Function $f_{ACF}(tx,ty)$ Analysis

The autocorrelation function f_{ACF} searches for correspondences of a surface with itself. If the surface heights are white noise, the only shift (tx, ty) that matches the original surface is $(0,0)$: $f_{ACF}(0,0) = 1$ and 0 elsewhere. Conversely if the surface exhibits "macro" features, like peaks, valleys, scratches, etc. then f_{ACF} decreases more slowly from its maximum value 1 (reached in tx=0, ty=0). If the surface is isotropic (no preferred direction) f_{ACF} is axisymmetric, otherwise f_{ACF} has a higher decreasing rate across the pattern direction. Therefore f_{ACF} decreasing behavior is a means to catch the direction of anisotropy when it occurs, and especially it quantifies the amount of anisotropy, as explained on figure S1.

The aforementioned method needs a height z for the cutting plane. The ISO 25178 norm suggests z=0.2, however it is not suitable (it is too low) for numerous anisotropic surfaces. We propose to average the values of *Sal* and *Rmax* for z=0.3, 0.4 and 0.5. A complementary way of catching the decreasing behavior of f_{ACF} is to directly study its slope around $(0,0)$, figure S2.
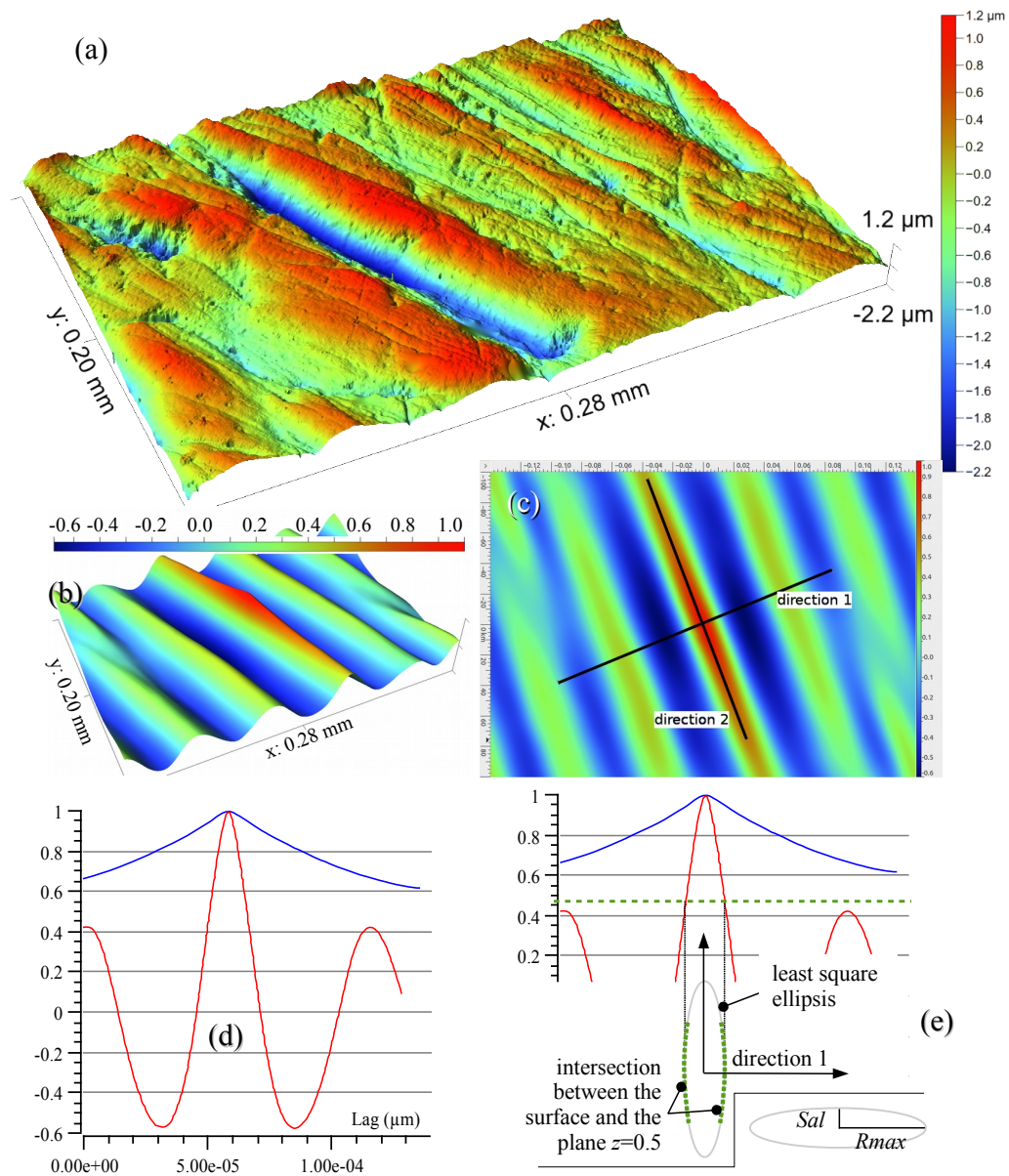
Figure S1. (a) A typical dental surface to be analyzed.
(b) The normalized 2D autocorrelation function.
(c) Two profiles are extracted along directions #1 and #2; #1 across the scratches and #2 along the scratches.
**(d)** The profile #1 (red curve) is more self repeating than the profile #2 because of shorter wavelengths.
(e) A plane that cuts the 2D f_{acf} surface at height z, defines an ellipsis—or just a part of it—with the small axis in direction #1 and the big axis in direction #2. The anisotropy can be quantified by *Rmax/Sal*, and the groove length by *Rmax*—*Rmax*: semi-major axis, *Sal*: semi-minor axis.
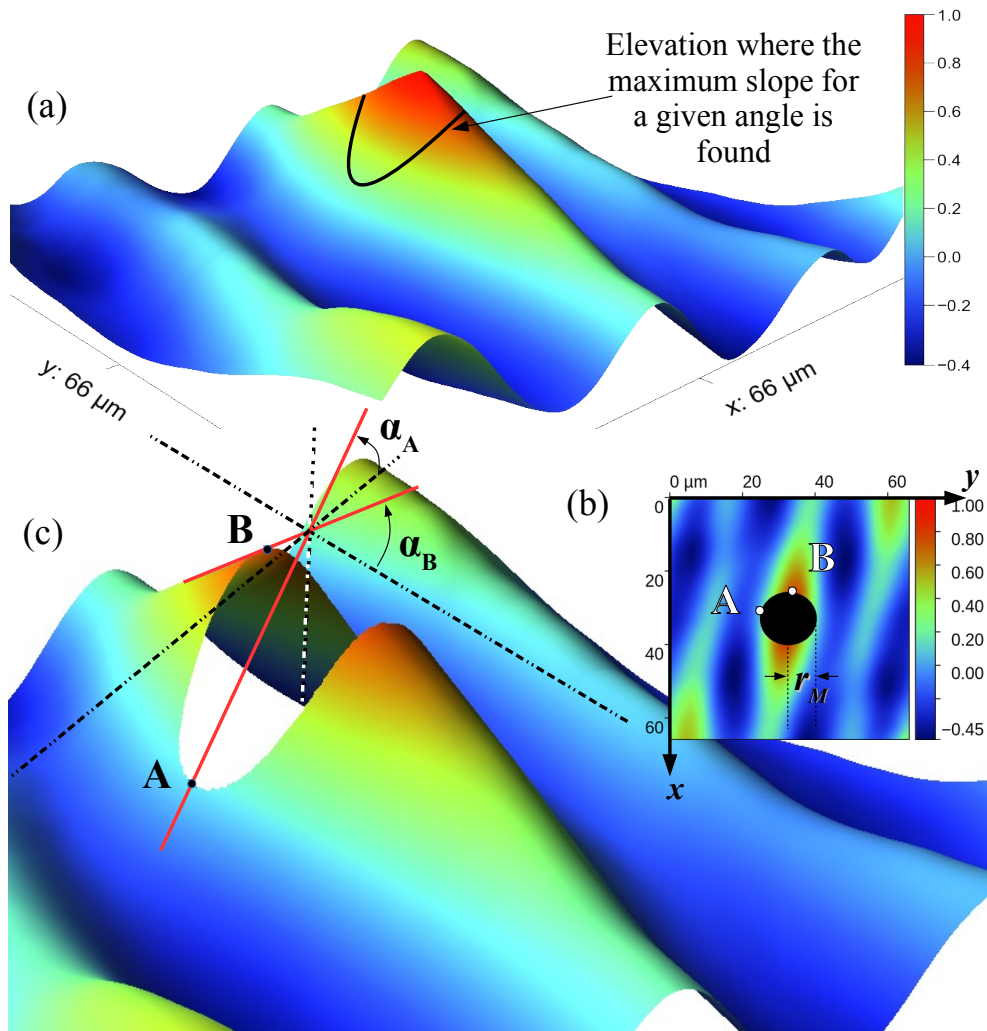
Figure S2. (a) In each direction, the point of maximum slope is recorded.
(b) At the minimum radius of the curve, the slopes are determined. The highest is located in A and the lowest in point B.
(c) Three parameters are built: $b.sl=\alpha_A$, $s.sl=\alpha_B$ and $r.sl = b.sl/s.sl$

1.2. Introduction of a Topological Parameter

A mixed parameter, dealing both with surface heights and height spatial repartition, is introduced. Discarding the 15% lowest heights, the points above x% of the height amplitude are masked. Then morphological operations are carried out on the mask in order to discard small cells. The remaining cells are therefore surface contiguous areas above a given threshold. Three parameters are built: the number of cells Snb, the median cell relative area Smc and the masked relative area Sk, figure S3.
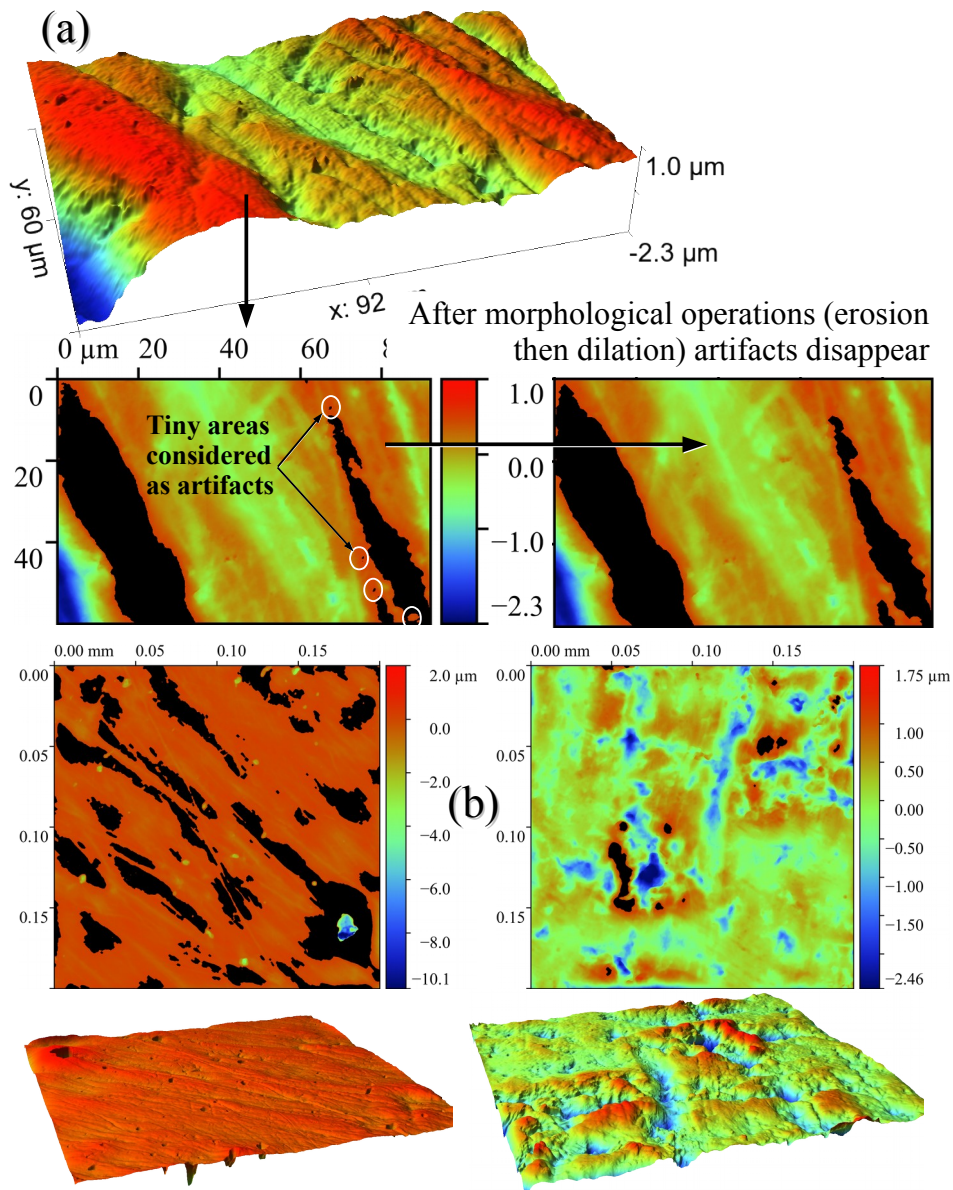
Figure S3. (a) The lowest 15% of the heights are ignored and the highest *x*% of the remaining are masked (left). After basic treatments, tiny masked cells are removed (right).
(b) Comparison between a grazer (left) and a fruit/seed feeder (right)

1.3. A Fast and Accurate Way of Determining Asfc

The ISO 25178-2:2012(E) states that "*The observed area is calculated as a function of scale by a series of virtual tiling exercises covering the measured surface in a patchwork fashion. The areas of the tiles, or patches, represent the areal scales of observation. The tiling exercises are repeated with tiles of progressively smaller areas to determine the observed areas to determine the observed areas as a function of the areal scales of observation.*" Then, the function log(*relative area*)=f(log(*element area*)) can be determined. The area-scale fractal analysis complexity parameter *Asfc* is defined as a thousand times minus the line slope, figure S4.
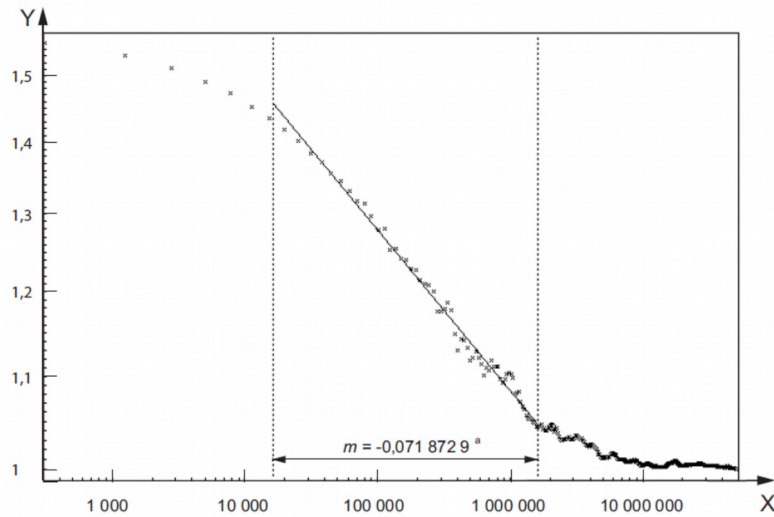
Figure S4. Typical relative area plot—ISO 25178-2:2012(E)

1.3.a. Determining the steepest part of the curve in a robust way

It can be reasonably stated that most of relative area plots are S-shaped, even if the trail can be more linear than curved. Instead of arbitrarily defining a "central" region where to calculate the maximum of the slope, the whole curve is fitted with a family Tn, of monotonic functions.

$$Tn(x) = y_0 + a . \tanh\left(\frac{x - x_0}{b}\right)^n \qquad (1)$$

In most cases $n=1$ leads to an accurate fit, especially in the linear part of the curve, where the studied surface exhibits self-similarity properties. The proposed model is parsimonious—only four parameters have to be determined, with a least square procedure for instance—and then doesn't suffer from overfitting. Some surfaces have better fit results with $n=2$ but it doesn't change much the *Asfc* value. The location x_s of the steepest part is analytically determined by canceling the second derivative of Tn. A simple calculus yields:

$$x_s = x_0 + b . atanh\left(\sqrt{\frac{n-1}{n+1}}\right) \qquad (2)$$

when $n=1$, it reduces to $x_s=x_0$.

To address the computation efficiency question—induced by the series of tiling exercises—the surface itself isn't tiled as explained in ISO 25178. The different scales are those of the grids that are used to discretize the surface. The grids are chosen regular with lateral steps from (hx, hy) for the finest grid to (Hx, Hy) for the largest grid. The finest grid is the one of the original surface, and the coarsest grid is an 8 × 8 grid. The original grid is the only one for which the surface points match the grid points, for the other grids the surface heights are obtained by interpolation. The intermediate grids are defined so that the element size has a geometric progression, hence its location is evenly spaced on a logarithm axis.

Three surface interpolation methods are tested: linear (linear Finite Elements), interpolant cubic splines, and Hermite (Hermite Finite Elements), summarized in figure S5.
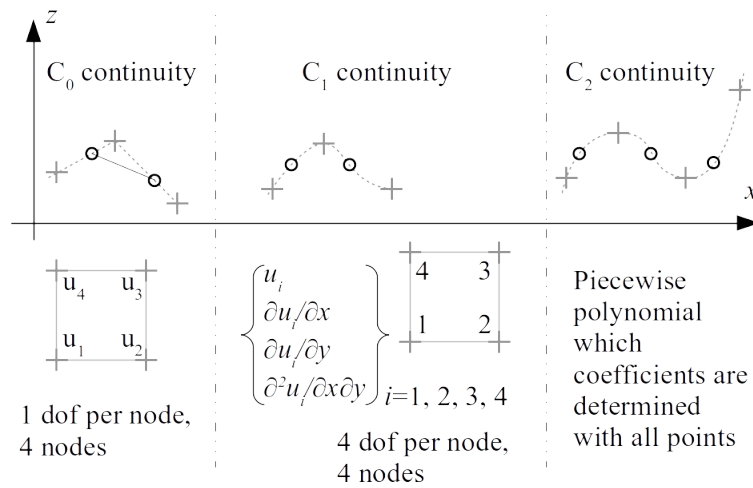
Figure S5. The chosen interpolation methods (left: linear FE, middle: Hermite FE, right: cubic spline). "dof" stands for "degree of freedom". In the upper part, one can see that from left to right, the smoothness increases. Below, calling *u* the unknown, it can be seen that increasing smoothness requires to take into account more derivatives. As the spline method guarantees curvature continuity, the unknown and its derivatives must be computed globally.

The surface heights are computed on the different grids with one of the three interpolation procedures that are proposed – linear (finite element style), Hermite (finite element style) and cubic spline. 128 grids are used to draw the relative area curve.

Provided the interpolating function *f(x,y)*, the area of a surface element defined on $[x_i, x_i+hx] \times [y_j, y_j+hy]$ is obtained by integration of $\sqrt{1+\left(\dfrac{\partial f}{\partial x}\right)^2+\left(\dfrac{\partial f}{\partial y}\right)^2}$. Such an expression being barely easy to analytically integrate, the Gauss integrating method is used with two points in each direction, which ensures the exact integration of a third degree polynomial on both directions, figure S6.
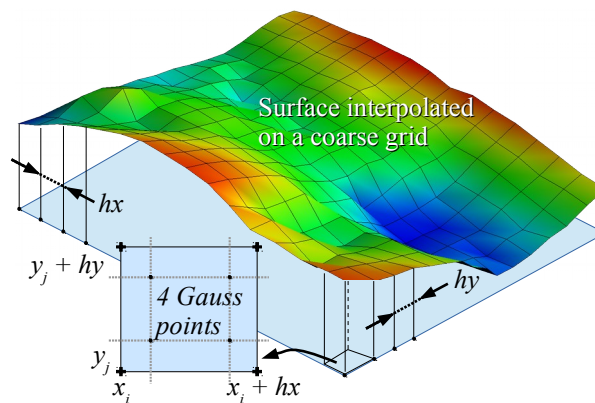
Figure S6. Gauss integrating points

Validation

The set of surfaces that we previously studied [1] is used to compare Toothfrax© results with the three methods proposed here.
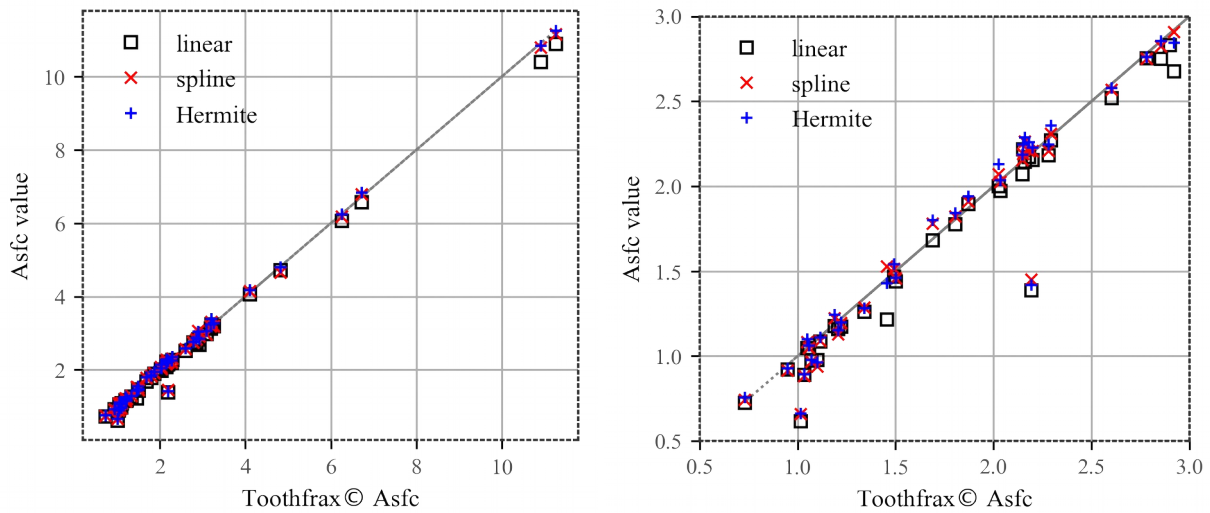
Figure S7. Global comparison of the presented methods with Toothfrax© results on the 45 SA surfaces **(left)**. Detailed view on small values pointing out some disparities **(right)**

Three observations are to be made on the figure S7. First, there is a satisfying concordance between the new values and Toothfrax© values, considered here as the reference. Second, even if the three methods (linear, spline and Hermite) yield close results, the smoothest methods (spline and Hermite) are slightly better. The spline and Hermite methods are very close, and because the fastest of the two is the Hermite one, it will be the preferred approach. Third, two points deserve further investigation because of the gap regarding the reference.

To understand why the values are so different, a detailed analysis of the *Asfc* is carried out.

It clearly appears on figure S8(left) that Toothfrax© locates the steepest part of the curve near the trail, where it is somewhat "chaotic". This zone is related to the biggest tiling elements for which it's hard to speak of self-similarity.
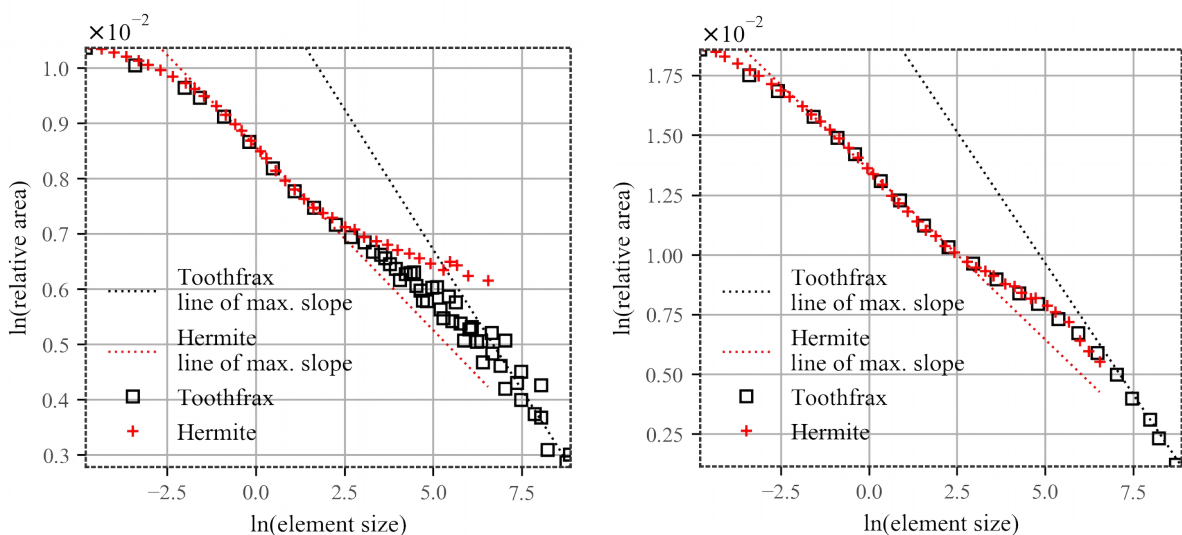
Figure S8. Relative areas for the surface #1 (left) and for surface #2 (right)

Hermite interpolation exhibits a much more S-shaped curve where the maximum of the tangent slope locates towards smaller elements. The second surface case – figure S8(right) – is different: the curves are the same, but the steepest parts are located differently. Once more, Toothfrax© identifies the curve trail as the steepest part of the curve whereas the hyperbolic tangent approximation locates it closer to the middle, as it can be clearly seen.

It seems that the maximum of slope for the surface #2 may be greater, but it's also one of the benefits of the

approximation: it brings stability when the curve trail deviates from an "S" trail.

A way to make Toothfrax© and Hermite results coincide is to subtract a "macro-geometric" shape from the surface. The reason is quiet simple: increasing the surface global flatness makes the tiling mesh closer to its projection on the horizontal plane. Hence the surface patchwork used to assess the developed area is close to the points of the same scale grid used in our approach.

If the previous study is carried out on SC surfaces, the conclusions are therefore different: the figure S9 shows very small differences between Toothfrax© results and a Hermite-based *Asfc* calculus. Using Hermite interpolation, *SA* and *SC* surfaces lead to very close results, figure S10.
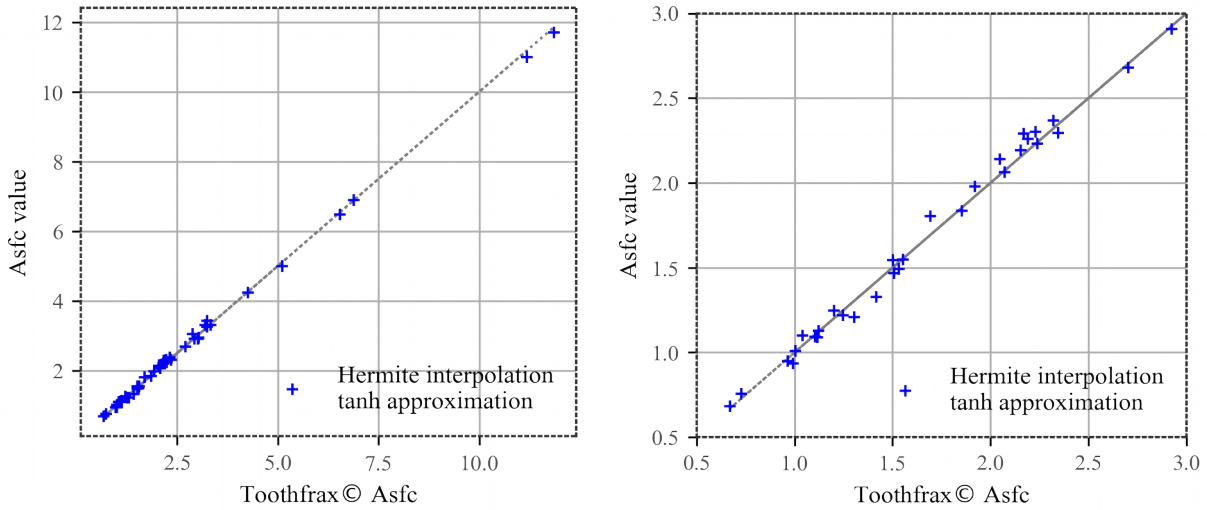
Figure S9. Global comparison of Hermite method with Toothfrax© on the 45 SC surfaces **(left)**. Detailed view on small values **(right)**
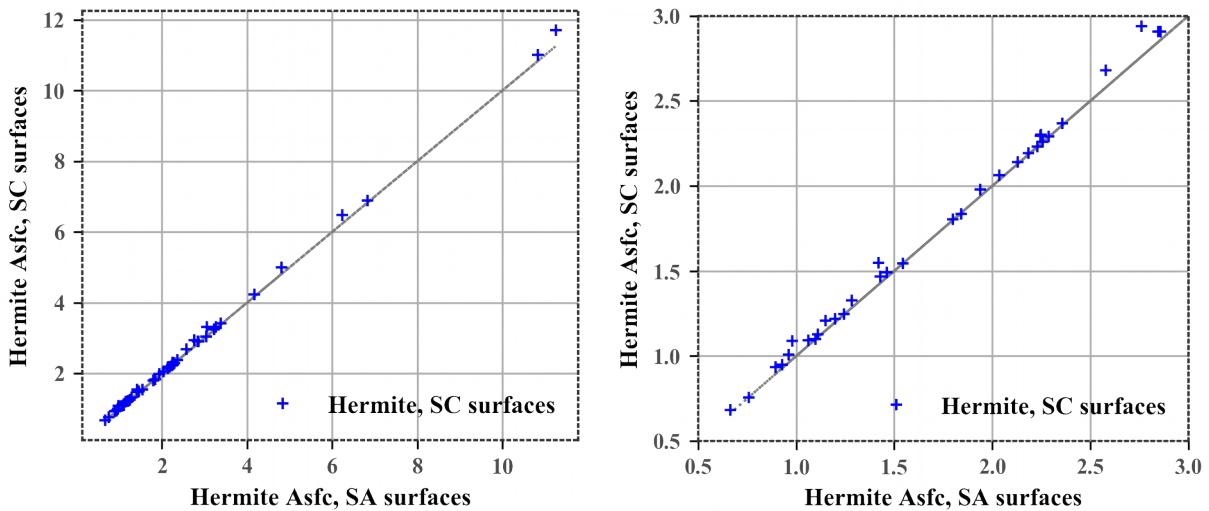
Figure S10. Global comparison of the Hermite method on the 45 *SA* and *SC* surfaces **(left)**. Detailed view on small values **(right)**

Table S1. Interpolation method comparison.

Interpolation method	Computational speed	Accuracy and smoothness	Implementation	Continuity	Interpolation scope
Linear FE	+++	+	+++	C_0	local
Cubic splines	+	+++	++	C_2	global
Hermite FE	++	+++	+	C_1	local

The pros and cons of the interpolation methods are summarized in table S1. As demonstrated above, determining *Asfc* using both Hermite FE interpolation and a hyperbolic tangent function approximation is an efficient and accurate method. It can be concluded that the *Asfc* parameter is worth being computed with the Hermite approach because of its accuracy and speed (less than 8 seconds on a single thread for a 2168×1555 px surface against several minutes for Toothfrax© with the recommended parameters) on *SC* surfaces.

1.3.b. Sample Size Influence on Asfc

Another question to address about the *Asfc* parameter is the right surface size. As for the other parameters—height parameters, spatial parameters, etc.—it's easier to decide. For none of them a multiscale analysis is needed, so providing a large enough surface sample ensures consistent values. In the preceding study [1] it has been stated that 512×512 px samples are large enough for the whole set of parameters, because it represents a 66×66 µm² area. Throughout the tests carried out by the authors on several surface sets, it appears that this size can be reduced to 256×256 px without consistency loss. However, as the *Asfc* parameter must apply to a sample that embeds enough scales, a closer look is required.

A *SC* surface, belonging to the African fruit-browsing duiker is cut according the scheme presented in figure S11. This surface is chosen because it exhibits local reliefs that ensure a globally high *Asfc* value, but what about the surface parts that are flatter? *Asfc*—as well as *Rmax* and *SKu* parameters, for comparison purposes—is calculated and normalized with its reference value, calculated on the sample 1037×743. The reference is chosen so that it seems to be the largest one that remains nearly isotropic. Larger surfaces include important reliefs that unavoidably change the *Asfc* value.
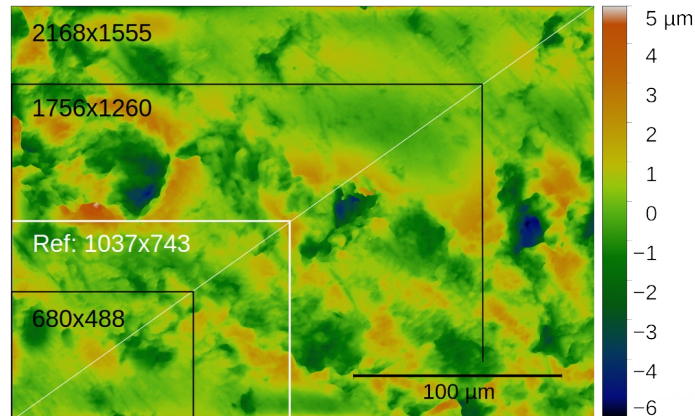
Figure S11. Different surface sizes are tested to assess *Asfc* robustness against the sample size. The white sample results are used to normalize the other sample results.
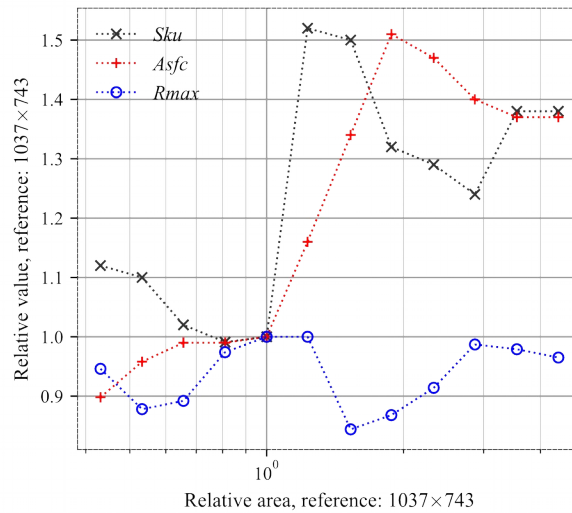
Figure S12. *Asfc*, *Sku* and *Rmax* as a function of the subsurface size.

The figure S12 shows that *Asfc* strongly depends on the subsurface chosen: when the reference subsurface is enlarged, a wide hole is included and, as expected, the fractal parameter increases. Then, larger subsurfaces dilute this inclusion and the parameter value decreases. Similar variations are observed for *Sku*, whereas *Rmax* is less affected.

Generally speaking, it's clear that choosing too small a sample—it's about definition, not its size—would give a rough *Asfc* relative area curve. Moreover, catching features like the hole above the reference surface needs at least 512×512 px. In the other hand, as previously seen, the curve shows that expanding the sample dilutes the hole effect. Different subsamplings are therefore tested: 512×512 px and 1024×1024 px subsurfaces along with 81, 144 and 256 samples.

We found that sampling a surface with 256 subsurfaces 256×256 evenly spaced, is a satisfactory trade-off between computing time and information quality regarding the height and spatial parameters. A small parametric study is carried out to show the influence of the sample size and the subsurface size, figure S13. Assuming that *Asfc* should be determined on subsurfaces of at least 512×512 px – to guarantee enough scales – the results suggest that for 81 and higher samples, the *Asfc* range of values is 11.7–22.9 As statistics are to be determined – in particular quantiles 5% and 95% – and for consistency purposes, the surfaces are finally sampled with 256 subsurfaces 512×512 for *Asfc*.
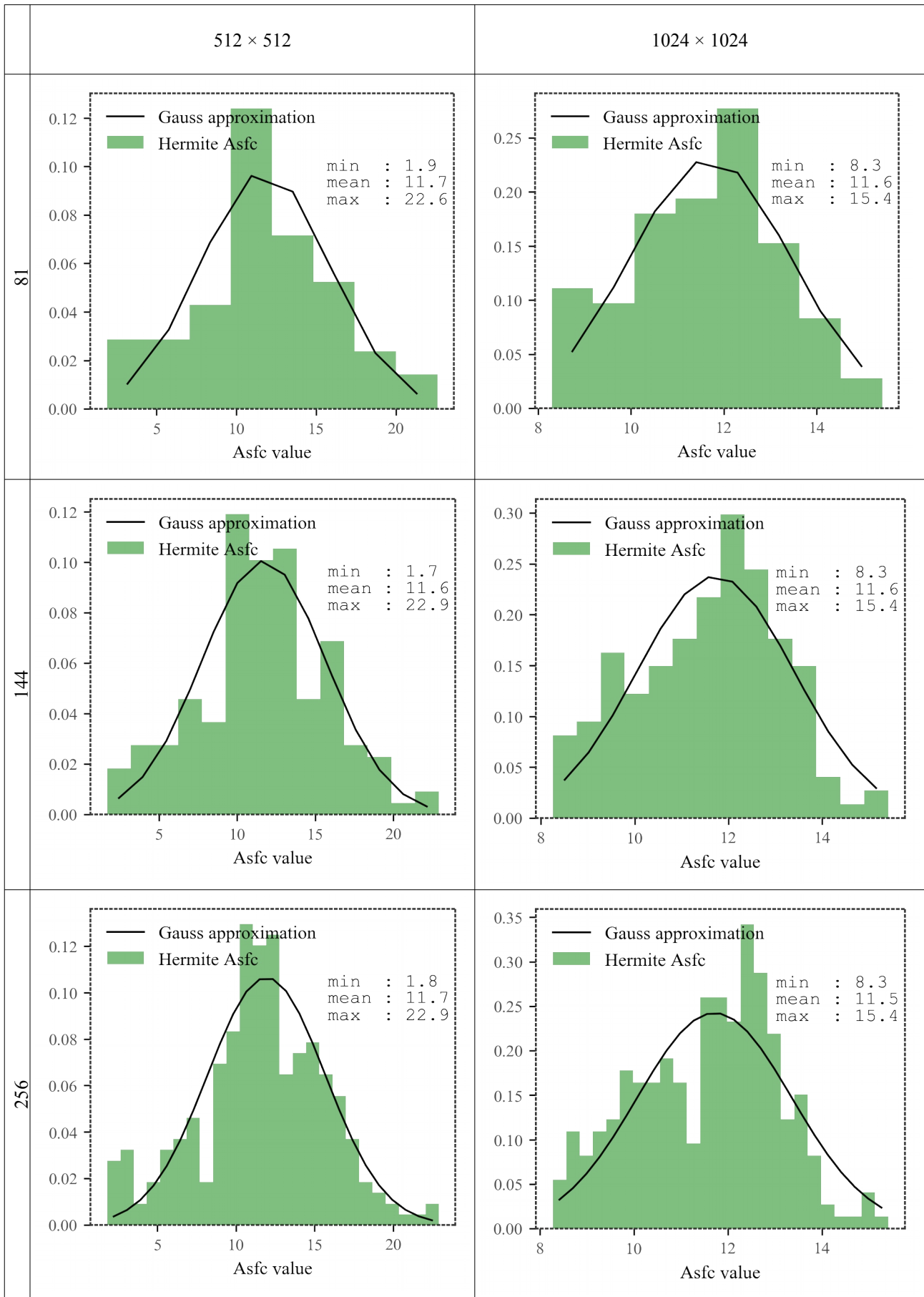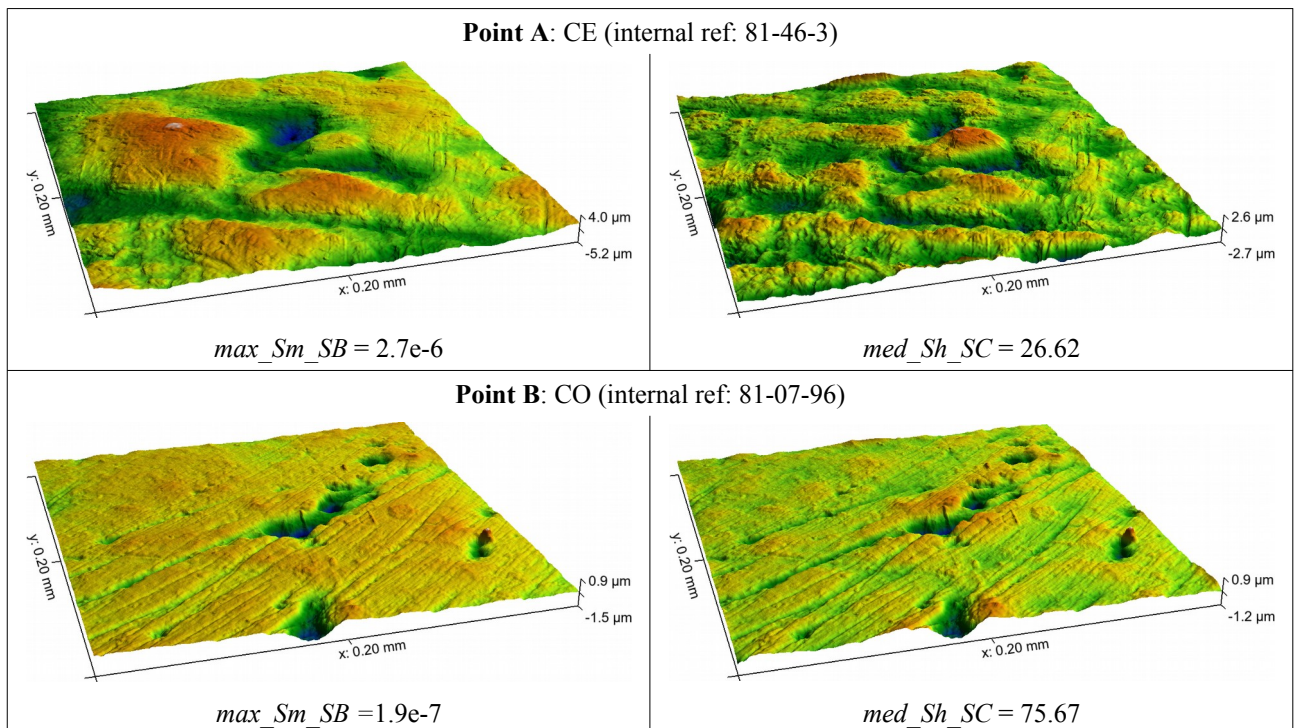
Figure S13. Effect of sample size and subsurface size on *Asfc*

2. Results

All links below were verified on April 28, 2018.

2.1. T1, Old World Monkeys

Colobus polykomos, CO	*Cercocebus atys lunulatus,* CE	*Papio hamadryas,* PA
Michal Sloviak https://www.biolib.cz/en/image/id230348/	Michal Sloviak https://www.biolib.cz/en/image/id131978/	Lubomír Prause https://www.biolib.cz/en/image/id287925/

Point A: CE (internal ref: 81-46-3)

max_Sm_SB = 2.7e-6

med_Sh_SC = 26.62

Point B: CO (internal ref: 81-07-96)

max_Sm_SB =1.9e-7

med_Sh_SC = 75.67

SB: cleaned surface - second order polynomial surface. *SC*: cleaned surface - eighth order polynomial surface.

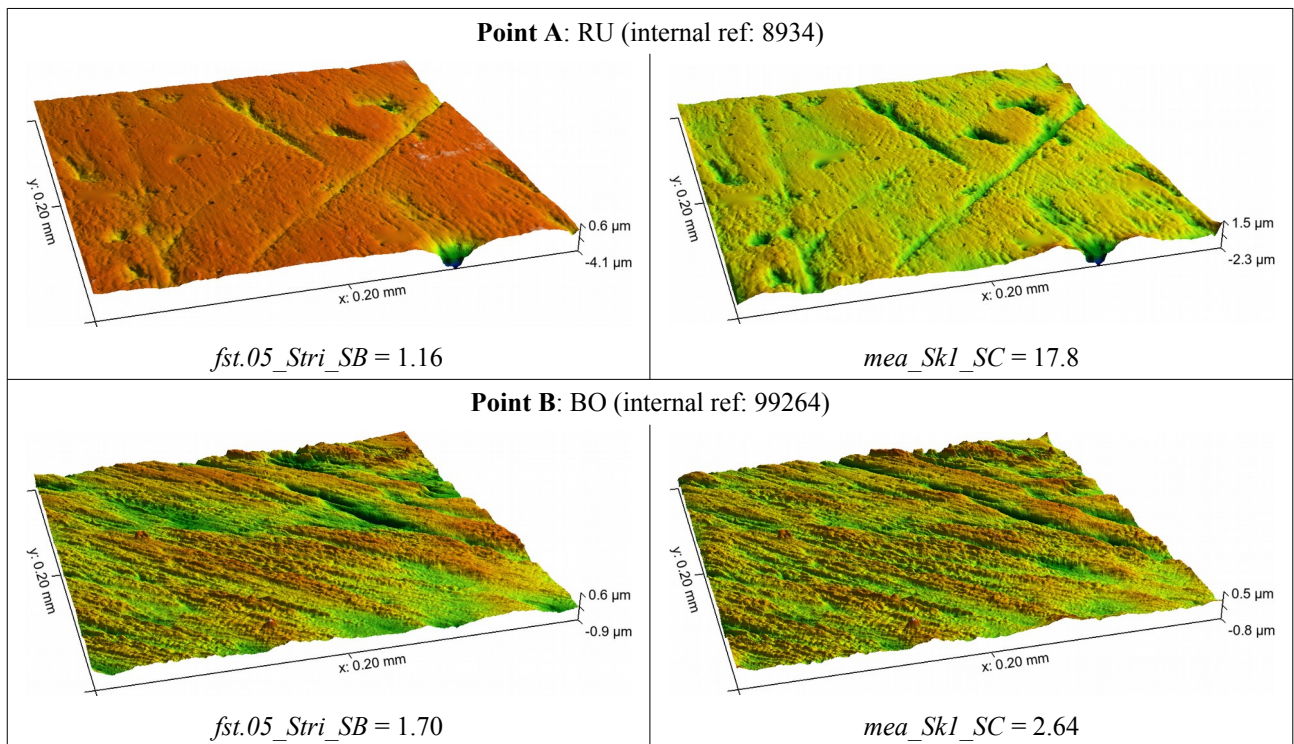*Sm*: (sub)surface mean. *Sh:* percentage of surface nearly horizontal.

max: absolute maximum. *med*: median value.

→ Point A exhibits more elevated samples on the *SB* surface than Point B but less flat zones on *SC*.

2.2. T2, European Ruminants

Cervus elaphus, CE	*Bos Taurus (Camargue),* BO	*Rupicapra rupicapra (*Alpes*),* RU
Luc Viatour https://commons.wikimedia.org/w/index.php?curid=16566391	No known author https://www.lestaxinomes.org/media1991	Michal Sloviak https://www.biolib.cz/en/taxonimage/id125571/?taxonid=249237

Point A: RU (internal ref: 8934)

fst.05_Stri_SB = 1.16

mea_Sk1_SC = 17.8

Point B: BO (internal ref: 99264)

fst.05_Stri_SB = 1.70

mea_Sk1_SC = 2.64

SB: cleaned surface - second order polynomial surface. *SC*: cleaned surface - eighth order polynomial surface.

Stri: anisotropy ratio. *Sk1:* percentage of surface above 85% of the height amplitude (once cleared from the first 15%)
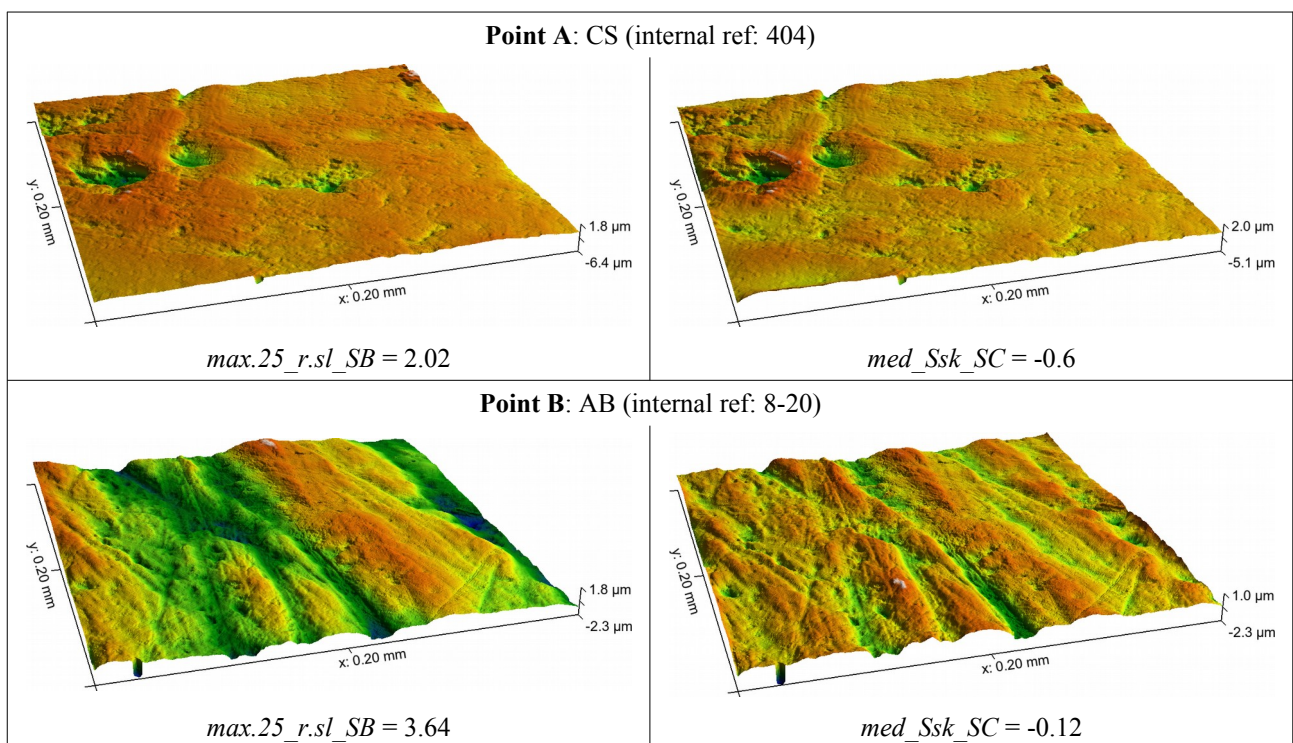
fst.05: quantile 5%. *mea*: mean value.

→ Point A exhibits less striations on the *SB* surface than Point B but more elevated plateaus on *SC*.

2.3. T3, African Ruminants

Alcelaphus buselaphus, AB	*Tragelaphus scriptus,* TS	*Cephalophus silvicultor,* CS
No known author https://commons.wikimedia.org/w/index.php?curid=5304683	Bernard Dupond https://commons.wikimedia.org/w/index.php?curid=40724289	Klaus Rudloff https://www.biolib.cz/en/image/id260658/

Point A: CS (internal ref: 404)

max.25_r.sl_SB = 2.02

med_Ssk_SC = -0.6

Point B: AB (internal ref: 8-20)

max.25_r.sl_SB = 3.64
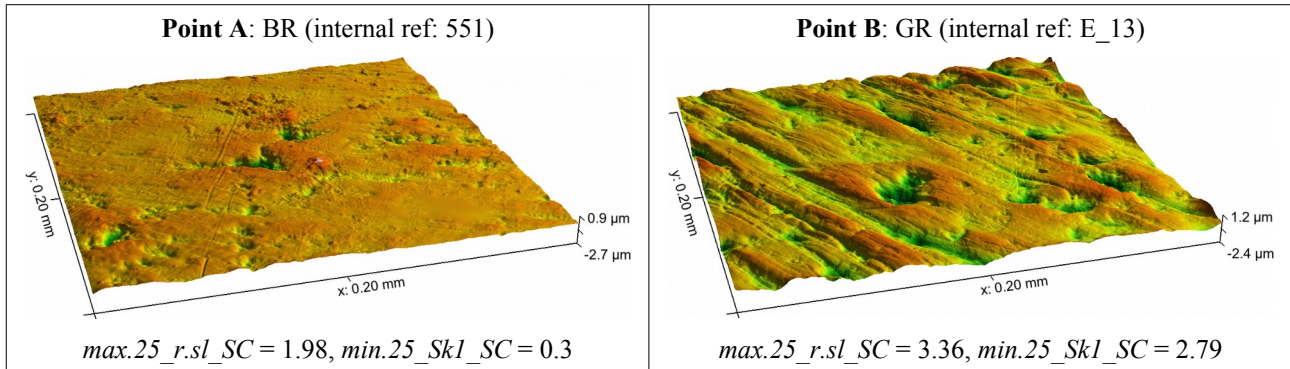

med_Ssk_SC = -0.12

SB: cleaned surface - second order polynomial surface. *SC*: cleaned surface - eighth order polynomial surface.

r.sl: anisotropy ratio based on f_{ACF} slopes. *Ssk:* skewness

max.25: mean above last quartile. *med*: median value.

→ Point A exhibits less striation on the *SB* surface than Point B but more height distribution asymmetry on *SC* samples.

2.4. Q1, Cervus

Point A: BR (internal ref: 551)	**Point B**: GR (internal ref: E_13)
max.25_r.sl_SC = 1.98, *min.25_Sk1_SC* = 0.3	*max.25_r.sl_SC* = 3.36, *min.25_Sk1_SC* = 2.79

SC: cleaned surface - eighth order polynomial surface.

r.sl: anisotropy ratio based on f_{ACF} slopes. *Sk1:* percentage of surface above 85% of the height amplitude (once cleared from the first 15%)

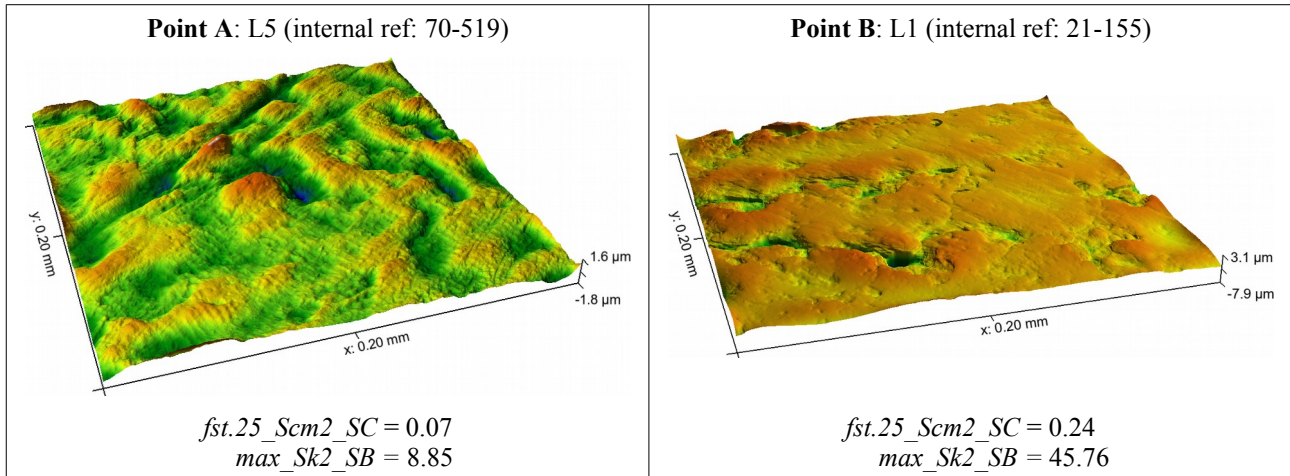*max.25*: mean above last quartile. *min.25*: mean below first quartile

→ Point A exhibits less striation and less elevated samples than Point B. Even if the Point A samples seem flatter, when the heights of a sample are normalized, there are fewer heights above 85% than for Point B samples.
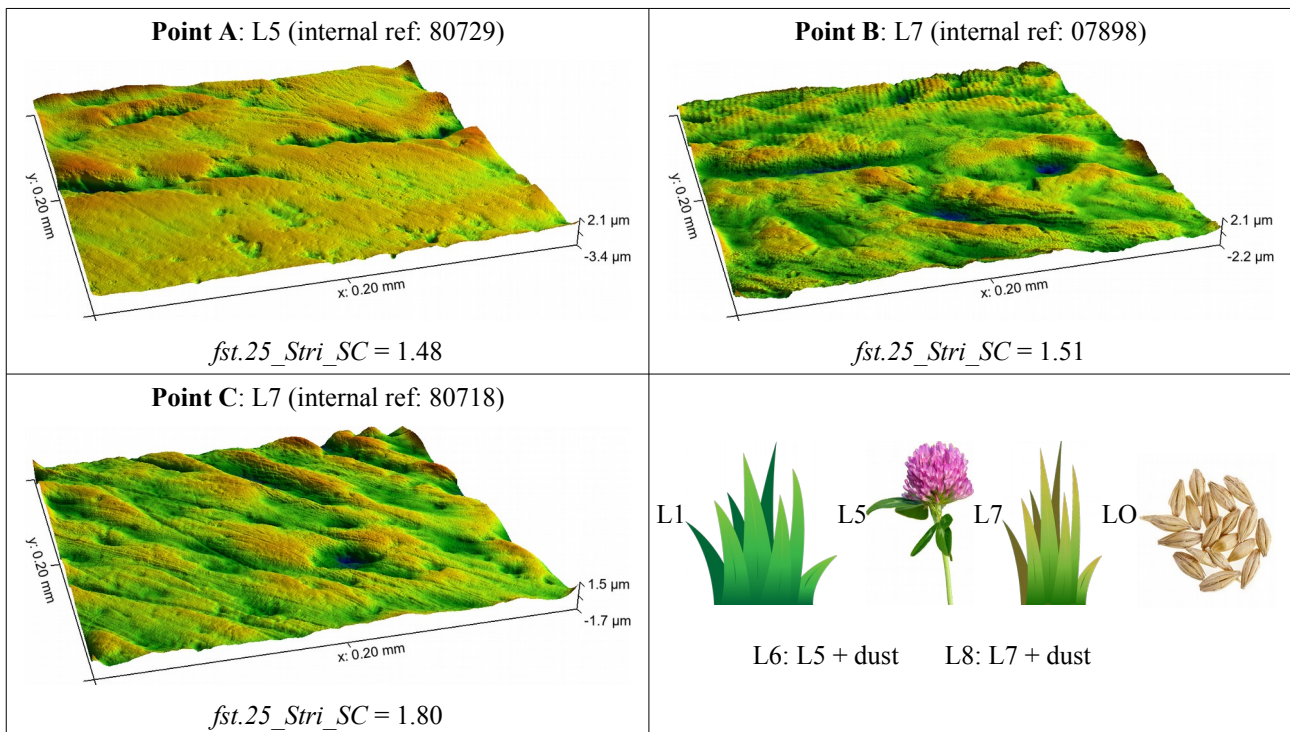
2.5. T4, Cervids

Alces alces, AA	*Capreolus capreolus,* CC	*Cervus elaphus,* CE
Jerzy Strzelecki https://commons.wikimedia.org/wiki/File:Alces_alces_14(js),_Biebrza_National_Park_(Poland).jpg	Lucille Billon https://inpn.mnhn.fr/espece/cd_nom/61057	No known author http://www.mapama.gob.es/es/red-parques-nacionales/nuestros-parques/donana/visita-virtual/fauna/Ciervo.aspx

2.6. Q2, Browse, Grass and Dust (Sheep Experiment)

Point A: L5 (internal ref: 70-519)

fst.25_Scm2_SC = 0.07
max_Sk2_SB = 8.85

Point B: L1 (internal ref: 21-155)

fst.25_Scm2_SC = 0.24
max_Sk2_SB = 45.76

2.7. T5, Seeds, Browse, and Grass (Sheep Experiment)

Point A: L5 (internal ref: 80729)

fst.25_Stri_SC = 1.48

Point B: L7 (internal ref: 07898)

fst.25_Stri_SC = 1.51

Point C: L7 (internal ref: 80718)

fst.25_Stri_SC = 1.80

L1   L5   L7   LO

L6: L5 + dust   L8: L7 + dust

References

1. Francisco, A.; Blondel, C.; Brunetière, N.; Ramdarshan, A.; Merceron, G. Enamel surface topography analysis for diet discrimination. A methodology to enhance and select discriminative parameters. *Surf. Topogr. Metrol. Prop.* **2017**, doi:10.1088/2051-672X/aa9dd3.