ON THE PRECISION OF CURVE LENGTH ESTIMATION IN THE PLANE

Ana I. Gómez✉, Marcos Cruz and Luis M. Cruz-Orive

Department of Mathematics, Statistics and Computation, Faculty of Sciences. University of Cantabria, E-39005. Santander, Spain.
e-mail: gomezanab@gmail.com, marcos.cruz@unican.es, luis.cruz@unican.es

ABSTRACT

The estimator of planar curve length based on intersection counting with a square grid, called the Buffon-Steinhaus estimator, is simple, design unbiased and efficient. However, the prediction of its error variance from a single grid superimposition is a non trivial problem. A previously published predictor is checked here by means of repeated Monte Carlo superimpositions of a curve onto a square grid, with isotropic uniform randomness relative to each other. Nine curvilinear features (namely flattened DNA molecule projections) were considered, and complete data are shown for two of them. Automatization required image processing to transform the original tiff image of each curve into a polygonal approximation consisting of between 180 and 416 straight line segments or 'links' for the different curves. The performance of the variance prediction formula proved to be satisfactory for practical use (at least for the curves studied).

Keywords: Buffon-Steinhaus estimator, Cauchy estimator, curve length, error variance prediction, Monte Carlo resampling, test grid.

INTRODUCTION

Consider a bounded, planar, piecewise smooth curve $Y \subset \mathbb{R}^2$ of finite length B, which is the target parameter. The Buffon-Steinhaus unbiased estimator of B is based on intersection counting with a square grid of test lines which is IUR relative to Y, namely,

$$\widehat{B} = \frac{\pi}{4} \cdot T \cdot I, \qquad (1)$$

(Steinhaus, 1930 – for references see Baddeley and Jensen, 2005) where T is the gap length between test lines and I the total number of intersections. The abbreviation 'IUR', introduced in Miles and Davy (1976), means 'isotropic uniform random' – details are given below. In this paper we address the problem of predicting the variance $\mathrm{Var}(\widehat{B})$ from a single grid superimposition. The interest in the problem is old. Moran (1966) gives an exact variance expression when Y is a straight line segment. For a general curve, he gives an approximation for the variance component due to the grid position, but not for the one due to orientation – see also Cruz-Orive (1989). To our knowledge the only estimators hitherto available for both variance components were given in Cruz-Orive and Gual-Arnau (2002). Such estimators are generally not unbiased – they are only theoretical approximations. Our purpose is to check the performance of such approximations by means of Monte Carlo simulations.

Often Y is a model for profile boundary (hence the notation 'B' for its length), namely the boundary of the intersection between a closed surface (*e.g.*, a cell membrane, or a grain boundary) and a sectioning plane. A planar curve may also be the orthogonal projection of a spatial curve onto an observation plane, or even a flexible linear feature in space flattened onto a planar surface, in which case the real and the observed lengths should approximately coincide. This is the case for the material used in this paper, namely DNA molecules which appear as open linear features on the observation plane (Podestà *et al.*, 2004). Fig. 1a below, kindly provided in tiff format by Professor Alessandro Podestà, is Fig. 1(A) from the latter paper.

The laborious steps leading to the conversion of the tiff image features into polygonal curves in vector graphics (Section *Processing of the curve images*), which constitute necessary prerequisites to perform automatic measurements, suggest that the intersection counting method to estimate feature length directly on the original images is an attractive option – first because it is fast, design unbiased and efficient, second because the error variance can be predicted and last, but not least, because the processed curve lengths will seldom coincide with the lengths of the original projections (automatic measurements are generally biased by a wide variety of artifacts depending on curve shape, image resolution, etc.).
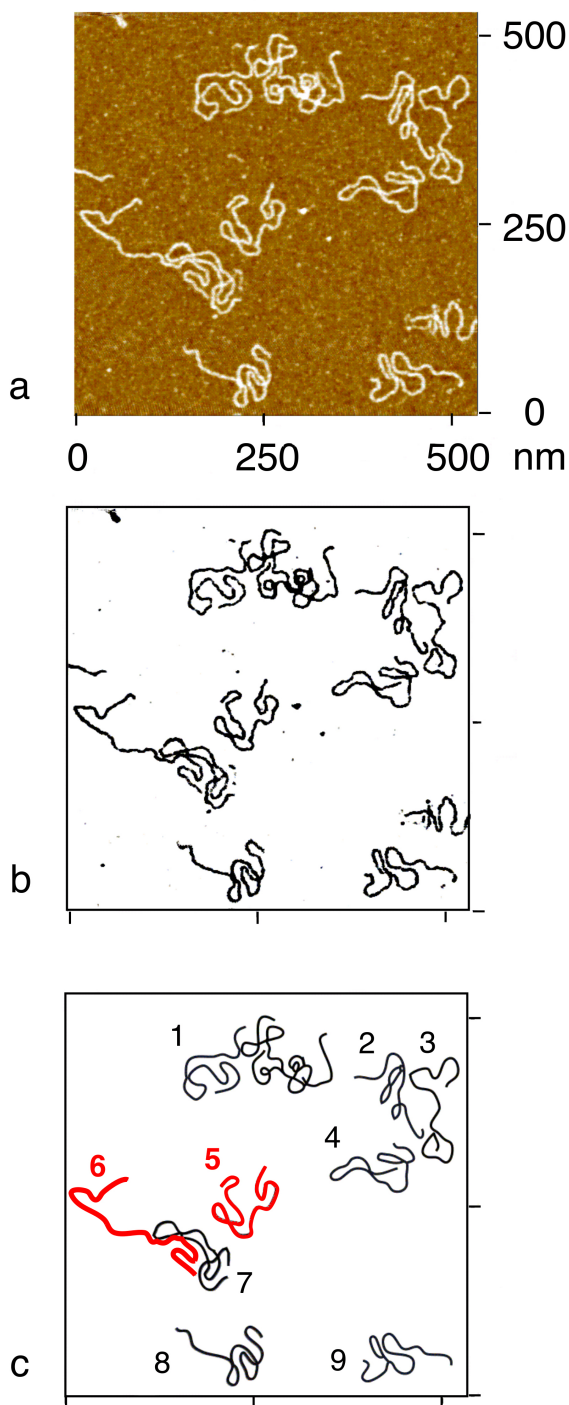
a

b

c

Fig. 1. *(a) The linear features, used as model curves in this paper, represent flattened projections of DNA molecules onto the observation plane. Redrawn from Podestà et al. (2004), with permission from John Wiley and Sons. (b) Raster bitmap image version of the linear features. (c) Vector graphics image with polygonal approximations of the original curves, ready for automatic Monte Carlo experiment. The two curves marked in red were selected for the results shown in Figs. 3, 4.*

CURVE MODEL AND LENGTH ESTIMATION

CURVE MODEL

Each of the selected curvilinear features shown in Fig. 1, namely each of the selected DNA molecules, was converted from tiff format into a polygonal curve in vector graphics following the steps described in the section *Processing of the curve images*. A fixed polygonal curve Y, henceforth called a 'curve', for short, may be represented as follows,

$$Y = \bigcup_{i=1}^{N} y_i ,\qquad (2)$$

where y_i denotes the ith straight line segment or 'link' of Y, and N is the finite total number of links. The total length of Y is

$$B = \sum_{i=1}^{N} b_i ,\qquad (3)$$

where b_i, denotes the length of y_i.

Fix a rectangular frame Ox_1x_2 in the plane of the curve. The position and orientation of the curve is determined by a unit vector (x, ω) rigidly attached to the curve, emanating from a point $x \in \mathbb{R}^2$, and making an angle $\omega \in [0, 2\pi)$ with Ox_1. The point $x = (x_1, x_2)$ is called the associated point (AP) of Y, (Miles, 1974, see also Baddeley and Jensen, 2005) and the vector (x, ω) is called the associated vector (AV). The choice of the AV is arbitrary, but once it is chosen it must remain rigidly attached to the curve. When the AV is at its initial position ($x = 0$, $\omega = 0$), then the corresponding initial position of the curve is denoted by $Y_{0,0}$, so that the AP of $Y_{0,0}$ is at the origin O. Here this AP was the lower left corner of the smallest rectangle (with its sides parallel to the reference axes) enclosing $Y_{0,0}$, see Fig. 2a. Under a rigid motion defined by the vector (x, ω), (namely a rotation ω about the origin followed by a translation x, or equivalently a translation x followed by a rotation ω about the point x) the curve $Y_{0,0}$ is transformed into the curve $Y_{x,\omega}$ see Fig. 2b. The triplet (x_1, x_2, ω) constitute the 'coordinates' of the curve.

The generation of a IUR curve in the plane requires the use of the motion invariant density associated with the coordinates of the curve. This density is called the kinematic density (Santaló, 1976), namely,

$$dY_{x,\omega} = dx\, d\omega ,\qquad (4)$$

where $dx = dx_1\, dx_2$ is the area element in the plane and $d\omega$ is the arc element in the unit circle. In practice, the preceding concept implies that the AP of the curve

2

must be uniform random (UR) in any bounded region of the plane, whereas the orientation of the AV must be isotropic random (IR) namely UR in the unit circle, and independent from the position of the AP.
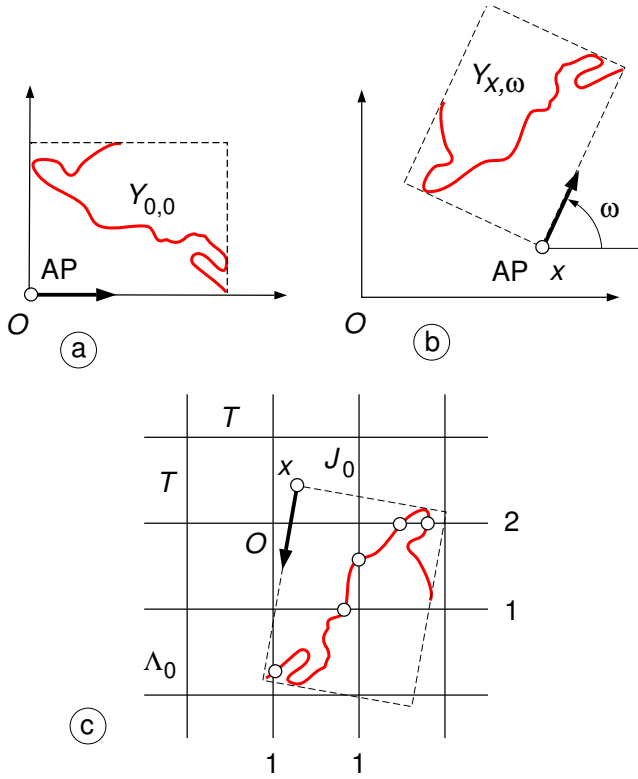

Fig. 2. *(a) Curve number 6 from Fig. 1c, with its rigidly attached associated point (AP) and associated vector (arrow). Each of the straight line link segments constituting the curve (indistinguishable by eye) share the same AP and the same associated vector. (b) Result of applying a rigid motion to the curve, namely a translation of the AP to the point x followed by a rotation ω about x. (c) A IUR superimposition of the curve onto a square grid Λ_0. The AP is UR in the fundamental tile J_0 of the grid, whereas the rotation angle ω is independent and UR in the interval $[0,2\pi)$. Here the relevant intersection counts read: $I_{11} = 1$, $I_{12} = 1$, $I_{21} = 1$, $I_{22} = 2$. The procedure was programmed in Python to generate automatic replications.*

CAUCHY (ONE STAGE) ESTIMATION OF CURVE LENGTH

A special case of Cauchy's projection formula (Santaló, 1976) expresses the length B of the curve $Y_{0,\omega}$ in terms of its total orthogonal projected length $l(\omega)$ onto Ox_1, with all points counted in their multiplicity, see Cruz-Orive (1989, Fig. 1), or Cruz-Orive and Gual-Arnau (2002, Fig. 4a). Let $\{\alpha_1, \alpha_2, ..., \alpha_N\}$ denote the fixed angles of the oriented links of $Y_{0,0}$ with Ox_1. Then,

$$l(\omega) = \sum_{i=1}^{N} b_i |\cos(\alpha_i - \omega)| . \qquad (5)$$

Integration from $\omega = 0$ to $\omega = 2\pi$ yields Cauchy's formula,

$$B = \frac{1}{4} \int_0^{2\pi} l(\omega) \mathrm{d}\omega . \qquad (6)$$

Suppose that the angle ω is UR in the interval $[0, 2\pi)$, with probability element,

$$\mathbb{P}(\mathrm{d}\omega) = \frac{\mathrm{d}\omega}{2\pi}, \ \omega \in [0, 2\pi) , \qquad (7)$$

and suppose also that $l(\omega)$ can be measured exactly. Then, by Cauchy's formula

$$B = \frac{\pi}{2} \mathbb{E}\{l(\omega)\} , \qquad (8)$$

and therefore,

$$\widetilde{B}(\omega) = \frac{\pi}{2} l(\omega) \qquad (9)$$

is an unbiased estimator (UE) of B. Here ω may alternatively be UR in the interval $[0, \pi)$, *i.e.*, $\omega \sim$ UR$[0, \pi)$, because $l(\omega) = l(\omega + \pi)$ for all ω. Estimation precision may be gained by measuring also the projected length $l(\omega + \pi/2)$ of $Y_{0,\omega}$ onto the Ox_2 axis, whereby,

$$\widehat{B}(\omega) = \frac{\pi}{4} \big[l(\omega) + l(\omega + \pi/2) \big] , \qquad (10)$$

is also unbiased for B, and it can be expected to be more precise than $\widetilde{B}(\omega)$. Note that here we may take $\omega \sim$ UR$[0, \pi/2)$ because $\widehat{B}(\omega) = \widehat{B}(\omega + \pi/2)$.

The estimator $\widehat{B}(\omega)$ is based on $n = 2$ systematic observations from the measurement function $l(\omega)$ in the interval $[0, \pi)$, *i.e.*, with period $\pi/2$, in the unit semicircle. A predictor of Var$(\widehat{B})$ is available from Eq. 4.4 of Cruz-Orive and Gual-Arnau (2002) with $r = \pi$, $n = 2$ and $m = 1$. Here m is a parameter of the global model adopted for the covariogram of the measurement function $l(\omega)$. In principle, the smoother $l(\omega)$, the more appropriate should be the choice $m = 1$. The local error is $\widehat{v}_2 = 0$ because $l(\omega)$ is measured without error. Further, $C_0 - C_1 = [l(\omega) - l(\omega + \pi/2)]^2/4$. Thus,

$$\mathrm{var}\{\widehat{B}(\omega)\} = \frac{\pi^2}{240} \cdot [l(\omega) - l(\omega + \pi/2)]^2 . \qquad (11)$$

Note: In this paper a true variance is denoted by Var$(\cdot)$, whereas the corresponding predictor or estimator is denoted by var$(\cdot)$.

BUFFON-STEINHAUS (TWO STAGE) ESTIMATION OF CURVE LENGTH

For a given orientation ω of the curve, the total orthogonal projected length $l(\omega)$ and $l(\omega + \pi/2)$ can be estimated without bias at a second stage by intersection counting with a square grid of test lines. More precisely:

(i) Fix a square grid Λ_0 of test lines parallel to the Ox_1x_2 axes, with gap length $T > 0$. It is convenient to fix the origin O at a central vertex of the grid. The adopted fundamental tile of the grid is the square $J_0 = [0,T)^2$, see Fig. 2c.

(ii) Rotate the curve $Y_{0,0}$ (with its AP at O) isotropically about O into $Y_{0,\omega}$, where $\omega \sim \text{UR}[0,2\pi)$. Next, shift the AP of $Y_{0,\omega}$ into a point $x \sim \text{UR}(J_0)$, thus translating the curve together into $Y_{x,\omega}$, (Fig. 2c).

(iii) Score the intersections counts $\{I_{1j}(\omega,x), j = 1,2,...,n_1\}$ between the curve and the vertical lines of the grid in an ordered sequence, where $I_{11}(\omega,x), I_{1n_1}(\omega,x)$ denote the first, and the last non zero counts, respectively. Score similarly the intersection counts $\{I_{2j}(\omega + \pi/2,x), \ j = 1,2,..,n_2\}$ with the horizontal lines.

The Buffon-Steinhaus estimator $\widehat{B}(\omega,x)$ of the curve length B is the result of replacing the projection lengths in the right hand side of Eq. 10, (i.e of the Cauchy estimator), with the corresponding Cavalieri type estimators (Eqs. 12b, 12c below) based on the intersection counts, namely,

$$\widehat{B}(\omega,x) = \frac{\pi}{4}\left[\widehat{l}(\omega,x) + \widehat{l}(\omega + \pi/2,x)\right], \qquad (12a)$$

$$\widehat{l}(\omega,x) = T\sum_{j=1}^{n_1} I_{1j}(\omega,x), \qquad (12b)$$

$$\widehat{l}(\omega + \pi/2,x) = T\sum_{j=1}^{n_2} I_{2j}(\omega + \pi/2,x). \qquad (12c)$$

The estimator $\widehat{B}(\omega,x)$ is therefore a two stage UE of B. The fist stage design is the Cauchy design generating two mutually perpendicular, total projections of the curve, whereas the second stage incorporates Cavalieri sampling to estimate the lengths of these projections by intersection counting. By the unbiasedness of the Cavalieri design, for each $\omega \in [0,2\pi)$ the conditional mean of the second stage estimator is equal to the fist stage (Cauchy) estimator, namely,

$$\mathbb{E}_x\{\widehat{B}(\omega,x)|\omega\} = \widehat{B}(\omega). \qquad (13)$$

Because the Cauchy estimator is unbiased, namely because $\mathbb{E}_\omega\{\widehat{B}(\omega)\} = B$, it follows that $\mathbb{E}\{\widehat{B}(\omega,x)\} = \mathbb{E}_\omega\mathbb{E}_x\{\widehat{B}(\omega,x)|\omega\} = B$, which verifies the unbiasedness of $\widehat{B}(\omega,x)$.

The exact variance of $\widehat{B}(\omega,x)$ may be represented using the standard variance decomposition formula. For simplicity set,

$$\begin{aligned}\widehat{B}_1 &= \widehat{B}(\omega), \\ \widehat{B}_2 &= \widehat{B}(\omega,x).\end{aligned} \qquad (14)$$

Then,

$$\begin{aligned}\text{Var}(\widehat{B}_2) &= \text{Var}_\omega\{\mathbb{E}_x(\widehat{B}_2|\omega)\} + \mathbb{E}_\omega\{\text{Var}_x(\widehat{B}_2|\omega)\} \\ &= \text{Var}(\widehat{B}_1) + \mathbb{E}_\omega\{\text{Var}_x(\widehat{B}_2|\omega)\}. \qquad (15)\end{aligned}$$

The first term in the right hand side of the preceding identity represents variance component due to orientations, whereas the second term represents the component due to intersection counting, averaged over orientations. The latter component, also called the nugget, or local error component, may be predicted using the standard Matheron's formula for systematic sampling along an axis (generally called Cavalieri sampling). For each $\omega \in [0,2\pi)$ set

$$\begin{aligned}\sigma_1^2 &= \text{Var}_x\left\{\widehat{l}(\omega,x)|\omega\right\}, \\ \sigma_2^2 &= \text{Var}_x\left\{\widehat{l}(\omega + \pi/2,x)|\omega\right\},\end{aligned} \qquad (16)$$

namely the local error variances due to intersection counting with the vertical and the horizontal test lines of the grid, respectively. Then, the second term in the right hand side of Eq. 15 is estimated as follows,

$$\text{var}(\widehat{B}_2|\omega) = \left(\frac{\pi}{4}\right)^2 (\widehat{\sigma}_1^2 + \widehat{\sigma}_2^2), \qquad (17a)$$

$$\widehat{\sigma}_i^2 = \frac{T^2}{12} \cdot (3C_{0i} - 4C_{1i} + C_{2i}), \ i = 1,2, \ n_i \geq 3, \qquad (17b)$$

$$\widehat{\sigma}_i^2 = \frac{T^2}{6} \cdot (C_{0i} - C_{1i}), \ i = 1,2, \ n_i = 2, \qquad (17c)$$

$$C_{ki} = \sum_{j=1}^{n_i-k} I_{ij}I_{i,j+k}, \ k = 0,1,2; \ i = 1,2, \qquad (17d)$$

where,

$$\begin{aligned}I_{1j} &= I_{1j}(\omega,x), \\ I_{2j} &= I_{2j}(\omega + \pi/2,x).\end{aligned} \qquad (18)$$

On the other hand, the first term in the right hand side of Eq. 15, namely the orientations or Cauchy component, cannot be estimated with Eq. 11 directly

4

because the curve projections are no longer measured exactly. Instead,

$$\text{var}_2(\widehat{B}_1) = \frac{\pi^2}{240} \cdot \left\{ \left[\widehat{l}(\omega, x) - \widehat{l}(\omega + \pi/2, x) \right]^2 \right. \tag{19}$$
$$\left. - (\widehat{\sigma}_1^2 + \widehat{\sigma}_2^2) \right\},$$

where the subscript '2' in $\text{var}_2(\cdot)$ indicates that the estimator is computed from the two stage data, namely from the intersection counts. Thus the total variance predictor of the two stage estimator is (Cruz-Orive and Gual-Arnau, 2002)

$$\text{var}(\widehat{B}_2) = \text{var}_2(\widehat{B}_1) + \text{var}(\widehat{B}_2|\omega) . \tag{20}$$

REMARKS

1. Provided that the curve $Y_{0,\omega}$ is connected, its orthogonal projection onto Ox_1 will be a bounded interval Y'_ω, say. Define the integer valued measurement function $I_{Y'_\omega}(z)$ as the number of intersection between $Y_{0,\omega}$ and the vertical straight line $x_1 = z$. Then its integral

$$l(\omega) = \int_{\mathbb{R}} I_{Y'_\omega}(z) \mathrm{d}z \tag{21}$$

   is the total orthogonal projected length defined above, and Eq. 12b is the standard Cavalieri estimator of it. Moreover, because the measurement function $I_{Y'_\omega}(z)$ is integer valued it will exhibit jumps, and therefore its smoothness constant will be $q = 0$ (Kiêu *et al.*, 1999; García-Fiñana and Cruz-Orive, 2004). In consequence, the $\widehat{\sigma}_i^2$ are computed with $q = 0$.

2. While $l(\omega) = l(\omega + \pi)$ for all ω, the corresponding Cavalieri estimators, see Eqs. 12b, 12c, will in general be different for each ω. This justifies the choice of the range $[0, 2\pi)$ for ω in step (ii) of the preceding subsection.

PROCESSING OF THE CURVE IMAGES

Here we describe the steps involved in the conversion of each original tiff curve image (Fig. 1a) into a polygonal curve in Scalar Vector Graphics (SVG, Fig. 1c) which can be imported into a programming environment to perform automatic Monte Carlo experiments. Preference was given to free, GPL licensed software.

Step 1. The original tiff image (containing all the relevant curves) was edited using the GIMP software package (http://www.gimp.org/). The image was desaturated and converted into a black-and-white (B/W) one. Brightness, contrast and B/W colour threshold were set at -46, $+6$ and 108, respectively. The resulting image was a raster (or 'bitmap') image consisting of pixels which were coloured in black (Fig. 1b). Each curve, or group of nearly adjacent curves, was selected and recorded separately into a JPEG output file, which was submitted to the next step.

Step 2. Each of the preceding (bitmap) JPEG images was submitted to AutoTrace (http://autotrace.sourceforge.net/) to transform it into vector graphics. A 'centerline' option was used whereby a pixelized curve was approximated by a 'spine' curve consisting of the union of Bèzier arcs. The output files were exported as SVG files.

Step 3. The preceding SVG files were edited manually with the aid of Inkscape (http://inkscape.org/) to remove background noise, and to eventually adjust the curve arcs properly.

Step 4. The result (Fig. 1c) was imported into Blender V 2.68a (http://www.blender.org/) to convert it into a connected polygonal of small linear segments or 'links' with known endpoint XY coordinates. The number of links ranged between 180 and 416, depending on the local curvatures of the original curves.

With the final polygonal curves a Python script was written to compute curve length (Eq. 3), to rotate and shift the curves as prescribed in the preceding section, and to perform the Monte Carlo procedures described in the next section.

MONTE CARLO EXPERIMENT TO TEST THE PERFORMANCE OF THE VARIANCE PREDICTORS

PURPOSES

From the curvilinear features (DNA molecules) shown in Fig. 1a, the two SVG curves numbered 5 and 6 in Fig. 1c were selected to illustrate the results. The choice was based on the fact that, qualitatively, these two curves give the visual impression of being 'fairly isotropic' and 'fairly anisotropic', respectively. For each curve the following aspects were studied.

(1) Behaviour of the Cauchy estimator $\widetilde{B}(\omega)$ as a function of the angle $\omega \in [0, \pi)$. In polar coordinates the graph of $l(\omega)$ is the rose of total orthogonal projected lengths of the curve, (*i.e.*, the 'rose of projections', for short).

(2) Behaviour of the Cauchy estimator $\widehat{B}(\omega)$ based on two mutually orthogonal projections.

(3) Replicates of the Buffon-Steinhaus estimator $\widehat{B}(\omega, x)$, which is based on intersection counts with a square grid, as a function of the mean total number of intersections $\mathbb{E}(I)$.

(4) Comparison of the variance predictor $\text{var}\{\widehat{B}(\omega)\}$ of the Cauchy estimator, see Eq. 11, against the empirical variance $\text{Var}_e\{\widehat{B}(\omega)\}$.

(5) Comparison of the variance predictor $\text{var}\{\widehat{B}(\omega, x)\}$ of the Buffon-Steinhaus estimator, and of the two variance component predictors in the right hand side of Eq. 20, against the corresponding empirical variances.

In the corresponding graphs, each empirical error variance was divided by the square of the true curve length in order to represent the square coefficient of error. For instance,

$$\text{CE}_e^2\{\widehat{B}(\omega)\} = \frac{\text{Var}_e\{\widehat{B}(\omega)\}}{B^2} \,. \qquad (22)$$

For convenience the corresponding variance predictors were normalized in the same way, *e.g.*,

$$\text{ce}^2\{\widehat{B}(\omega)\} = \frac{\text{var}\{\widehat{B}(\omega)\}}{B^2} \,. \qquad (23)$$

If the interest was focused on the sample coefficient of error itself, however, then B^2 should be replaced with its sample version $\widehat{B}^2(\omega)$ in the denominator of the preceding expression.

CAUCHY ESTIMATOR

For each of the selected curves, the Cauchy estimator $\widehat{B}(\omega)$ was computed according to Eq. 10 at each of $M = 32^2 = 1024$ values of ω in the interval $[0, \pi/2)$, namely

$$\begin{aligned} &\{\omega_k = (U + k - 1)\frac{\pi}{2M}, \ k = 1, 2, ..., M\} \,, \\ &U \sim \text{UR}[0, 1) \,. \end{aligned} \qquad (24)$$

From the corresponding M replicates of $\widehat{B}(\omega)$, the empirical mean and the error variance of the estimator were computed respectively as follows,

$$\mathbb{E}_e\{\widehat{B}(\omega)\} = \frac{1}{M}\sum_{k=1}^{M}\widehat{B}(\omega_k) \,, \qquad (25a)$$

$$\text{Var}_e\{\widehat{B}(\omega)\} = \frac{1}{M}\sum_{k=1}^{M}\left[\widehat{B}(\omega_k) - \mathbb{E}_e\{\widehat{B}(\omega)\}\right]^2 \,. \qquad (25b)$$

On the other hand the corresponding M replicates $\{\text{var}\{\widehat{B}(\omega_k)\}, k = 1, 2, ..., M\}$ of the variance predictor given by Eq. 11 were computed for comparison against the empirical variance.

The Cauchy estimator $\widetilde{B}(\omega)$ was also computed in a similar way at $M = 1024$ points in the interval $[0, \pi)$. Here, however, no variance predictor exists because the estimator is based on a single projection.

BUFFON-STEINHAUS ESTIMATOR

For a given curve $Y_{0,0}$ and a fixed square grid Λ_0 of test lines of gap $T > 0$, a total of $M = K^2 = 32^2 = 1024$ replicates of the Buffon-Steinhaus estimator $\widehat{B}(\omega, x)$ were computed from Eq. 12 by means of M superimpositions of the curve onto the grid, (Fig. 2c), according to a systematic design as follows. Recall that the AP of the curve is UR in the fundamental square of the grid, namely $x \in J_0 = [0, T)^2$, whereas ω is UR in $[0, 2\pi)$. The Cartesian coordinates of the $M = K^2$ positions of the AP, *e.g.*,

$$\{(x_{1i}, x_{2j}), \ i, j = 1, 2, ..., K\} \,, \qquad (26)$$

were the vertices of a fine systematic square grid of gap T/K generated within the fundamental square J_0, namely,

$$\begin{aligned} x_{1i} &= (U_1 + i - 1) \cdot T/K \,, \\ x_{2j} &= (U_2 + j - 1) \cdot T/K \,, \end{aligned} \qquad (27)$$

where U_1, U_2 are two independent UR numbers in the interval $[0,1)$. Renumber the resulting sequence of M associated points as $\{x_1, x_2, ..., x_M\}$ in any convenient way. Next, generate M systematic rotation angles in the interval $[0, 2\pi)$, namely,

$$\left\{(U_3 + i - 1)\frac{2\pi}{M}, \ i = 1, 2, ..., M\right\} \,, \qquad (28)$$

where U_3 is a third UR number in the interval $[0, 1)$, independent from U_1, U_2. Now, generate a random permutation of the preceding set of M systematic angles, and denote it by $\{\omega_1, \omega_2, ..., \omega_M\}$. Then, the curve coordinates of the M superimpositions of the curve onto the grid are

$$\{(x_k, \omega_k), \ k = 1, 2, ..., M\} \,. \qquad (29)$$

Note that the angle permutations avoid any correlation between curve location (x_k) and orientation (ω_k).

Different values of the gap T of the grid were chosen so that the expected total number $\mathbb{E}(I)$ of intersections did not exceed about 130. With the abbreviation defined in Eq. 14, and bearing in mind

that $\mathbb{E}(\widehat{B}_2) = B$, for a desired value of $\mathbb{E}(I)$ the gap was computed by the following formula,

$$T = 4B/(\pi\mathbb{E}(I)) . \tag{30}$$

For each curve and each value of T, the M replicates

$$\left\{ \widehat{B}(\omega_k, x_k), \ k = 1, 2, .., M \right\} , \tag{31}$$

were obtained as described above, and the empirical mean and variance of the Buffon-Steinhaus estimator were thereby computed similarly as in Eqs. 25a, 25b, namely,

$$\mathbb{E}_e\left\{\widehat{B}(\omega, x)\right\} = \frac{1}{M} \sum_{k=1}^{M} \widehat{B}(\omega_k, x_k) , \tag{32a}$$

$$\mathrm{Var}_e\left\{\widehat{B}(\omega, x)\right\} = \frac{1}{M} \sum_{k=1}^{M} \left[\widehat{B}(\omega_k, x_k) - \mathbb{E}_e\left\{\widehat{B}(\omega, x)\right\} \right]^2 . \tag{32b}$$

From Eqs. 15, 25b, the empirical value of the Cavalieri (intersection counting) contribution was computed as follows,

$$\mathbb{E}_e \mathrm{Var}\{\widehat{B}(\omega, x)|\omega\} = \mathrm{Var}_e\{\widehat{B}(\omega, x)\} - \mathrm{Var}_e\{\widehat{B}(\omega)\} . \tag{33}$$

Further, for each pair (ω_k, x_k) the predictors of the Cavalieri and the Cauchy error variance components were computed from Eq. 17 and 19, respectively, as well as the total error variance predictor given by their sum (Eq. 20).

RESULTS

CAUCHY ESTIMATOR

Each of the two curves studied, and their corresponding roses of projections, are shown at the top of Fig. 3.

The graph of the single projection estimator $\widetilde{B}(\omega)$ is displayed for each curve in Fig. 3a,b, whereas those for the double projection estimator $\widehat{B}(\omega)$ are displayed in Fig. 3c,d. For either curve the superior precision of $\widehat{B}(\omega)$ is apparent.

The more flattened the rose of projections, the more anisotropic is the curve. This is visually reflected by the different degrees of accuracy of each of the two Cauchy estimators for each of the two curves.

The empirical $\mathrm{CE}_e^2\{\widehat{B}(\omega)\}$, (Fig. 3e, f, pink straight lines), computed via Eq. 25b, is negligible

for the first curve ($3.86 \cdot 10^{-5}$), and rather small for the second ($5.06 \cdot 10^{-4}$). The graph of the predictor $\mathrm{ce}^2\{\widehat{B}(\omega)\}$, see Eq. 11, is also displayed for each curve in Fig.3e,f, (red oscillating curves). The corresponding means ($6.00 \cdot 10^{-4}$ and $1.30 \cdot 10^{-3}$, respectively) are shown as red dotted lines. The biases are small in absolute terms ($5.62 \cdot 10^{-4}$ and $7.99 \cdot 10^{-4}$ respectively), but very large in relative terms (14.56 and 1.58 respectively) because the actual empirical errors are small.

BUFFON-STEINHAUS ESTIMATOR

For each of the two curves considered (see top of Fig. 4), the $M = 1024$ Monte Carlo replicates of the Buffon-Steinhaus estimator $\widehat{B}(\omega, x)$, see Eq. 12a, are displayed in Fig. 4a, b, respectively, at each of 10 values of the mean total number of intersections $\mathbb{E}(I)$. As a check of the Monte Carlo procedure, the corresponding means (red dots), computed via Eq. 32a, visually coincide with the known value of B in each case, as expected by unbiasedness. Approximate confidence bands of 95% (coloured in grey), and of 100%, are also displayed for the individual realizations of $\widehat{B}(\omega, x)$.

The empirical $\mathrm{CE}^2\{\widehat{B}(\omega, x)$ was computed for each curve, via Eq. 32b, at each of the 10 values of $\mathbb{E}(I)$ and joined by a polygonal line (thick black line in Fig. 4c, d).

The individual realizations of the corresponding predictor $\mathrm{ce}^2\{\widehat{B}(\omega, x)\}$ (computed via Eq. 20) are also displayed in Fig. 4c, d. The corresponding 10 group means of the predictors were joined by a polygonal curve (thick red line). Approximate 95% (coloured in grey), and 100% confidence bands are displayed together. It is seen that, in the range of $\mathbb{E}(I)$ considered, the empirical curve is always captured by the 95% confidence band. For $\mathbb{E}(I) \approx 50$ the empirical $\mathrm{CE}\{\widehat{B}(\omega, x)\}$ is about 5% for either curve (horizontal, blue broken line). Such degree of accuracy is however unnecessary if the target quantity is the mean length of a population of curves. For a simple random sample of n curves, the stereological contribution to the total square coefficient of error of the population mean length estimate would be reduced to $\mathrm{CE}^2\{\widehat{B}(\omega, x)\}/n$.

The graph of the empirical Cavalieri component $\mathbb{E}_e\mathrm{CE}^2\{\widehat{B}(\omega, x)|\omega\}$, computed via Eq. 33, is shown for each curve in Fig. 4e, f, respectively (thick blue polygonal), together with its mean predictor computed via Eq. 10, (broken blue line). On the other hand, the corresponding Cauchy components are represented in pink and in red (broken line), respectively.
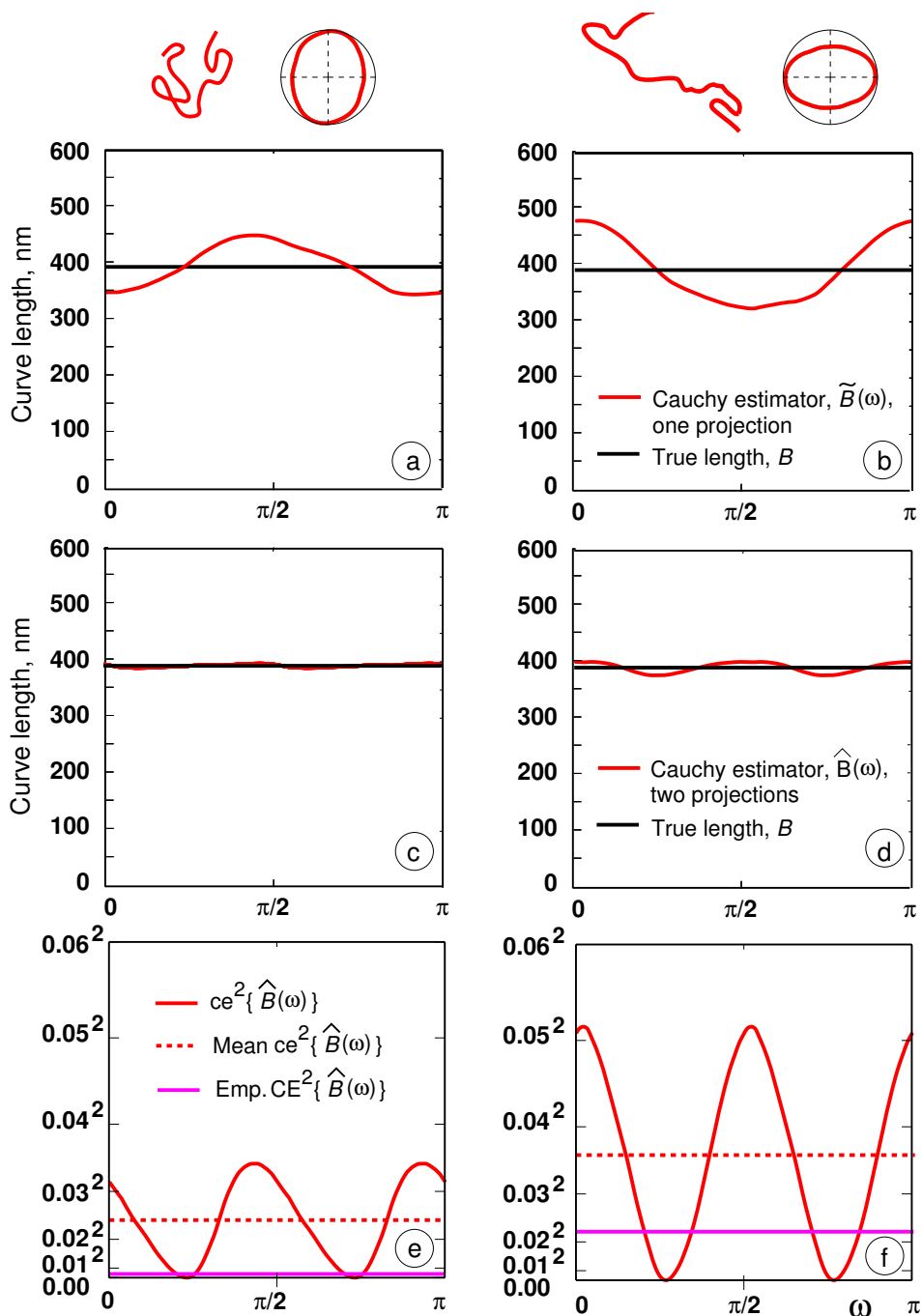
Fig. 3. *Top: curves numbered 5 (with corresponding results (a, c, e)) and 6 (with corresponding results (b, d, f)), from Fig. 1c, respectively, with their corresponding roses of total projected length. (a) (b) Cauchy curve length estimates from a single total projected length (red curves) as a function of the orientation of the projection axis. (c), (d) Idem for the two projection Cauchy estimates. Note the drastic increase in precision with respect to the single projection estimates. (e, f) The red wavy curves represent the model based predictors of the square coefficient of error of the two projection Cauchy estimator (Eq. 11).*

As anticipated in Fig. 3 for the Cauchy estimator, the predictor of the Cauchy (orientations) component shows a relatively poor accuracy, due in part to the fact that the true value of such component was very small relative to the total error variance. Thus, the total error variance practically coincides with the Cavalieri component in either case. Conditional on a given orientation ω, the measurement function $I_{Y'_\omega}(z)$, see Eq. 21, exhibits jumps because it is an intersection count, namely an integer. For this reason the smoothness constant used to obtain Eq. 17b was $q = 0$. Theory establishes that, under certain conditions, the trend or 'extension term' of the variance of a Cavalieri estimator is of order $O(T^{2q+2})$, (Kiêu *et al.*, 1999; García-Fiñana and Cruz-Orive, 2004). In the present study the expected trend of $\mathrm{CE}^2\{\widehat{B}(\omega,x)\}$ should therefore be approximately of $O(\mathbb{E}(I)^{-(2q+2)}) = O(\mathbb{E}(I)^{-2})$. A linear regression of $\log \mathrm{CE}^2\{\widehat{B}(\omega,x)\}$ versus $\log \mathbb{E}(I)$ yielded slope values of -1.90 and -1.49 for curves 5 and 6 respectively. Direct non linear least squares fitting of a exponential curve yielded exponent values of -2.45 and -2.13 respectively. For the Cavalieri components (Fig. 4e, f) the corresponding exponent estimates were -1.93 and -2.14 with the linear log-log regression, and $-2.45, -2.22$ with the non linear (exponential) regression, respectively. As a reference, a straight line segment corresponding to an exponent of -2 is represented in Fig. 4c-f.

CASE OF A POPULATION OF CURVES

DATA

For the sake of illustration suppose that Fig. 1a is a UR quadrat from a large observation region of the plane containing the curves of interest. For instance, this quadrat could be one from an extensive grid of systematic quadrats encompassing the whole population. The curves analyzed from a properly sampled quadrat should be chosen according to some unbiased rule, such as the unbiased frame (also called the forbidden line) rule (Gundersen, 1977, see also Howard and Reed, 2005, or Baddeley and Jensen, 2005). An unbiased counting – or sampling – rule, combined with UR quadrat sampling, warrants a UR sample of items because all the items in the population will have identical a priori probabilities of being included in the sample. A guard area should also be uniquely defined for all quadrats, to warrant that any curve that is sampled can be entirely observed for measurement. This means that the effective quadrat

analysed (not shown in Fig. 1a) will generally be smaller than the raw photograph. Suppose that the 9 curves numbered in Fig. 1c constitute a proper UR sample from the relevant population. The curve next to curve number 1 is apparently the union of two curves that could not be separated, and it has been removed as convenient for the present illustration. Imagine a horizontal line sweeping the quadrat from top to bottom. The curves tagged $1, 2, \ldots, 9$ were the 1st, 2nd, ..., 9th met by such line. (Actually, the sweeping line rule – see for instance Howard and Reed, 2005, Fig. 5.4 – is also unbiased for sampling, or counting, bounded objects in a bounded region).

Table 1. *Data corresponding to the curves labelled 1–9 in Fig. 1c. B: curve length computed via Eq. 3. $\widehat{B}$: curve length estimate, based on intersection counting, computed via Eq. 12a-c with a square grid of gap length $T = 25$ nm. $\mathrm{var}_w(\widehat{B}|B)$: error variance of $\widehat{B}$ computed by Eq. 20 via Eq. 16–19. $\mathrm{ce}(\widehat{B}|B) = 100 \cdot \{\mathrm{var}_w(\widehat{B}|B)\}^{1/2}/\widehat{B}$.*

Curve	B (nm)	$\widehat{B}$ (nm)	$\mathrm{var}_w(\widehat{B}|B)$	$\mathrm{ce}(\widehat{B}|B)$ (%)
1	404	438	2615	11.7
2	397	430	2156	10.8
3	385	399	1254	8.9
4	399	339	2558	14.9
5	393	333	1484	11.6
6	392	351	1532	11.2
7	427	402	1081	8.2
8	417	354	2114	13.0
9	389	385	951	8.0
Mean	400.3	381.2	1749.44	-
Var	187.75	1515.94	-	-

The data pertinent to the 9 tagged curves in Fig. 1c are collected in Table 1. The 'exact' length B of each curve was computed from Eq. 3. The Buffon-Steinhaus estimator $\widehat{B}(\omega,x)$ of B was computed in all cases from a single, computer generated IUR superimposition of the curve onto a fixed square grid of gap length $T = 25$ nm. This gap length was approximated from Eq. 30 to warrant $\mathbb{E}(I) \approx 20$ intersections, with B replaced with the sample mean of the 9 true lengths, namely 400.3 nm. Further, a single predictor $\mathrm{var}\{\widehat{B}(\omega,x)|B\}$ was computed for each curve via Eq. 20 using the automatically scored intersection counts.

ASSESSMENT OF THE MODEL ERROR VARIANCE PREDICTOR

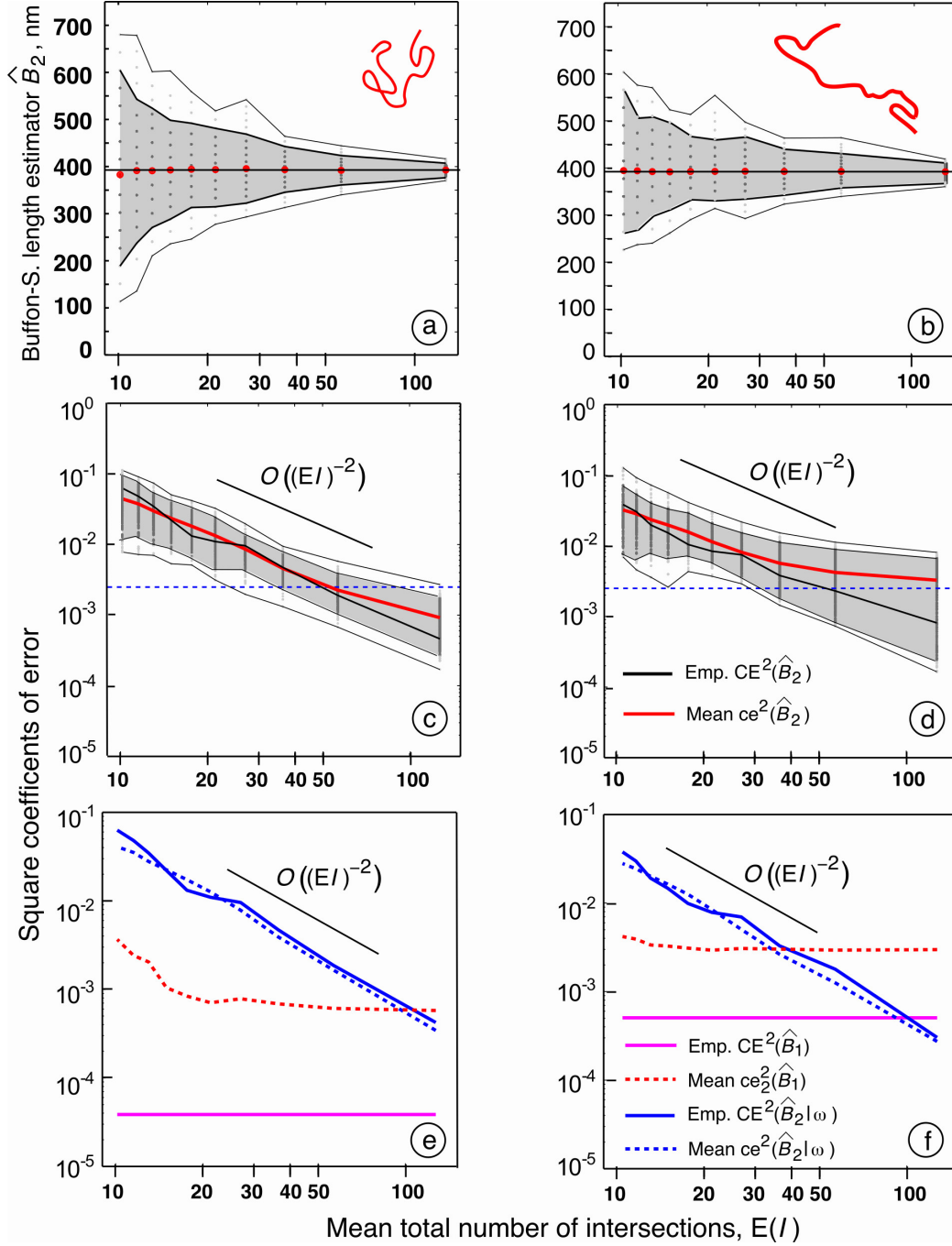The data in Table 1 allows an indirect assessment of the quality of the stereological error variance

Fig. 4. *(a), (b) Empirical Monte Carlo replicates of the intersection counting curve length estimates with* 10 *square grids of different sizes (ranging from* $T = 25$ *nm to the left, to* $T = 4$ *nm to the right end of the horizontal axis). A total of* 1024 *replications were generated in each case. The red dots represent the empirical means, which approximately coincide with the true curve length in all cases (as expected because the estimator is unbiased). The grey region is a* 95% *confidence band, that is, it encloses* 95% *of the replications; the outer line bounds enclose* 100% *of them. (c, d) The grey region is a* 95% *confidence band for the* 1024 *Monte Carlo replications of the model based predictor given by Eq. 20. The outer limits enclose all the replications. It is seen that the* 95% *confidence band always contains the empirical (i.e, the nearly true)* CE^2, *(black polygonal line) for the entire range considered, and for each of the two curves. (e, f) Empirical (continuous curves) and corresponding model based means (dotted) squared coefficient of error components. The Cauchy components are in pink and in red colour, respectively, whereas the Cavalieri components are in blue.*

predictor proposed in this paper (Eq. 20).

The variance $\mathrm{Var}_b(\widehat{B})$ of the unbiased length estimator $\widehat{B} = \widehat{B}(\omega, x)$ between curves (hence the subscript 'b') may be decomposed as follows,

$$\mathrm{Var}_b(\widehat{B}) = \mathrm{Var}_b(B) + \mathbb{E}_b \mathrm{Var}_w(\widehat{B}|B) , \qquad (34)$$

where $\mathrm{Var}_w(\widehat{B}|B)$ represents the stereological error variance within a curve (hence the subscript 'w'). The sample version of the preceding decomposition reads,

$$\mathrm{var}_b(\widehat{B}) = \mathrm{var}_b(B) + \mathrm{mean}_b\{\mathrm{var}_w(\widehat{B}|B)\} . \qquad (35)$$

From the data displayed in Table 1 we obtain $\mathrm{var}_b(\widehat{B}) = 1515.94$ and $\mathrm{var}_b(B) = 187.75$, whereby Eq. 35 yields the indirect, model free estimate,

$$\begin{aligned} \mathrm{mean}_b\{\mathrm{var}_w(\widehat{B}|B)\} &= 1515.94 - 187.75 \\ &= 1328.19 . \end{aligned} \qquad (36)$$

Note that the preceding estimate is model free because Eq. 34 is universally valid provided that $\widehat{B}$ is unbiased, namely that $\mathbb{E}(\widehat{B}|B) = B$. On the other hand, from the fourth column of Table 1 the direct sample mean of the individual model based variance predictors is,

$$\mathrm{mean}_b\{\mathrm{var}_w(\widehat{B}|B)\} = 1749.44 , \qquad (37)$$

which deviates from the model free estimate by a reasonable 32%.

ESTIMATION OF THE POPULATION MEAN CURVE LENGTH

Often, the target parameter is the mean length of a population of curves such as the DNA molecules in Fig. 1a. Let $\mathbb{E}(B)$ represent the population mean curve length. If the curves were digitized and measured automatically, then $\mathbb{E}(B)$ could be estimated by

$$\overline{B} = 400.3 \text{ nm} ,$$
$$\mathrm{ce}(\overline{B})\% = 100 \cdot \frac{\sqrt{187.75/9}}{400.3} = 1.14\% . \qquad (38)$$

On the other hand, if curve length was estimated from about 20 intersections each, then $\mathbb{E}(B)$ would be estimated by

$$\overline{\overline{B}} = 381.2 \text{ nm} ,$$
$$\mathrm{ce}(\overline{\overline{B}})\% = 100 \cdot \frac{\sqrt{1515.94/9}}{381.2} = 3.40\% . \qquad (39)$$

The preceding result would be similar if the intersections were counted directly by hand on the original images, with no image processing. Apart from time and effort, biases arising in the digitization procedure would be avoided.

Note that the within curve error variance estimates $\mathrm{var}_w(\widehat{B}|B)$ do not enter in the preceding calculations. Their knowledge, however, may be used to design the sampling protocol. The required coefficient of error of the mean depends on the purposes of the experiment – see Section 8.4 *How many animals?* from Cruz-Orive *et al.* (2004) for details, (here the latter title should better read *How many curves?*). For the present purposes we simplify the situation as follows. Suppose that we want to count a total of 400 intersections. If we count 20 intersections in each of 20 curves, then we might expect,

$$\mathrm{ce}(\overline{B})\% = 100 \cdot \frac{\sqrt{1515.94/20}}{381.2} = 2.28\% . \qquad (40)$$

On the other hand, if we rather want to count 40 intersections in each of 10 curves, then $\mathrm{Var}_w(\widehat{B}|B)$ may be expected to be reduced by a factor of about 4, because the latter variance is roughly proportional to $(\mathbb{E}I)^{-2}$, (Fig. 4), and we are doubling the number of intersections per curve. Thus, in this case,

$$\mathrm{mean}_b\{\mathrm{var}_w(\widehat{B}|B)\} = 1328.19/4 = 332.05 , \qquad (41a)$$
$$\mathrm{var}_b(\widehat{B}) = 187.75 + 332.05 = 519.80 \qquad (41b)$$
$$\mathrm{ce}(\overline{\overline{B}})\% = 100 \cdot \frac{\sqrt{519.80/10}}{381.2} = 1.89\% . \qquad (41c)$$

The predicted precision of the mean is therefore of about 2% in either case. The former option (*i.e.*, working less in more curves) should generally be the better one (Gundersen and Østerby, 1981). First, counting too many intersections per curve is tedious. Second, the foregoing formulae for the coefficient of error of the mean require independence between curve estimates. Sampling a few curves in a relatively small region is likely to violate this condition.

It is worth mentioning that, if the target parameter is $\mathbb{E}(B)$, then it is not necessary to measure individual curve lengths, as above. A more convenient (and probably more efficient) design consists in superimposing – on the (extensive) field of interest, see for instance Fig. 1 of Podestà *et al.* (2005) – systematic quadrats combined with test lines. The upper edge of each quadrat, for instance, could be adopted as a test line, although the test lines may be placed outside the quadrats. The superimpositions should be IUR relative to the curves, or just UR if the test lines consisted of half circles. Let l/a cm^{-1} denote the known test line length per quadrat area. Further, let Q_i denote the number of individual curves counted in the ith quadrat according to the unbiased frame rule, say. Also, let I_i denote the corresponding number of intersections scored with test lines associated with the ith quadrat,

and let n represent the total number of quadrats. Then, a ratio unbiased estimator of $\mathbb{E}(B)$ is,

$$\overline{B} = \frac{\pi}{2} \cdot \frac{a}{l} \cdot \sum_{i=1}^{n} I_i \bigg/ \sum_{i=1}^{n} Q_i \text{ cm}. \qquad (42)$$

In this case $\text{Var}(\overline{B})$ may be predicted by Cochran's formula (Cochran, 1977; Howard and Reed, 2005, p.158), provided that the distance between quadrats is large enough to assume independence between the n data pairs $\{(I_i, Q_i)\}$. Similar designs are illustrated in Chapter 12 of the latter book. Note that: (a) If a quadrat does not lie entirely inside the reference area, then we count curve(s) just in the available quadrat area anyway, always respecting the unbiased rule adopted. (b) The intersections are counted on any curve that is intersected by any test line, whether it was counted in the quadrats, or not. In other words, there will generally be counted curves that do not contribute any intersections, and curves that do contribute intersections but are not counted.

DISCUSSION

The main purpose of this paper was to check the accuracy of a previously published error variance prediction formula for curve length estimation by intersection counting with a square grid. Each of nine finite linear features (namely flattened DNA molecule images, see Fig. 1a) was repeatedly superimposed with isotropic uniform randomness onto fixed square grids of various sizes. The subsequent replicates of the variance predictors were thereby compared against the empirical (*i.e.*, close to true) error variance. The overall performance of the prediction formula was satisfactory – the results corresponding to two of the curves (namely a fairly isotropic and a fairly anisotropic one) are displayed in Fig. 4c, d.

The error variance has a rotation (Cauchy) and a translation (Cavalieri) component. The prediction of the former (Eq. 11) was relatively poor, whereas the one of the latter was satisfactory (Fig. 4e, f). Because the Cauchy contribution was relatively very low, however, the overall prediction was satisfactory.

It may be didactic to consider the extreme case in which the target curve is a straight line segment of length B. Here the Cauchy estimator based on two mutually orthogonal projections (Eq. 10) reads,

$$\widehat{B}(\omega) = \frac{\pi}{4} B \cdot (\cos \omega + \sin \omega),$$
$$\mathbb{P}(\mathrm{d}\omega) = \frac{2}{\pi} \mathrm{d}\omega, \quad (0 \le \omega < \pi/2), \qquad (43)$$

and its exact square coefficient of error is,

$$\text{CE}^2\{\widehat{B}(\omega)\} = \frac{\pi^2 + 2\pi}{16} - 1, \qquad (44)$$

so that $\text{CE}\{\widehat{B}(\omega)\}\% \approx 9.77\%$, which is non negligible. As soon as the curve is not too anisotropic, however, the preceding CE decreases rapidly. On the other hand, for a straight line segment the Cauchy error variance predictor given by Eq. 11 reads

$$\text{var}\{\widehat{B}(\omega)\} = \frac{\pi^2}{240} B^2 \cdot (\cos \omega - \sin \omega)^2, \qquad (45)$$

and its relative bias is,

$$\frac{\mathbb{E}_\omega \text{var}\{\widehat{B}(\omega)\}}{\text{Var}\{\widehat{B}(\omega)\}} - 1 = \frac{1}{15} \cdot \frac{\pi^2 - 2\pi}{\pi^2 + 2\pi - 16} - 1 \qquad (46)$$
$$\approx 0.56,$$

namely 56%. It should be borne in mind, however, that $\widehat{B}(\omega)$ is based on only two observations.

The Monte Carlo experiment involved 1024 automatic IUR superimpositions of each curve onto each fixed grid. To do this each curve was approximated as described in Section *Processing of the curve images* by a polygonal curve with known vertex coordinates. The procedure involved the successive use of four software packages (with some manual editing at the later stages) and took about 10–15 min per curve on average. Moreover, the manipulations will introduce some bias in the curve length. On the contrary, intersection counting estimation can be implemented directly on each original curve in less than 20s, and the only bias present may be due to the possible difference between the length of a 'free' DNA molecule, and its flattened projection onto the observation plane. It is in fact pausible that all the DNA molecules considered here had the same length, and that the different lengths reported in the first column of Table 1 were due to the sum of two biases, namely the flattening bias plus the image processing one. This is impossible to ascertain: the best way to handle bias is to avoid it. Note also that the curves themselves might not be separable automatically – see for instance the curve lying between curves 1 and 2 in Fig. 1c. A human, however, will soon realize that this curve must be the union of two. Thus, apart from the possibility of performing automatic experiments to investigate higher order properties, as illustrated in this study, automatic image processing cannot be recommended if the only purpose is to estimate mean curve length in a not too extensive case study.

The present study is relatively academic, in the sense that it is restricted to a single bounded curve. Our main concern was the prediction of the within curve

error variance $\mathrm{Var}_w(\widehat{B}|B)$. Often, however, the target quantity may be the population mean curve length (Subsection *Estimation of the population mean curve length*).

On the other hand, only the square grid was considered here. In practice, grids consisting of separate straight line segments, half circles, etc., can be more efficient than the square grid (which often tends to yield too many intersections). In addition, test systems consisting of half circles, such as the Merz grid (Weibel, 1979; Howard and Reed, 2005) may be convenient because they do not require a random orientation relative to the curve. Unfortunately, to our knowledge no general error variance prediction formula exists for a general grid. It is worth mentioning, however, that the 'fakir predictor' proposed for a test grid of wavy cycloids in Cruz-Orive *et al.* (2014, Eq. 42) proved to be satisfactory for digitized brain sections (Fig. 5 of that study). The formula could be tried also for separate, systematically arranged test segments, half circles, or more general bounded test curves. The formula requires recording the intersection counts separately for each fundamental test curve, and arranging them into a rectangular matrix (p. 132 of the latter paper). The problem, however, remains basically open.

In short, the intersection counting method is highly recommended to estimate curve length because it is direct (*i.e.*, it does not require image processing as long as the intersection points can be scored unambiguously in the original images), design unbiased, and rather efficient. Also, if a square grid is used, then this study shows that the error variance may be predicted fairly reliably from a single grid superimposition on the target curves considered. Semiautomatic stereological devices (several brands of which are available on the market) may be of considerable help here because they will generate the desired test grid automatically on the computer screen, with IUR position relative to the target curve images. The intersection counts are then recorded manually, and may be exported for instance into the free software R (http://www.r-project.org/) to compute the estimates.

ACKNOWLEDGEMENTS

REFERENCES

Baddeley AJ, Jensen EBV (2005). Stereology for statisticians. London: Chapman & Hall/CRC.

Cochran W (1977). Sampling Techniques. 3rd Ed. New York: J. Wiley & Sons.

Cruz-Orive LM (1989). Precision of systematic sampling on a step function. In: Hubler A, Nagel W, Ripley B, Werner G, eds., Geobild'89, pp. 185–193. Berlin: Akademie Verlag.

Cruz-Orive LM, Gelšvartas J, Roberts N (2014). Sampling theory and automated simulations for vertical sections, applied to human brain. J Microsc 253:119–50.

Cruz-Orive LM, Gual-Arnau X (2002). Precision of circular systematic sampling. J Microsc 207:225–42.

Cruz-Orive LM, Insausti A, Insausti R, Crespo D (2004). A case study from Neuroscience involving Stereology and Multivariate Analysis. In: Evans S, Janson A, Nyengaard J, eds., Quantitative Methods in Neuroscience, chap. 2. Oxford: Practical Neuroscience Series, Oxford University Press, 16–64.

García-Fiñana M, Cruz-Orive LM (2004). Improved variance prediction for systematic sampling on R. Statistics 38:243–72.

Gundersen HJG (1977). Notes on the estimation of the numerical density of arbitrary profiles: the edge effect. J Microsc 111:219–23.

Gundersen HJG, Østerby R (1981). Optimizing sampling efficiency of stereological studies in biology: or 'do more less well!'. J Microsc 121:65–73.

Howard V, Reed MG (2005). Unbiased stereology: Three-dimensional measurement in microscopy, 2nd Ed. Oxford: Bios/ Taylor & Francis.

Kiêu K, Souchet S, Istas J (1999). Precision of systematic sampling and transitive methods. J Stat Plan Inference 77:263–79.

Miles R (1974). On the elimination of edge effects in planar sampling. In: Kendal E, Harding D, eds., Stochastic Geometry: A tribute to the memory of Rollo Davidson. London: John Wiley and Sons, 228–47.

Miles R, Davy P (1976). Precise and general conditions for the validity of a comprehensive set of stereological fundamental formulae. J Microsc 107:211–26.

Moran PAP (1966). Measuring the length of a curve. Biometrika 53:359–64.

Podestà A, Imperadori L, Colnaghi W, Finzi L, Milani P, Dunlap D (2004). Atomic force microscopy study of DNA deposited on poly l-ornithine-coated mica. J Microsc 215:236–40.

Podestà A, Indrieri M, Brogioli D, Manning GS, Milani P, Guerra R, Finzi L, Dunlap D (2005). Positively charged surfaces increase the flexibility of DNA. Biophys J

89:2558–63.

Santaló L (1976). Integral geometry and geometric probability. Reading, MA: Addison-Wesley.

Steinhaus H (1930). Zur Praxis der Rektifikation und zum Längenbegriff. Berichte Sachsischen Akad Wiss Leipzig 82:120–30.

Weibel ER (1979). Stereological methods. Practical methods for biological morphometry, vol. 1. London: Academic Press.