ON THE PRECISION OF THE NUCLEATOR

JAVIER GONZÁLEZ-VILLA✉, MARCOS CRUZ AND LUIS M. CRUZ-ORIVE

Department of Mathematics, Statistics and Computation, Faculty of Sciences, University of Cantabria, Avda.
Los Castros 48, E-39005. Santander, Spain.
e-mail: javier.gonzalezv@alumnos.unican.es, marcos.cruz@unican.es, luis.cruz@unican.es

ABSTRACT

The nucleator is a design unbiased method of local stereology for estimating the volume of a bounded object. The only information required lies in the intersection of the object with an isotropic random ray emanating from a fixed point (called the pivotal point) associated with the object. For instance, the volume of a neuron can be estimated from a random ray emanating from its nucleolus. The nucleator is extensively used in biosciences because it is efficient and easy to apply. The estimator variance can be reduced by increasing the number of rays. In an earlier paper a systematic sampling design was proposed, and theoretical variance predictors were derived, for the corresponding volume estimator. Being the only variance predictors hitherto available for the nucleator, our basic goal was to check their statistical performance by means of Monte Carlo resampling on computer reconstructions of real objects. As a plus, the empirical distribution of the volume estimator revealed statistical properties of practical relevance.

Keywords: Monte Carlo resampling, nucleator, rat brain, stereology, variance prediction.

INTRODUCTION

The nucleator (Gundersen, 1988) is a geometric sampling method to estimate the volume of a particle. In stereology the term 'particle' refers to a compact subset of three dimensional Euclidean space with piecewise smooth boundary. The particle should bear an observable associated point, called the pivotal point, from which one or more isotropically oriented rays can be generated. Details are given in *Nucleator representation of volume* section.

Because the sampled rays emanate from a common pivotal point, the nucleator is a method of local stereology (Jensen, 1998). The direction of an isotropic random (IR) ray corresponds to a uniform random point on the unit sphere. If the ray is chosen directly in space, then we have the 3D nucleator. In practice, the nucleator is implemented in two stages: first, a pivotal plane (namely a IR plane through the pivotal point) is generated, and then a IR ray is generated in the pivotal plane. The latter version (Gundersen, 1988), generally called the IR nucleator, was called the pivotal nucleator by Cruz-Orive and Gual-Arnau (2015) to distinguish it from the 3D nucleator. Another two stage nucleator is the vertical nucleator: first, a local vertical plane (namely a plane containing an arbitrary fixed vertical axis through the pivotal point), is generated, and then a sine weighted ray is sampled in this plane (Tandrup, 1993; Jensen, 1998; Howard and Reed, 2005; West, 2012). The three nucleator versions are unbiased for particle volume, but they are different estimators, and

will generally have different variances for a given number of rays.

Gual-Arnau and Cruz-Orive (2002) proposed a systematic design based on coaxial vertical sections a fixed angle apart (see *Estimation of volume with the nucleator* section), and obtained theoretical variance prediction formulae for the corresponding volume estimator (Section *Nucleator error variance predictors considered here*). The purpose of this paper is to check the performance of these variance prediction formulae by means of Monte Carlo replications of vertical systematic nucleator samples (Section *Assessment of the nucleator variance predictors by Monte Carlo resampling*) on digitized versions of various test objects (Fig. 1 and Section *Material and its processing*). As a plus, the empirical distribution of a popular version of the nucleator based on two opposite rays exhibited statistical properties of practical interest (Subsection *Distribution of the nucleator estimators*). The empirical variances computed on the different objects for different sample sizes, and the performance of the variance predictors, are displayed in Subsection *Precision of the nucleator estimators* and Subsection *Performance of the variance predictors*, respectively. The conclusions pertaining to the performance of the variance predictors are mainly negative (Section *Discussion and conclusions*).

While the nucleator estimator is unbiased for the volume of an object irrespective of the orientation of the vertical axis relative to the object, for the proposed design the distribution of the estimator generally

depends on that orientation. For this reason, two mutually perpendicular orientations were considered for one of the objects (Fig. 1b, c).
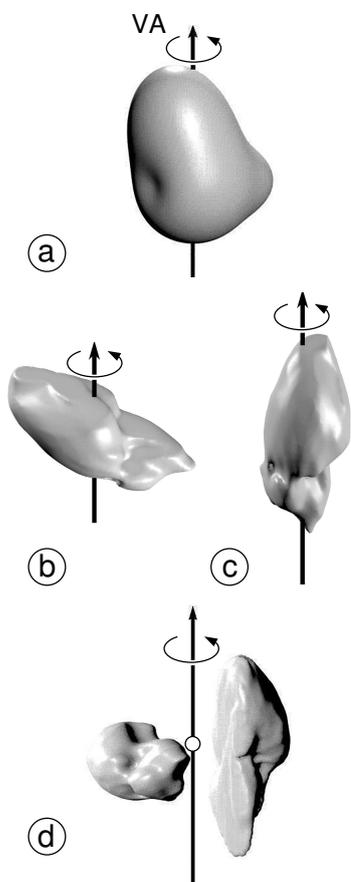
Fig. 1. *Test objects studied in this paper, see Material and its processing section. (a) Synthetic particle whose radius vector has a known explicit expression. VA: vertical or polar axis. (b, c) Reconstruction of the right hemisphere (named H_1 in the tables) of a rat brain, with approximately transversal (b) and longitudinal (c) vertical axis through the centre of mass of the object. (d) Reconstructions of the right (H_1) and left (H_2) hemispheres of the same rat brain, placed in antithetic position. The white dot represents the centre of mass of the union $H_1 \cup H_2$ of both hemispheres.*

NUCLEATOR REPRESENTATION OF VOLUME

Consider a particle $Y \subset \mathbb{R}^3$ of volume V, which is the parameter of interest. Let O represent a fixed pivotal point associated with Y.

The orientation of a ray emanating from O is determined by a point u on the unit sphere $\mathbb{S}^2$ centred at O. Henceforth u will indistinctly refer to a point

on $\mathbb{S}^2$, or to a spatial orientation. If a ray emanating from O hits Y, then the corresponding intersection will generally consist of say $m(u)$ separate intercept segments. The distances of the end points of these intercepts from O, arranged in increasing order of magnitude, may be denoted as follows,

$$\{l_{i-}(u), l_{i+}(u); i = 1, 2, ..., m(u)\} . \tag{1}$$

Thus, $(l_{i-}(u), l_{i+}(u))$ denote respectively the closer and the farther distances of the end points of the i-th intercept segment from O. Integration of the conic volume element associated with a ray (see, *e.g.* Cruz-Orive, 1987, Appendix B) leads to the general nucleator representation of volume, namely,

$$V = \int_{\mathbb{S}^2} f(u) \, du , \tag{2}$$

where

$$f(u) = \frac{1}{3} \sum_{i=1}^{m(u)} \left(l_{i+}^3(u) - l_{i-}^3(u) \right) \tag{3}$$

is the measurement function. Note that, if $O \in Y$, then $l_{1-}(u) = 0$ for all u. Further, if Y is star shaped with respect to $O \in Y$, then $m(u) = 1$, and

$$V = \frac{1}{3} \int_{\mathbb{S}^2} l^3(u) \, du , \tag{4}$$

where $l(u) \equiv l_{1+}(u)$ is the length of the radius vector of Y.

ESTIMATION OF VOLUME WITH THE NUCLEATOR

An unbiased estimator (UE) of V may be easily constructed by observing f at an isotropic orientation $u \in \mathbb{S}^2$. Isotropy means that u is a uniform random (UR) variable on $\mathbb{S}^2$, whereby its probability element is proportional to the area element du on the unit sphere, namely,

$$\mathbb{P}(du) = \frac{du}{4\pi}, \quad u \in \mathbb{S}^2 . \tag{5}$$

From Eq. 2 it follows that

$$\widehat{V} = 4\pi f(u) \tag{6}$$

is a UE of V from a single isotropic ray.

To describe a sampling design involving more than one ray it is convenient to parametrize u by its spherical polar coordinates, namely the longitude angle ϕ, and the colatitude angle θ, see Fig. 2a, so that $du = \sin\theta d\phi d\theta$. The origin of θ is the polar axis Ox_3, which is regarded as the vertical axis. The origin O is the pivotal point. Now Eq. 2 may be written as follows,

$$V = \int_0^{2\pi} d\phi \int_0^{\pi} f(\phi, \theta) \sin\theta \, d\theta . \tag{7}$$
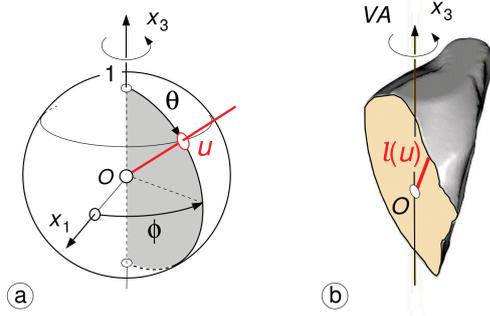
Fig. 2. *Basic definitions for the nucleator design. (a) A direction u, or equivalently a point u on the unit sphere, of spherical polar coordinates* (ϕ, θ) *with respect to an orthogonal frame* $Ox_1x_2x_3$*, where* Ox_3 *is the polar vertical axis (VA). (b) A local vertical section of the object shown in Fig. 1b, through a fixed pivotal point O which is the centroid of the object. Within this section a nucleator ray of length* $l(u)$ *has been drawn along a direction u.*

In Gual-Arnau and Cruz-Orive (2000) it was proposed to estimate V with a systematic design in which ϕ would be sampled at an integer number $n_1 \geq 1$ of systematic UR points in $[0, 2\pi)$, whereas $y = \cos\theta$ would be sampled at an integer number $n_2 \geq 1$ of systematic UR points in $[-1, 1)$. The fact that $f(\phi, \theta)$ is periodic modulo 2π in ϕ, but not periodic modulo π in θ, however, makes it difficult to represent the variance of the estimator of V. For this reason it was proposed to replace the measurement function $f(u)$ with $F(u) = \frac{1}{2}[f(u) + f(-u)]$, namely,

$$F(\phi, \theta) = \frac{1}{2}[f(\phi, \theta) + f(\phi + \pi, \pi - \theta)] , \quad (8)$$

because $-u = (\phi, \theta + \pi) = (\phi + \pi, \pi - \theta)$. It is easy to verify that $F(\phi, \theta)$ is periodic in both ϕ and θ, namely,

$$F(\phi, \theta) = F(\phi + 2k\pi, \theta + l\pi), \quad k, l \in \mathbb{Z} , \quad (9)$$

which, as shown in Gual-Arnau and Cruz-Orive (2002), renders the mathematical problem tractable. Now the target parameter may be written,

$$V = \int_0^{2\pi} d\phi \int_0^{\pi} F(\phi, \theta) \sin\theta \, d\theta$$
$$= \int_0^{2\pi} d\phi \int_{-1}^{1} F(\phi, \cos^{-1} y) \, dy. \quad (10)$$

The systematic sampling design consists in observing F at $n_1 n_2$ points

$$\{F_{ij} \equiv F(\phi_i, \theta_j), i = 0, 1, ..., n_1 - 1, j = 0, 1, ..., n_2 - 1\} \quad (11)$$

where

$$\phi_i = (U_1 + i)T_1, \ T_1 = 2\pi/n_1 ,$$
$$\theta_j = \cos^{-1}[1 - (U_2 + j)T_2], \ T_2 = 2/n_2 , \quad (12)$$

and (U_1, U_2) is a pair of independent UR numbers in the interval $[0, 1)$. The preceding procedure uses the fact that ϕ_0 and $\cos\theta_0$ are independent UR variables in the intervals $[0, T_1)$ and $[1 - T_2, 1)$, respectively, where T_1, T_2 are the corresponding sampling periods. The resulting UE of V is,

$$\widehat{V}_{n_1, n_2} = T_1 T_2 \sum_{i=0}^{n_1-1} \sum_{j=0}^{n_2-1} F_{ij} , \quad (13)$$

which is random via the UR pair (U_1, U_2).

The following two special cases are of practical interest because the data can be collected on two axes through a single vertical section containing the vertical axis Ox_3 through the pivotal point O, namely,

$$\widehat{V}_{12} = 2\pi[F(\phi_0, \theta_0) + F(\phi_0, \theta_1)],$$
$$\phi_0 = 2\pi U_1, \theta_j = \cos^{-1}(U_2 - j) , \quad j = 0, 1 . \quad (14)$$

see Fig. 3a, and

$$\widehat{V}_{21} = 2\pi[F(\phi_0, \theta_0) + F(\phi_1, \theta_0)],$$
$$\phi_i = \pi(U_1 + i), \theta_0 = \cos^{-1}(1 - 2U_2) , \quad i = 0, 1 , \quad (15)$$

see Fig. 3b. Out of the preceding two nucleator estimators, the more usual one is $\widehat{V}_{12}$, see for instance Tandrup (1993).
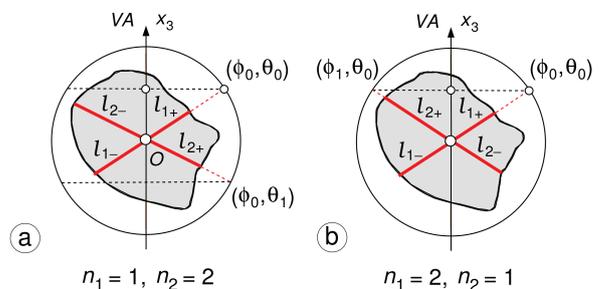
Fig. 3. *Special cases of the vertical nucleator design on a single vertical section through the pivotal point O. The design in (a) corresponds to the estimator* $\widehat{V}_{12}$*, the one in (b) to* $\widehat{V}_{21}$*.*

NUCLEATOR ERROR VARIANCE PREDICTORS CONSIDERED HERE

Our purpose is to explore the performance of the predictor $\text{var}_{m_1,m_2}(\widehat{V}_{n_1,n_2})$ of the error variance $\text{var}(\widehat{V}_{n_1,n_2})$ of the nucleator volume estimator $\widehat{V}_{n_1,n_2}$, see Eq. 13, proposed in Gual-Arnau and Cruz-Orive (2002). The predictor depends on the value of two constants m_1, m_2 entering in the underlying covariogram model. The latter was the product of two polynomials of degrees $2m_1 + 1$ and $2m_2 + 1$ corresponding to the individual covariogram models for ϕ and θ, respectively. Here we are considering only two cases: the choice $m_1 = m_2 = 1$ should in principle be more adequate than $m_1 = m_2 = 0$ for smoother, relatively more regular objects. It should be stressed that the predictors are theoretical approximations, and not unbiased estimators of $\text{var}(\widehat{V}_{n_1,n_2})$, in general. Their performance depends on the shape of the object Y of interest.

For $n_1, n_2 \geq 2$ the predictors considered here have the following expressions,

$$
\begin{aligned}
\text{var}_{00}(\widehat{V}_{n_1,n_2}) = & \frac{4\pi^2}{9n_1n_2(n_1-1)(n_2-1)} \\
& \cdot \left[6(n_1-1)(C_{00}-C_{01}) \right. \\
& + 6(n_2-1)(C_{00}-C_{10}) \\
& - (n_1^2 + n_2^2 - 1) \\
& \left. \cdot (C_{00}-C_{01}-C_{10}+C_{11}) \right] ,
\end{aligned}
\tag{16}
$$

$$
\begin{aligned}
\text{var}_{11}(\widehat{V}_{n_1,n_2}) = & \frac{4\pi^2}{225n_1n_2(n_1-1)^2(n_2-1)^2} \\
& \cdot \left[30(n_1-1)^2(C_{00}-C_{01}) \right. \\
& + 30(n_2-1)^2(C_{00}-C_{10}) \\
& - (n_1^4 + n_2^4 - 1) \\
& \left. \cdot (C_{00}-C_{01}-C_{10}+C_{11}) \right] ,
\end{aligned}
\tag{17}
$$

where

$$
\begin{aligned}
C_{kl} = & \sum_{i=0}^{n_1-1} \sum_{j=0}^{n_2-1} F_{ij}F_{i+k,j+l} , \\
& k = 0,1,...,[n_1/2] , \quad l = 0,1,...,[n_2/2] , \\
F_{i+n_1,j} = & F_{i,j+n_2} = F_{i+n_1,j+n_2} = F_{ij} .
\end{aligned}
\tag{18}
$$

For the special case $(n_1 = 1, n_2 \geq 2)$ the following formulae are used,

$$
\begin{aligned}
\text{var}_{00}(\widehat{V}_{1,n_2}) = & \frac{8\pi^2}{3} \cdot \frac{C_{00}-C_{01}}{n_2(n_2-1)} , \\
\text{var}_{11}(\widehat{V}_{1,n_2}) = & \frac{8\pi^2}{15} \cdot \frac{C_{00}-C_{01}}{n_2(n_2-1)^2} .
\end{aligned}
\tag{19}
$$

For $(n_1 \geq 2, n_2 = 1)$, the corresponding formal expressions are obtained by replacing C_{01} with C_{10} in the preceding ones. In particular,

$$
\begin{aligned}
\text{var}_{00}(\widehat{V}_{12}) = & \frac{4\pi^2}{3}(F_{00}-F_{01})^2 , \\
\text{var}_{11}(\widehat{V}_{12}) = & \frac{4\pi^2}{15}(F_{00}-F_{01})^2 ,
\end{aligned}
\tag{20}
$$

and for $\widehat{V}_{21}$ it suffices to replace F_{01} with F_{10} in the preceding formulae.

MATERIAL AND ITS PROCESSING

Here we describe the imaging procedures necessary to perform automatic Monte Carlo resampling of the nucleator in each of the following three dimensional objects, (see Fig. 1).

(1) *Synthetic particle model (Fig. 1a) studied in Section 6 from Gual-Arnau and Cruz-Orive (2002).* The radius vector of this particle had a known mathematical expression consisting of a linear combination of three non random exponential terms of Gaussian type (Foley *et al.*, 1990). The position of the vertical axis (VA) was kept unchanged. In the former paper the volume of the particle (namely 6.919) and the nucleator replications, were obtained with *Mathematica*® using exact expressions. Here, however, we have applied the same reconstruction techniques and the same numerical resampling procedures to all the objects. The previously published results therefore enabled us to check the correctness of our software. For instance, the approximate volume was now found to be 6.914, only 0.07% less than the more exact one.

(2) *Rat brain no.5 illustrated in Fig.6 from Cruz-Orive et al. (2010).* The input consisted of the projections of an exhaustive series of 19 nonempty, 1.2 mm thick serial slabs, obtained on a 1.5 T MRI scanner (Gyroscan Intera, Philips). The right hemisphere, consisting of 14 serial slabs, was studied in detail for each of two different orientations of the VA, see Fig. 1b,c. The left hemisphere consisted of 9 slabs, see Table 1 from the latter paper. Finally, the union of both hemispheres was also studied in the antithetic arrangement suggested in Fig.2 of the latter paper, see Fig. 1d here.

For convenience both brain hemispheres were processed simultaneously because they were scanned together. The reconstructions were separated at the end. The relevant steps may be summarized as follows.

Step 1. The consecutive slab projections were regarded as parallel thin sections 1.2 mm apart. The field of view of each section was of 190×190 pixels, and the side length of each pixel represented $2/7 \approx 0.29$ mm. The gray scale resolution was of 8bpp (bits per pixel) where the extreme values 0 and 255 corresponded to black and white pixels, respectively. Initially each section image was thereby represented by a 190 x 190 matrix of grey scale values. Using 'PIL' (Python Imaging Library) the latter values were filtered to retain values less than 150 – this thresholding led to a black-and-white image which was judged to be satisfactory to represent the type of sections analyzed here.

Step 2. Within Blender V 2.68a (www.blender.org), each pixel of each section was replaced with a point, whereby each section was represented by one or more clouds of points in the same plane. To each point belonging to the section, a vector (x, y, z) was assigned its Cartesian coordinates. The Oz axis was normal to the sections. If a point did not belong to the section, the corresponding coordinate vector was replaced with '0'. Then a script was written to remove the interior points of each cloud, leaving only the points at the periphery of the cloud. The software did this automatically, first scanning the rows, and then the columns of the matrix.

Step 3. The surviving points at the periphery of each section were interpolated by means of Bezier curves using the 'Auto Handle' facility from Blender. The curves were automatically fit according to an optimality smoothing criterion.

Step 4. The initial set of sections thereby became a set of parallel smooth Bezier contours in serial planes 1.2 mm apart. Moreover the number of control points of each contour was forced to be the same for all the contours. In this way it was possible to apply a 'knitting' or 'needlework' algorithm (Klein *et al.*, 1999) whereby the closest nodes from each pair of consecutive contours were automatically linked one-to-one, to form a closed ribbon of triangles between the corresponding planes. The result was a closed three dimensional triangular mesh constituting a primary reconstruction of the boundary of the object.

Step 5. From the primary triangulation, the enclosed object volume was computed with the command 'calc_volume()' from the 'bmesh' Blender library. Further, the command 'origin_set(type='ORIGIN_CENTER_OF_MASS')' from the 'bpy' Blender library computed the centre of mass of the object and adopted it as the new origin, which played the role of the pivotal point of the object from which the nucleator rays were generated. Thereafter 'Auto Handle' refined and optimized the

triangulation automatically, respecting the previously computed volume. The secondary triangulation of the complete rat brain consisted of 21764 triangles, whose areas lied in the interval $[4.43 \cdot 10^{-5}, 1.80 \cdot 10^{-3}]$ cm^2. The total surface area was 16.022 cm^2, split into 7.682 cm^2 for the right hemisphere (Fig. 1b,c) and 8.340 cm^2 for the left one. The corresponding volumes were 2.856, 1.419 and 1.437 cm^3 respectively.

Step 6. The vertices and edges of the secondary triangulation were further recalculated to obtain a smoother version of the object boundary (Fig. 2) by reducing the number of triangles and optimizing their shape. This was done automatically with the aid of the 'Remesh Dual Contour Algorithm' provided by Blender (Ju *et al.*, 2002). Here the following settings were used: Octree Depth = 7; Scale = 0.99; Sharpness = 1.000; Threshold = 1.000; Smooth Shading = Enabled. This step had no influence on the volume and the position of the centroid of the object.

ASSESSMENT OF THE NUCLEATOR VARIANCE PREDICTORS BY MONTE CARLO RESAMPLING

For each object and each pair (n_1, n_2) of sample sizes, the volume estimator $\widehat{V}_{n_1,n_2}$, and its variance predictor $\text{var}_{m,m}(\widehat{V}_{n_1,n_2})$, $m = 0, 1$, were replicated a number N of times according to a systematic protocol. The rectangular tile $[0, T_1) \times [1 - T_2, 1)$ was partitioned into $N = N_1 N_2$ rectangular subtiles of horizontal and vertical side lengths $t_1 = T_1/N_1$ and $t_2 = T_2/N_2$, respectively, namely,

$$\{[kt_1, (k+1)t_1) \times [1 - (l+1)t_2, 1 - lt_2), \\ k = 0, 1, ..., N_1 - 1, l = 0, 1, ..., N_2 - 1\}. \tag{21}$$

For the first sample replication, the starting point $(\phi_0(0), \cos\theta_0(0))$ was UR in the first subtile $[0, t_1) \times [1 - t_2, 1)$, and so forth, see below.

Most of the computing time was taken by calculating the nucleator ray lengths. Thus, we chose $N_1 = 180/n_1, N_2 = 90/n_2$ to ensure that all the replications of any given sample involved the same number of rays, namely $N = 16200$. The sample sizes n_1, n_2 were chosen so that N_1, N_2 were natural numbers. The kl-th replication of $\widehat{V}_{n_1,n_2}$ was computed as follows,

$$\widehat{V}_{n_1,n_2}(k,l) = T_1 T_2 \sum_{i=0}^{n_1-1} \sum_{j=0}^{n_2-1} F_{ij}(k,l) ,$$

$$F_{ij}(k,l) = F(\phi_i(k), \theta_j(l)) ,$$

$$\phi_i(k) = \left(\frac{U_1+k}{N_1} + i \right) T_1 , \tag{22}$$

$$\theta_j(l) = \cos^{-1} \left[1 - \left(\frac{U_2+l}{N_2} + j \right) T_2 \right] ,$$

where U_1, U_2 are two independent UR numbers in the interval $[0,1)$. Thus, all the replications of a given sample depended on a single pair (U_1, U_2). The relevant ray lengths involved in the computation of $F_{ij}(k,l)$ via Eq. 8 were computed from the final triangulation of the object boundary using the 'intersect_ray_tri' facility available in the Blender library 'mathutils.geometry'.

The *kl*-th replication $\mathrm{var}_{m,m}(\widehat{V}_{n_1,n_2}(k,l)), m = 0,1$, was computed simultaneously by replacing each F_{ij} with $F_{ij}(k,l)$ in Eqs. 16-20. While the graphs display individual raw variance predictors, whether positive or negative, in practice one will take,

$$\mathrm{var}_{m,m}(\widehat{V}_{n_1,n_2}(k,l))_+ =$$
$$\max \left\{ \mathrm{var}_{m,m}(\widehat{V}_{n_1,n_2}(k,l)), 0 \right\} . \tag{23}$$

For the sake of convenience, renumber the double sequences $\left\{ \widehat{V}_{n_1,n_2}(k,l) \right\}$ and $\left\{ \mathrm{var}_{m,m}(\widehat{V}_{n_1,n_2}(k,l)) \right\}$ as $\left\{ \widehat{V}_{n_1,n_2}(i) \right\}$ and $\left\{ \mathrm{var}_{m,m}(\widehat{V}_{n_1,n_2}(i)) \right\}$, respectively, where $i = 1, 2, ..., N$. Then, the empirical mean and variance of $\widehat{V}_{n_1,n_2}$, representing accurate estimators of the true mean and variance, were computed as follows,

$$\mathbb{E}_e(\widehat{V}_{n_1,n_2}) = \frac{1}{N} \sum_{i=1}^{N} \widehat{V}_{n_1,n_2}(i) ,$$

$$\mathrm{var}_e(\widehat{V}_{n_1,n_2}) = \frac{1}{N} \sum_{i=1}^{N} \left[\widehat{V}_{n_1,n_2}(i) - \mathbb{E}_e(\widehat{V}_{n_1,n_2}) \right]^2 . \tag{24}$$

In the graphs, the preceding variance is normalized with the square of the true volume to obtain the square coefficient of error, namely,

$$\mathrm{CE}_e^2(\widehat{V}_{n_1,n_2}) = \frac{\mathrm{var}_e(\widehat{V}_{n_1,n_2})}{V^2} . \tag{25}$$

Likewise, the empirical mean and the mean square error of the variance predictors were computed as follows,

$$\mathbb{E}_e \left\{ \mathrm{var}_{m,m}(\widehat{V}_{n_1,n_2})_+ \right\}$$
$$= \frac{1}{N} \sum_{i=1}^{N} \mathrm{var}_{m,m}(\widehat{V}_{n_1,n_2}(i))_+ ,$$

$$\mathrm{MSE}_e \left\{ \mathrm{var}_{m,m}(\widehat{V}_{n_1,n_2})_+ \right\}$$
$$= \frac{1}{N} \sum_{i=1}^{N} \left[\mathrm{var}_{m,m}(\widehat{V}_{n_1,n_2}(i))_+ - \mathrm{var}_e(\widehat{V}_{n_1,n_2}) \right]^2 . \tag{26}$$

The corresponding coefficient of bias, and the relative mean square error, were computed as follows,

$$\mathrm{CB}_e \left\{ \mathrm{var}_{m,m}(\widehat{V}_{n_1,n_2})_+ \right\}$$
$$= \frac{\mathbb{E}_e \left\{ \mathrm{var}_{m,m}(\widehat{V}_{n_1,n_2})_+ \right\}}{\mathrm{var}_e(\widehat{V}_{n_1,n_2})} - 1,$$

$$\mathrm{RMSE}_e \left\{ \mathrm{var}_{m,m}(\widehat{V}_{n_1,n_2})_+ \right\} \tag{27}$$
$$= \frac{\mathrm{MSE}_e \left\{ \mathrm{var}_{m,m}(\widehat{V}_{n_1,n_2})_+ \right\}}{\left[\mathrm{var}_e(\widehat{V}_{n_1,n_2}) \right]^2} .$$

For simplicity, the square root of either side of the second Eq. 27 is denoted by $\mathrm{CE}_e \left\{ \mathrm{var}_{m,m}(\widehat{V}_{n_1,n_2})_+ \right\}$ in Table 3.

If desired, the actual error variance of the variance predictors can be easily recovered via the following identity,

$$\mathrm{MSE}(X) = \mathrm{var}(X) + \mathrm{Bias}^2(X), \tag{28}$$

where X represents a random variable standing for $\mathrm{var}_{m,m}(\widehat{V}_{n_1,n_2})_+$ in this case.

RESULTS

DISTRIBUTION OF THE NUCLEATOR ESTIMATORS

Consider first the simplest nucleator estimators, namely,

$$\widehat{V}_1 = \frac{4\pi}{3} \cdot l_+^3 , \qquad \text{(one ray)},$$

$$\widehat{V}_{11} = 4\pi F_{00} = \frac{2\pi}{3} \cdot (l_+^3 + l_-^3) , \quad \text{(two opposite rays)}. \tag{29}$$

The corresponding Monte Carlo histograms for H_1 are plotted in Fig. 4. In accordance with *Assessment of the nucleator variance predictors by Monte Carlo resampling* section, each of the two histograms incorporated 16 200 replicated estimates. The following features are worth mentioning.

(a) Each density function is independent of the orientation of the VA relative to the object, because the Monte Carlo procedure samples UR points on the unit sphere according to Eq. 5. In fact, for each of the two estimators $\widehat{V}_1$ and $\widehat{V}_{11}$, the histograms corresponding to either orientation could hardly be visually distinguished. In Fig. 4, both histograms correspond to the transversal orientation.

(b) The density functions are rather asymmetric, with long tails. Only about 27% and 24% of the $\widehat{V}_1$ and the $\widehat{V}_{11}$ replicates, respectively, exceeded the true volume V. The conditional means of the estimates that 'can be seen' in Fig. 4, namely,

$$
\begin{aligned}
\mathbb{E}_e(\widehat{V}_1 | \widehat{V}_1 < 6) &\approx 1.235, \\
\mathbb{E}_e(\widehat{V}_{11} | \widehat{V}_{11} < 6) &\approx 1.226,
\end{aligned}
\tag{30}
$$

are about -13% biased, and this bias has to be compensated by only about 3% of all the possible estimates namely by the estimates greater than 6.
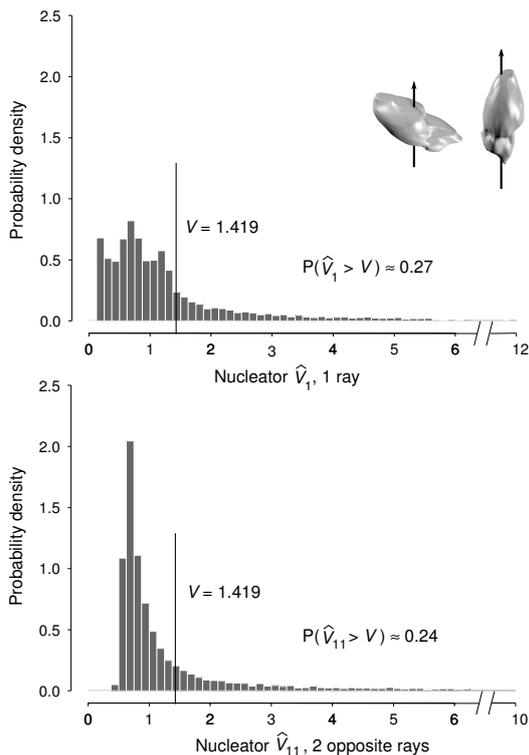
Fig. 4. *Empirical probability densities of the nucleator estimator of the volume of the right rat brain hemisphere from one ray and from two opposite rays, see Eq. 29. Neither of the two density functions depends on the relative position of the polar VA relative to the object. As soon as $n_1 > 1$, or $n_2 > 1$, however, the corresponding densities will generally depend on the choice of the vertical axis, see Fig. 5.*

(c) As soon as $n_1 > 1$, or $n_2 > 1$, however, the distribution of $\widehat{V}_{n_1,n_2}$ may depend significantly on the relative orientation of the VA, see Fig. 5.
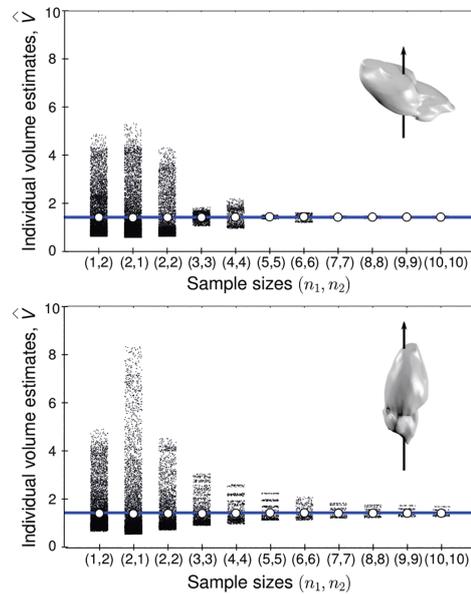
Fig. 5. *Individual Monte Carlo replications of the nucleator estimator $\widehat{V}_{n_1,n_2}$ of the volume of the right rat brain hemisphere for various sample sizes (n_1,n_2). Each dot represents a replication. As expected, the means (white circles) practically coincide with the true volume in all cases, because the estimator is unbiased. Note, however, that for each pair (n_1,n_2) the distribution of the estimates depends on the choice of the vertical axis relative to the object.*

PRECISION OF THE NUCLEATOR ESTIMATORS

Percent empirical coefficients of error — namely 100 times the square root of Eq. 25 — are displayed in Table 1 for each of the objects represented in Fig. 1. The following remarks are opportune.

(a) As a cross check of our Monte Carlo procedure, the data corresponding to the Gaussian particle (Fig. 1a) coincide with those in Table 1 from Gual-Arnau and Cruz-Orive (2002).

(b) For small sample sizes (n_1,n_2), the large CEs are a consequence of the long tails of the probability densities of the corresponding estimators, see Fig. 4.

(c) For small sample sizes the orientation of the VA relative to the object had a small effect on the precision of the estimator.

Table 1. *Empirical percent coefficient of error (Eq. 25) of the nucleator volume estimator (Eq. 13) for each of the four test objects illustrated in Fig. 1, and for various sample sizes. Gaussian: synthetic particle model (Section Material and its processing). H_1, H_2: right and left rat brain hemispheres respectively. $H_1 \cup H_2$: union in antithetic arrangement, see Fig. 1d. V: empirical, nearly true object volume.*

Object	V	$CE_e(\widehat{V}_{n_1,n_2})\%$ for sample sizes (n_1,n_2)							
		(1)	(1,1)	(1,2)	(2,1)	(2,2)	(3,3)	(4,4)	(5,5)
Gaussian	6.914	73.0	45.1	27.3	38.6	22.5	10.4	7.0	4.3
H_1 transverse	1.419	110.1	102.0	64.2	69.3	62.0	12.1	24.3	2.6
H_1 longitudinal	1.419	112.1	102.4	61.5	96.8	59.2	38.9	30.2	23.0
$H_1 \cup H_2$	2.856	154.7	127.0	99.2	120.0	97.7	13.7	23.4	6.6

(d) An object like the union of both rat brain hemispheres in antithetic position is not ideal for the nucleator. First, in practice it may be inconvenient to apply the nucleator to the union of two or more separate objects. And second, the corresponding estimator may also be inefficient because, if the pivotal point O lies outside the object (Fig. 1d), then a proportion of the volume estimators may be equal to zero whenever a ray misses the object. In fact, for that object,

$$\mathbb{P}(\widehat{V}_1 = 0) \approx 0.629, \quad \text{(one ray)},$$
$$\mathbb{P}(\widehat{V}_{11} = 0) \approx 0.506, \quad \text{(two opposite rays).} \quad (31)$$

PERFORMANCE OF THE VARIANCE PREDICTORS

Statistical properties of the theoretical error variance predictors $\text{var}_{m,m}(\widehat{V}_{n_1,n_2})$ given by Eqs.(16), (17), are illustrated in Fig. 6 for the right brain hemisphere, and for a number of sample sizes. Numerical summaries are displayed in Tables 2, 3 for each of the four object cases shown in Fig. 1.

A salient feature of Fig. 6 is that the variance predictions can be negative, specially for $(n_1,n_2) = (2,2)$ in the transverse VA case. To a lesser extent this was also observed for all the objects, see Table 4. An explanation of this fact is provided in the next subsection.

For the particular estimators $\widehat{V}_{12}$ and $\widehat{V}_{21}$, see Fig. 3, the error variance predictors are always non negative (Eq. 20) and their performance is illustrated separately in Fig. 7 for the right brain hemisphere. With reference to the notation used in Fig. 3, for a star shaped object the corresponding predicted square coefficients of error simplify as follows,

$$ce_{00}^2(\widehat{V}_{12}) = \frac{1}{3}\left(\frac{F_{00} - F_{01}}{F_{00} + F_{01}}\right)^2,$$
$$ce_{11}^2(\widehat{V}_{12}) = \frac{1}{15}\left(\frac{F_{00} - F_{01}}{F_{00} + F_{01}}\right)^2, \quad (32)$$
$$F_{00} = \frac{1}{6}(l_{1+}^3 + l_{1-}^3),$$
$$F_{01} = \frac{1}{6}(l_{2+}^3 + l_{2-}^3).$$

For $\widehat{V}_{21}$ the corresponding formulae are formally the same because F_{01} reads the same as F_{10} in this case.

In practice, a negative prediction will be taken to be equal to zero, hence the statistical summaries were computed with the non negative variance predictors only, see Eq. 23. Also, because the variance predictors were generally biased, the CE's displayed in Table 3 were computed via the mean square errors, see the second Eq. 27.

NEGATIVE VARIANCE PREDICTION: AN EXPLANATION

The fact that $\text{var}_{00}(\widehat{V}_{22})$, and to some extent $\text{var}_{11}(\widehat{V}_{22})$, may take negative values in a non negligible proportion of cases (Table 4 and Fig. 6) calls for a brief scrutiny.

Set

$$F_{11} = F_{00} + e_{00},$$
$$F_{10} = F_{01} + e_{01}. \quad (33)$$

Then, Eq. 16 becomes,

$$\text{var}_{0,0}(\widehat{V}_{22}) = \frac{\pi^2}{9}\left[-4(F_{00} - F_{01})(F_{11} - F_{10})\right.$$
$$\left. + (e_{00} + e_{01})^2 + 4(e_{00}^2 + e_{01}^2)\right]. \quad (34)$$
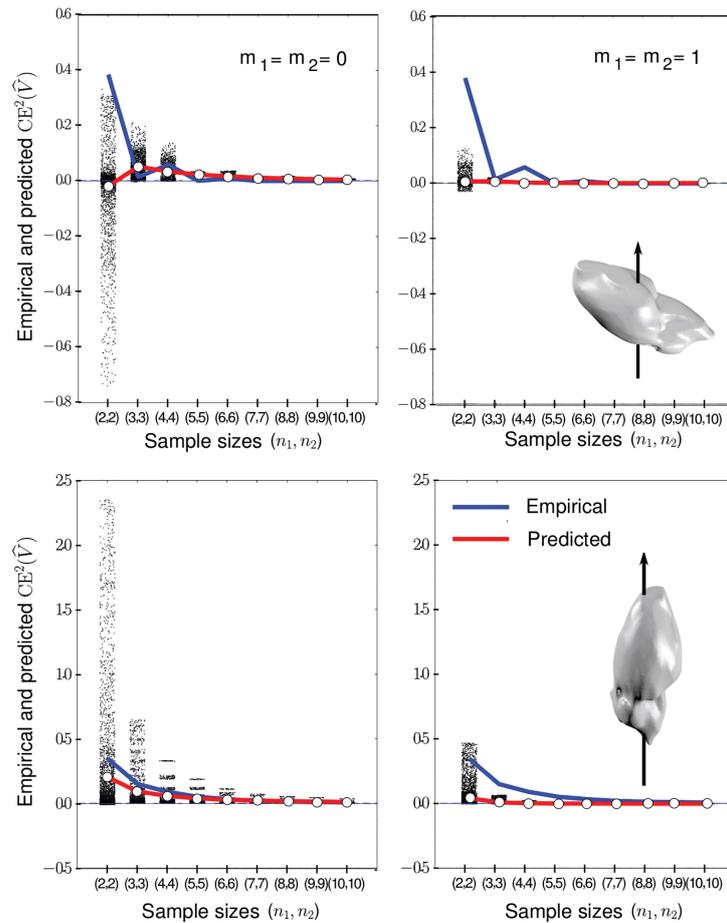
Fig. 6. *Performance of the predictors of* $\mathrm{CE}^2(\widehat{V}_{n_1,n_2})$ *based on Eqs. 16, 17, for each of the two positions of the vertical axis relative to the right brain hemisphere. Each dot represents a replicated predictor, and the white circles represent their means (negative replications included). Numerical data summaries are displayed in Tables 1–4.*

Table 2. *Empirical percent coefficient of bias (see the first Eq. 27) of the variance predictors given by Eqs. 16, 17. Negative predictions were replaced with zero, Eq. 23.*

Object	$\mathrm{CB}_e\{\mathrm{var}_{00}(\widehat{V}_{n_1,n_2})_+\}\%$					$\mathrm{CB}_e\{\mathrm{var}_{11}(\widehat{V}_{n_1,n_2})_+\}\%$				
	(1,2)	(2,1)	(2,2)	(3,3)	(4,4)	(1,2)	(2,1)	(2,2)	(3,3)	(4,4)
Gaussian	-42.6	-87.9	-14.2	64.1	115.5	-88.5	-97.6	-83.2	-83.0	-86.5
H_1 tr.	-49.3	-61.2	-92.2	269.5	-36.2	-89.9	-92.2	-97.8	-58.3	-96.4
H_1 lon.	-42.6	-96.7	-28.8	-37.3	-35.3	-88.5	-99.3	-86.9	-93.6	-95.8
$H_1 \cup H_2$	-78.8	-95.8	-81.7	712.2	103.6	-95.8	-99.2	-96.3	-16.3	-87.1
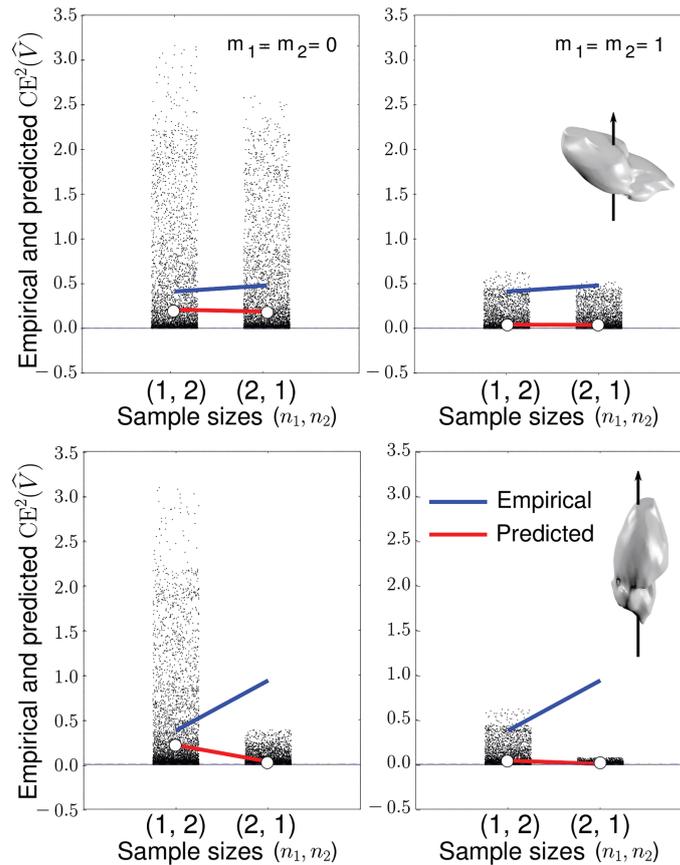
Fig. 7. *Completion of Fig. 5 with the special cases illustrated in Fig. 3. Corresponding numerical data summaries are displayed in Tables 1–3*

Table 3. *Empirical percent coefficient of error of the variance predictors given by Eqs. 16, 17. More precisely, the data are square root relative mean square error (namely the square root of the second Eq. 27).*

Object	$CE_e\{\mathrm{var}_{00}(\widehat{V}_{n_1,n_2})_+\}\%$					$CE_e\{\mathrm{var}_{11}(\widehat{V}_{n_1,n_2})_+\}\%$				
	(1,2)	(2,1)	(2,2)	(3,3)	(4,4)	(1,2)	(2,1)	(2,2)	(3,3)	(4,4)
Gaussian	97.5	91.1	100.1	169.8	192.7	90.2	97.7	85.6	84.5	87.1
H_1 tr.	127.4	108.2	93.8	391.1	69.0	92.9	93.9	97.9	66.3	96.5
H_1 lon.	129.8	96.9	141.4	107.6	98.1	91.9	99.3	90.9	94.2	96.0
$H_1 \cup H_2$	85.5	96.1	87.2	770.8	161.5	96.0	99.2	96.5	33.3	87.4

Table 4. *Empirical percent of negative error variance predictions.*

Object	$\mathrm{var}_{00}(\widehat{V}_{n_1,n_2})$					$\mathrm{var}_{11}(\widehat{V}_{n_1,n_2})$				
	(2,2)	(3,3)	(4,4)	(5,5)	(6,6)	(2,2)	(3,3)	(4,4)	(5,5)	(6,6)
Gaussian	13.9	0.0	0.0	0.0	0.0	6.0	0.0	0.0	0.0	0.0
H_1 tr.	37.2	1.6	4.7	0.0	0.0	19.8	0.6	11.7	305	40.9
H_1 lon.	16.6	0.0	0.0	0.0	0.0	7.7	0.0	0.0	2.2	0.0
$H_1 \cup H_2$	9.9	0.0	0.0	0.0	0.0	4.7	0.0	0.0	0.0	0.0
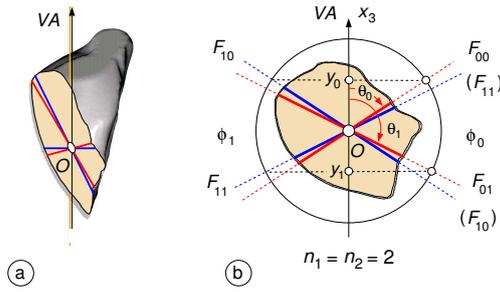
Fig. 8. *The special vertical nucleator design for $\widehat{V}_{22}$, whose variance predictors $\mathrm{var}_{00}(\widehat{V}_{22})$ and $\mathrm{var}_{11}(\widehat{V}_{22})$ may yield negative values, see Fig. 6, Table 4, and Negative variance prediction: an explanation subsection.*

If e_{00}, e_{01} are large enough, then $\mathrm{var}_{0,0}(\widehat{V}_{22}) \geq 0$. However, if e_{00}, e_{01} are small relative to the $F_{ij}, i, j = 0, 1$, then,

$$\mathrm{var}_{0,0}(\widehat{V}_{22}) \approx -\frac{4\pi^2}{9}(F_{00} - F_{01})(F_{11} - F_{10}) . \quad (35)$$

Thus, if F_{00}, F_{11} on the one hand, and F_{01}, F_{10} on the other, are not too different from each other (Fig. 8), then $F_{00} - F_{01}$ and $F_{11} - F_{10}$ are likely to be of the same sign, in which case we may get $\mathrm{var}_{0,0}(\widehat{V}_{22}) < 0$. In the extreme case in which $F_{00} \approx F_{11}$ and $F_{01} \approx F_{10}$ we have,

$$\mathrm{var}_{0,0}(\widehat{V}_{22}) \approx -\frac{4\pi^2}{9}(F_{00} - F_{01})^2 \leq 0 . \quad (36)$$

DISCUSSION AND CONCLUSIONS

Monte Carlo resampling. For each object Y with $O \in Y$, and for each direction $u \in \mathbb{S}^2$, the corresponding ray will hit one of more triangles from the reconstructed boundary of the object. The first step consisted in identifying the hit triangles, and the second was to compute each distance $l(u)$ from O of the intersection between the ray and the relevant triangle. This was computer intensive because, in the first step, every triangle was interrogated. There is scope for improving the efficiency of the procedure.

Statistical distribution of the more usual nucleator estimators. Because the nucleator estimator involves ray length raised to the power 3, its probability density may exhibit a long tail, specially for low sample sizes, see Fig. 4. Consequently, the correctness of ray sampling should be carefully controlled. In addition, the sample size (*e.g.*, the total number of

neurons analyzed) should not be too small because the coefficient of error of $\widehat{V}_{11}$, for instance, can be greater than 100% for a cell, see Table 1. The invariator (Cruz-Orive and Gual-Arnau, 2015) may be worth considering as an alternative, but this requires isotropically oriented pivotal sections.

Nucleator observations were strongly non normal, and trying to model their distribution via some transformation is likely to be futile because the data are strongly shape dependent. For this reason the model based approach suggested by Hobolth and Jensen (2002) is unlikely to succeed in this context.

Precision of the nucleator. The precision of $\widehat{V}$ increases rapidly for about $n_1, n_2 > 5$, specially for the right hemisphere with transverse VA. Here we have represented only a few data points with $n_1 = n_2$. We also resampled with $n_1 \neq n_2$, but the results did not add much information and their graphical representation was less convenient. In practice, the most interesting estimators are $\widehat{V}_{11}$, and $\widehat{V}_{12}$, because they can be implemented on a single vertical section (either physical, or virtual).

Performance of the theoretical error variance predictors. The variance of $\widehat{V}_{11}$ cannot be predicted because the sample is based on a single axis. For $\widehat{V}_{12}$, the relative bias of the predictor given by Eq. 20 was lower for $m_1 = m_2 = 0$ than for $m_1 = m_2 = 1$, see Table 2 and Fig. 7.

Under systematic sampling a certain degree of bias can always be expected from a variance predictor. In the present case, in addition to the bias the variance predictors exhibited a large degree of variation, see Table 3 and Figs. 6, 7. Still worse, the predictors were negative at times, see also Table 4. Actually, Fig. 8 suggests that the case $n_1 = n_2 = 2$ may provide a similar information as $n_1 = n_2 = 1$, which points toward inefficiency of the design in some cases. There is certainly scope to improve the design proposed in Gual-Arnau and Cruz-Orive (2002). Nucleator variance predictors will hardly ever be satisfactory, however, because the interesting estimators are based on very little information, and variance estimators (even the unbiased ones in classical statistics) tend to be unstable. On the other hand, for large sample sizes the true error variance soon becomes very small, and to predict it lacks interest.

Main conclusions:

(1) The nucleator is useful to estimate the mean individual volume of a population of particles. For the usual estimators $\widehat{V}_{11}, \widehat{V}_{12}$, it is important that ray sampling is implemented correctly, because for a single particle the probability density of these

estimators may exhibit a very long tail (for $\widehat{V}_{11}$ see Fig. 4). Thus, truncation of a mere 2–3% of the larger values may induce a bias of more than 10%, even if the pivotal point or 'nucleolus' has a central position in the particle. Moreover, the number of particles analyzed should be large enough, because the within cell coefficients of variation can be greater than 100%, see Table 1. Unless vertical sections are compulsory, the invariator (see Cruz-Orive and Gual-Arnau (2015), for a recent survey) may be a worthy alternative.

(2) As indicated in the preceding subsection, see also Fig. 8, for $n_1 \geq 2$ and $n_2 \geq 2$, the design studied here cannot be recommended because it may generate redundant data. Discontinuities can also be observed in the lower panel of Fig. 5 for $n_1 = n_2 \geq 3$, and in the corresponding empirical error variances, see the lower left panel of Fig. 6. This is probably due to interactions between the configuration of the ray sample and the shape of the particle.

(3) The performancae of the variance predictors was poor in general as far as bias and variation. A reason for this — apart from the deficiences of the design itself — may be that the covariogram model adopted in Gual-Arnau and Cruz-Orive (2002) exhibited bilateral symmetries — in general the true covariogram is symmetric only about the origin.

ACKNOWLEDGMENTS

REFERENCES

Cruz-Orive LM (1987). Particle number can be estimated using a disector of unknown thickness: the selector. J Microsc 145:121–42.

Cruz-Orive LM, Gual-Arnau X (2015). The invariator design: an update. Image Anal Stereol 34:147–59.

Cruz-Orive LM, Ramos-Herrera ML, Artacho-Pérula E (2010). Stereology of isolated objects with the invariator. J Microsc 240:94–110.

Foley A, Lane DA, Nielson GM, Franke R, Hagen H (1990). Interpolation of scattered data on closed surfaces. Comput Aided Geom D 7:303–12.

Gual-Arnau X, Cruz-Orive LM (2000). Systematic sampling on the circle and on the sphere. Adv Appl Prob 32:628–47.

Gual-Arnau X, Cruz-Orive LM (2002). Variance prediction for pseudosystematic sampling on the sphere. Adv Appl Prob 34:469–83.

Gundersen HJG (1988). The nucleator. J Microsc 151:3-21.

Hobolth A, Jensen EBV (2002). A note on design-based versus model-based variance estimators in stereology. Adv Appl Prob 34:484–90.

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

Jensen EBV (1998). Local Stereology. Singapore: World Scientific.

Ju T, Losasso F, Schaefer S, Warren J (2002). Dual contouring of hermite data. ACM T Graphic 21:339-46.

Klein R, Schilling A, Strasser W (1999). Reconstruction and simplification of surfaces from contours. In: Proc 7th IEEE Pacific Conf Comput Graph Appl, 198-207.

Tandrup T (1993). A method for unbiased and efficient estimation of number and mean volume of specified neuron subtypes in rat dorsal root ganglion. J Compar Neurol 329:269–76.

West MJ (2012). Basic stereology for biologists and neuroscientists. New York: CSH Laboratory Press, Cold Spring Harbor.