Stereology and Image Analysis. Ecs10 - Proceedings of the 10th European Congress of ISS,
(V.Capasso et al. Eds.), The MIRIAM Project Series, ESCULAPIO Pub. Co., Bologna, Italy, 2009

ADVANCES ON THE PRECISION OF SEVERAL STEREOLOGICAL VOLUME ESTIMATORS

KIÊN KIÊU[1] AND MARIANNE MORA[2]

[1]UR341, Mathématiques et Informatique Appliquées, INRA, F-78350 Jouy-en-Josas, France, [2]Modal'X, Université Paris Ouest Nanterre la Défense, 200 av. de la République, F-92000 Nanterre, France
e-mail: Kien.Kieu@jouy.inra.fr, mora@u-paris10.fr

ABSTRACT

Investigations of the precision of stereological estimators based on systematic test systems date back from early papers by D. Kendall and Georges Matheron published in 1950–1970. A widely used formula predicts the mean squared error of the planar area estimator based on point grid sampling. This formula is very simple since it involves only a standard shape parameter and the grid spacing. A similar formula for volume estimation has been proposed by Matheron. Also further formulae for other sampling schemes have been derived, in particular for serial sections, thick slabs, systems of quadrats...
All these formulae hold when the boundary of the target structure is isotropic. Extended formulae for anisotropic boundaries will be discussed. These extensions are derived from an idea already proposed by Matheron for the planar area estimator based on point grid sampling.
Computations of the MSE approximations and estimates can be performed using the R package pgs. Examples of calculations with pgs are provided.

Keywords: estimation, precision, sampling, stereology.

INTRODUCTION

Assessing the precision of stereological estimators based on systematic sampling is not straightforward due to correlations between neighbour geometric probes. That problem is clearly the object of ongoing research. It should be noticed that the situation is different in other fields of spatial statistics. In particular, in geostatistics a key step consists in modelling spatial correlations via the variogram. With a variogram model in hand, assessing the precision of a given sampling scheme is rather straightforward. In stereology, much of the effort is put on sampling design. In general, one tries to avoid modelling the investigated spatial structures. In particular, classical stereological estimators are unbiased for very large classes of spatial structures. However when turning to precision, developping "universal" methods is clearly a difficult task.

When investigating a population of objects sampled independently, the issue is not so critical. Sampling is performed at different scales and systematic sampling is often used at the smallest scale. At the highest level (object sampling), estimates can be considered as independent and it is easy to predict the mean squared error of estimators of population mean features (volume, surface area, number). However assessing the part of variability du to stereological sampling is required in order to distinguish between sampling variability and biological variability. Also quantifying sampling variability is of interest when designing a sampling scheme.

This paper focuses on planar area and 3D volume estimation where a significant corpus of methods is now available. Our aim is to review what pratical methods are available at the moment. Toy examples will be used together with the R package pgs which provides mean squared error predictions.

The second section is devoted to pioneering work of Kendall and Matheron on sampling by point grids. The mean squared error formula for the planar area estimator based on sampling by a square grid is now widely used. Results for grids with other shapes and 3D grids are also available. In the third section, other sampling devices such as serial sections and lattices of quadrats are considered. The fourth section shows how the mean squared error predictions can be used for designing sampling schemes. The fifth section is devoted to the estimation a posteriori of mean squared errors using collected data. The case of spatial structures with anisotropic boundaries is the object of the sixth section.

POINT GRID SAMPLING

The total area A of a planar region can be estimated by superimposing a grid of points onto the region and counting the number of points hitting the region. The

area estimator $\widehat{A}$ can be written as

$$\widehat{A} = aP \tag{1}$$

where a is the area of a fundamental tile of the point grid and P is the number of grid points hitting the region.

An example is shown in Figure 1. The planar region under study is Italy. Italy perimeter B is equal to 49.6ℓ and its area A is equal to $30\ell^2$ where ℓ is the length of the scaling bar. The sampling is based on a square grid of points with grid spacing $u = \ell$. For this square grid the area of a fundamental tile is $a = \ell^2$. Here, 27 grid points hit Italy. Therefore the area estimate is $27a = 27\ell^2$.

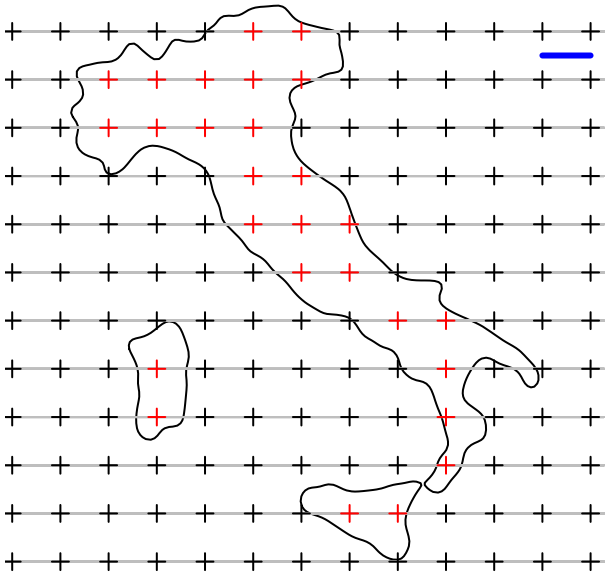

Fig. 1. *Square point grid superimposed onto Italy. The scaling bar is shown in blue. The planar area of Italy is estimated by counting the number of grid points hitting Italy (shown in red).*

It is well-known that the estimator $\widehat{A}$ is *unbiased* provided the sampling grid is Uniform Random (UR). Assessing the precision of $\widehat{A}$ through its Mean Squared Error (MSE) is not straightforward. One may expect the estimation error to be somewhat connected to the boundary length B of the target region, see Figure 2. Indeed, the MSE approximation as derived by Kendall (1948) and Matheron (1970; 1971) is as follows

$$\mathrm{MSE}\widehat{A} \simeq \frac{M}{4\pi^3} a^{3/2} B \tag{2}$$

where M is a scale-invariant grid parameter. For a square grid $M = 9.03$ and $M/(4\pi^3) = 0.0728$. Note that a slightly different constant 0.0724 may be found in the litterature instead of 0.0728, see e.g. Matheron (1971) or Gundersen and Jensen (1987). As noticed by Kendall, the grid parameter M is a multidimensional zeta function known as the Epstein zeta function. The value 0.0724 is based on a standard approximation involving only the more standard unidimensional Riemann zeta function. The difficult part in the approximation formula lies in M.

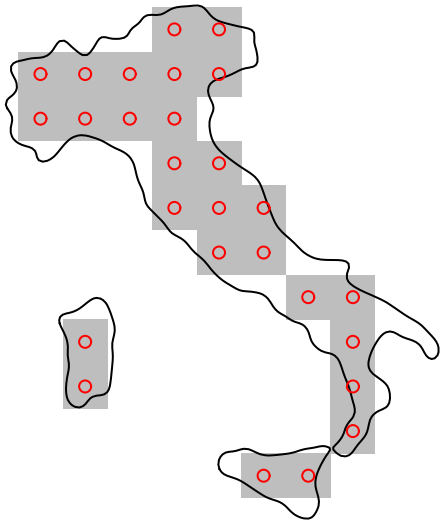

Fig. 2. *The area estimate based on point counts coincides with the total area of a square reconstruction of Italy (shown in grey). The reconstruction differs from Italy near the border.*

The pgs function area.mse implements the computation of Equation (2) for any planar point grid. It requires two main arguments : the point grid and the boundary length. In the example of Figure 1, the MSE approximation is obtained as follows

```
> plat = PPRectLat2(1,1)
> area.mse(plat,B=49.6)
[1] 3.612717
```

The first command line defines a rectangular lattice (function PPRectLat2) with horizontal and vertical spacings equal to 1. The square point grid is stored in a variable arbitrarily named plat. In the second command line, the point grid plat and a boundary length of 49.6 are provided to function area.mse for the computation of Equation (2). Note that pgs is unaware of units. Therefore it is the user responsability to use consistent units. In the code example above, both the grid spacings and the boundary length use ℓ as unit. The computed MSE must be interpreted as $3.61\ell^2$.

Using pgs, one can predict the MSE for any point grid. For instance, for a rectangular lattice with a horizontal spacing which is four times larger than the vertical spacing, the MSE computation yields

```
> plat = PPRectLat2(2,0.5)
> area.mse(plat,B=49.6)
[1] 9.007288
```

The Kendall-Matheron formula (2) is an approximation which holds for *dense* sampling grids: a must be "small enough" . However the theory does not precise how small a must be. In order to investigate that point, one may resort to simulations. By simulating independent realizations of the area estimator, one obtains Monte-Carlo unbiased approximations of the MSE to be compared with the Kendall-Matheron MSE approximation. In Figure 3, standard errors are plotted against the mean numbers $n = A/a$ of hitting points instead of a. For each value of n, 3000 replicates of the area estimator have been simulated in order to obtain a Monte-Carlo SE approximation. Grid spacing was chosen in order to get a range of 1–40 for n.

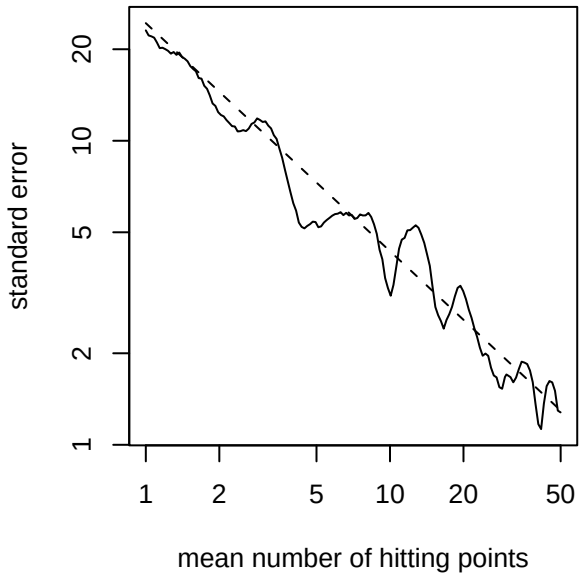

Fig. 3. *Log-log plot of the area estimator standard error against the mean number of hitting points. The region of interest is Italy as shown in Figure 1 sampled by square grids. Standard errors of the area estimator (continuous line) have been computed by means of simulations. The Kendall-Matheron SE approximation is shown as a dashed curve.*

According to the simulations, the MSE for the example of Figure 1 is equal to $1.70^2\,\ell^2 = 2.89\,\ell^2$. This result is to be compared with the MSE Kendall-Matheron approximation given above: $3.61\,\ell^2$.

In view of Figure 3, the grid spacing does not seem to be a serious issue. The "true" MSE shows oscillations, a phenomenon called *Zitterbewegung* by Matheron. It has been shown recently that the Zitterbewegung tends to vanish when one is considering a population of regions with varying sizes and shapes instead of a single region, see Kiêu and Mora (2004) for details.

MSE formula (2) has been derived under the assumption that the normals to the region boundary are isotropically distributed. It is not obvious whether this assumption is checked by Italy. There is an obvious anisotropy at a large scale with an extension along a preferred direction. However it is not quite clear whether there is a strong anisotropy of the normals (or tangents) to the boundary. Note that the isotropy condition requires the normals to the region boundary to be isotropically distributed with respect to the main grid directions. As a consequence, it is possible to fulfill the condition even if the boundary is anisotropic by just randomizing the grid orientation. Furthermore methods for predicting the MSEs for spatial structures with anisotropic boundary will be presented latter in the paper.

So far, we have focused on planar regions. One may expect the whole approach to extend to regions in spaces with higher-dimensions. This is indeed the case. In particular, the MSE formula for the 3D volume estimator based on sampling by a point grid is as follows

$$\mathrm{MSE}\,\widehat{V} \simeq \frac{M}{8\pi^3}\, v^{4/3} S, \tag{3}$$

where again M is a scale-invariant grid parameter, v is the volume of a fundamental tile of the point grid and S is the surface area of the region of interest. This formula holds for regions with an isotropic boundary. A method is provided by Matheron (1965) for deriving an approximation of the parameter M for any grid. The pgs function vol.mse can compute the approximation (3) for any 3D point grid. For example, for a cell nucleus with a diameter approximately equal to 8 microns sampled by a cubic grid with spacing 2 microns, pgs yields

```
> plat = PPRectLat3(2,2,2)
> vol.mse(plat,S=4*pi*4^2)
[1] 214.4094
```

Another standard 3D point grid is the body centered grid where a 2D square point grid lying on a plane is shifted halfway into the next plane. The MSE computation yields

```
> plat = PPBCRectLat3(2,2,2)
> vol.mse(plat,S=4*pi*4^2)
[1] 209.9755
```

OTHER SAMPLING SCHEMES

In practical stereology, commonly used sampling schemes are not based only on point grids. For instance, in order to estimate its area, one may sample Italy by finite point patterns contained in a systematic arrangement of quadrats, see Figure 4. Probes with higher dimensions are also used: instead of counting points within quadrats, one may measure the area of Italy contained in the quadrats. More generally, a region can be sampled by a lattice of figures (instead of points). At each lattice node, one measures the content of Italy within the figure.

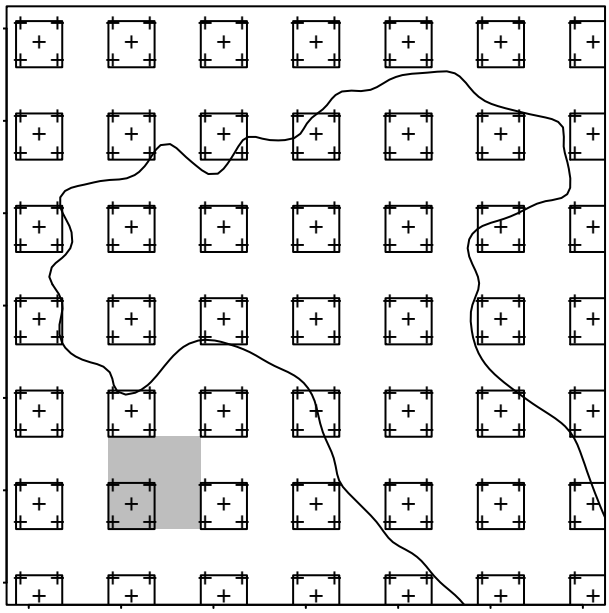

Fig. 4. *Sampling by figure lattices. The set of crosses defines a lattice of point patterns. The set of quadrats defines a lattice of quadrats. Grey: a fundamental tile (identical for both lattices).*

Matheron (1965; 1971) has designed a general framework for deriving MSE approximation formulae when sampling by figure lattices. Let us come back for a while to point grids and formulae (2–3). In order to derive these formulae, Kendall and Matheron have used Fourier analysis. Such an approach is standard when considering sampling of signals. A region like Italy can be considered as a binary function defined on the whole plane. This function, usually called indicator function, takes value 1 at a point inside Italy and value 0 outside. The squared modulus of the Fourier transform of the indicator function is called the spectral density. If the sampling grid is dense enough (u or a small enough), the MSE of the area estimator only depends on values of the spectral density at points (of the spectral space) far from the origin, that is at points associated with high frequencies. For an indicator function, high frequency variations occur along the region boundary. In particular, it is possible to derive approximations of the spectral density (for high frequencies) only depending on the region boundary.

A similar approach applies when one replaces a point lattice by a figure lattice. Instead of an indicator function, one must consider the regularized function on the whole plane that associates to each point the value of the content of the region seen in the figure centered at the considered point. Regularization is often called convolution. In the Fourier space, convolution translates as a simple multiplication. Using this key fact, it is possible to derive approximations of the Fourier transform of the regularized function involving only the region boundary.

Using Matheron's general framework, further MSE approximations for area and volume estimators have been derived for sampling by figure lattices (Gual Arnau and Cruz-Orive, 1998; Kiêu and Mora, 2005; Kiêu and Mora, 2006).

The MSE approximations for area estimation take again the form given in Equation (2) where M is a scale-invariant parameter depending on the figure lattice. Note that M is sensitive to any scaling changing only the fundamental tile or only the figure. It remains invariant only under scaling applied to the whole figure lattice. The parameter M involves the Epstein zeta function and the geometric covariogram of the figure (point pattern, quadrat, line segment...). Again the approximation (2) can be computed by pgs for a variety of figure lattices. Below is the R code providing MSE's for both figure lattices shown in Figure 4.

```
> pplat = PPRectLat2(1,1,5,0.4)
> area.mse(pplat,B=49.6)
[1] 0.6158561
> qlat = QRectLat2(1,1,0.5)
> area.mse(qlat,49.6)
[1] 0.7481244
```

Function PPRectLat2 defines a rectangular lattice of point patterns. The first two arguments are the horizontal and vertical spacings. The third argument defines the point pattern (4 points lying at a square corner and a middle point). The fourth argument defines the side length of the square containing the point pattern. The third command line defines a quadrat lattice using the function QRectLat2.

For the lattice of point patterns, the MSE given by 3000 Monte-Carlo simulations is equal to $0.47\ell^2$. This result is to be compared with the approximation MSE $\simeq 0.62$ obtained above.

MSE approximations are also available for 3D figure lattices. For instance, let us consider a 3D lattice of horizontal point patterns. Each point pattern consists of 5 points, 4 lying at a square corners and the fifth one in the square middle. Such a 3D lattice can be obtained by repeating a 2D lattice of point patterns on a series of parallel planes (with constant vertical spacing). The MSE of the volume estimator can be predicted using pgs as follows

```
> pplat = PPRectLat3(2,2,2,5,1)
> vol.mse(pplat,S=4*pi*4^2)
[1] 57.61168
```

Function PPRectLat3 defines a 3D lattice of horizontal point patterns. The first three arguments define the lattice spacings in the three orthogonal directions. The fourth argument defines the point pattern. The fifth argument defines the side length of the horizontal square containing the point pattern.

Series of parallel lines or planes are also considered as figure lattices. A major difference with the previous figure lattices is that lines (planes) are unbounded. MSE approximations are available when the unbounded figure is obtained by shifting continuously an initial bounded figure along a (generating) line or plane. In 2D, there are mainly two unbounded figures of interest: parallel lines and parallel strips. For such figure lattices, a fundamental tile is defined as a segment line parallel to the lines or strips and extending between neighbour lines or upper boundaries of neighbour strips. If l is the length of a fundamental tile, the MSE approximation is

$$\mathrm{MSE}\widehat{A} \simeq \frac{M}{4\pi^3} l^3 B. \tag{4}$$

For parallel lines, M is equal to 2.40 (twice the Riemann zeta function at 3). Note that the only difference with Equation (2) is that $a^{3/2}$ is replaced by l^3. Approximation (4) is computed by function area.mse:

```
> lser = LLat2(1)
> area.mse(lser,B=49.6)
[1] 0.9614509
```

The function LLat2 defines a series of equidistant parallel lines. The first argument is just the spacing between neighbour lines.

In 3D, standard lattices of unbounded figures are serial planes, series of thick slices and the so-called fakir bed. For planes and thick slices, the MSE approximation has the form

$$\mathrm{MSE}\widehat{V} \simeq \frac{M}{8\pi^3} l^4 S. \tag{5}$$

For parallel planes, M simplifies as $\pi^4/45$. For the fakir bed, the MSE approximation has the form

$$\mathrm{MSE}\widehat{V} \simeq \frac{M}{8\pi^3} a^2 S. \tag{6}$$

Both approximations are implemented in pgs.

Again all MSE approximations hold under the condition that the boundary normals are isotropically distributed or that the lattice of figures is oriented with an isotropic distribution.

COMPARING SAMPLING SCHEMES

All MSE approximations (2–6) have basically the same form: products of two terms, one depending on the sampling figure lattice, the other one being the boundary content. Therefore it is possible to compare sampling schemes *per se* independently of the investigated region.

This is useful when one has to choose between different sampling schemes. Back to the example of Figure 1, one may estimate Italy area by counting grid points hitting Italy or by measuring the total intercept length of Italy with say the horizontal lines (shown in grey). Comparing both sampling schemes through the ratio of their MSE's we get

```
> plat = PPRectLat2(1,1)
> lser = LLat2(1)
> area.mse(plat)/area.mse(lser)
[1] 3.757568
```

The SE for the point grid is about twice the SE for parallel lines.

The MSE approximations are also useful when one is determining some parameters of a sampling scheme in order to achieve a nominal precision. For instance, one aims at predicting the spacing u of a square point grid yielding a SE equal to $0.05A$.

$$0.05^2 A^2 \simeq \frac{M}{4\pi^3} u^3 B.$$

At this stage, neither the total area A nor the boundary length are available. They can be replaced by rough guesses. Or, as suggested by Gundersen and Jensen (1987), one may try to guess values for A and the

standard shape parameter $B/\sqrt{A}$. The assessment of the shape parameter can be done visually using a nomogram as given in (Gundersen and Jensen, 1987). The equation to be solved writes as

$$0.05^2 A^{3/2} \simeq \frac{M}{4\pi^3} u^3 \frac{B}{\sqrt{A}},$$

where A and $B/\sqrt{A}$ are replaced by "guestimates" and M must be replaced by its numerical value provided above. Using pgs function latscale, one can solve this type of problem for arbitrary lattices of figures. As an example, consider Italy to be sampled by a lattice of 5 point patterns as shown in Figure 4. The parameter one aims to tune in order to obtain a SE equal to $0.1A$ is the side length of the fundamental tile shown in grey. The problem is solved by the following commands:

```
> pplat = PPRectLat2(1,1,5,0.4)
> latscale(pplat,A=30,shape=10,
+          CE.n=0.1,upper=4)
[1] 1.768231
```

The argument upper defines the upper bound of the interval where the tile side length is searched.

Using pgs, it is also possible to design optimal sampling schemes under some constraints. For instance, consider a lattice of point patterns. The lattice is hexagonal with a fixed spacing say 1. There are 5 points per pattern and they are confined to a square of side length 0.3. Figure 5 shows how the 5 points must be arranged inside the square in order to minimize the MSE of the area estimator. Other examples of optimal grids are given in Kiêu and Mora (2006). However in general optimal grids do not improve significantly compared to most standard grids used in practice. For instance, moving the left middle point in Figure 5 to the square middle does not increase the MSE more than 5%.

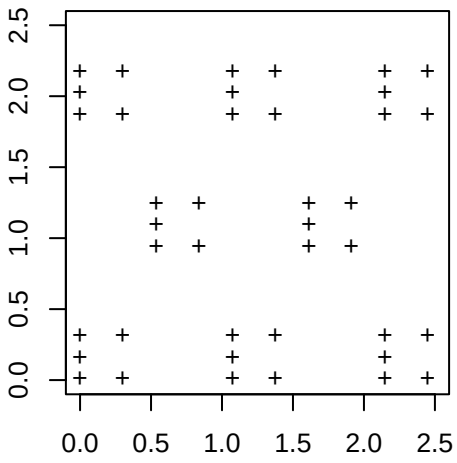

Fig. 5. *Optimal point pattern lattice.*

An intriguing problem in stereology is the comparison of nested figure lattices. For instance compare the square point grid and the vertical lines in Figure 1. The point grid is contained in the line grid and previous computations show that the line grid MSE is less than the point grid MSE. For a counter-example, consider both figure lattices in Figure 4. The lattice of point patterns is contained in the grid of quadrats. However, the MSE of the latter is more than the MSE of the former. Therefore enlarging a figure does not necessarily decrease the MSE. This statement is rather counterintuitive. For the comparison of the point and line grids, one may use Rao-Blackwell theorem for proving that the larger figure has a lower MSE (Lantuéjoul, 1988; Baddeley and Cruz-Orive, 1995). However there are many pairs of nested test systems that do not fulfill the conditions of the Rao-Blackwell theorem. Such cases are called Smit paradoxes. Paradoxes have been found by Jensen and Gundersen (1981); Baddeley and Cruz-Orive (1995); Voss and Cruz-Orive (2009). These counterexamples are based either on particular examples of deterministic regions or on specific probabilistic models of random sets. Using the MSE approximations (2–6), one can exhibit more paradoxes without making assumptions on the investigated region. This shows that Smit paradox is somewhat intrisic to sampling. Further examples of Smit paradox are provided in (Kiêu and Mora, 2006).

MSE ESTIMATION IN THE ISOTROPIC CASE

The MSE formulae (2–6) holding under the isotropy assumption involve two terms: one depending on the sampling scheme that can be computed a priori, the other one depending on the boundary (perimeter or surface area). If an estimate of the boundary perimeter (surface area) is available, it can be plugged in formulae (2–6) in order to get a MSE estimate.

A straightforward approach consists in designing a stereological sampling scheme for estimating simultaneously the area and the perimeter (the volume and the surface area in 3D). Note that sampling figures with dimension larger than 0 allow unbiased perimeter and surface area estimation. For instance, when sampling Italy by parallel lines, its perimeter can be estimated using the total number of intercepts. Another example in 3D is sampling by serial sections. In practice, measurement of profile areas on sections requires their segmentation by image processing. Once segmented, the profile perimeters can easily be

measured and the total surface area can be estimated from the profile perimeters.

An alternative approach consists in using only data collected for area or volume estimation. It is supposed that individual data for each sampling figure are available. In the examples of Figure 4, areas inside each sampling quadrat or number of hits for each point pattern should be available. For the figure of point patterns, the data to be used can be displayed as

```
0 0 1 4 5 5 0 0 0 0
0 5 5 5 5 2 0 0 0 0
0 5 4 5 5 0 0 0 0 0
0 0 0 2 5 5 0 0 0 0
0 0 0 0 5 5 3 0 0 0
0 0 0 0 0 5 5 3 0 0
0 0 3 0 0 0 1 5 5 1
0 0 5 0 0 0 0 1 5 0
0 0 4 0 0 0 0 0 4 0
0 0 0 0 0 0 0 1 4 0
0 0 0 0 0 1 5 3 0 0
```

First consider an easy case: area estimation based on point grid sampling. The available data are

```
0 0 0 1 1 0 0 0
1 1 1 1 1 0 0 0
1 1 1 1 0 0 0 0
0 0 0 1 1 0 0 0
0 0 0 1 1 1 0 0
0 0 0 0 1 1 0 0
0 0 0 0 0 0 1 1
0 1 0 0 0 0 0 1
0 1 0 0 0 0 0 1
0 0 0 0 0 0 0 1
0 0 0 0 0 1 1 0
```

The number of pairs 1-0 is related to the boundary length: for each such pair, the line segment joining the two corresponding grid points crosses the boundary (at least once). Hence the number of horizontal pairs 1-0 gives an approximation of the total projection length onto the vertical axis. Using pairs in different directions, one gets an estimate of the total perimeter. This idea was first suggested by Matheron (1971), see alsoCruz-Orive (1993). It should be noticed that this method tends to underestimate the perimeter. Obviously, the line segment joining a pair 1-0 may cross the boundary more than once. Also line segments joining pairs with identical values 0-0 and 1-1 may cross (at least twice) the boundary.

The pgs function area.mse.est computes both the perimeter and MSE estimates:

```
> flat = PPRectLat2(1,1,1)
> fldat = FigLatData(flat,hits)
> area.mse.est(fldat,iso=TRUE,
                mse.only=FALSE)
$B.est
[1] 39.26991

$mse.est
[1] 2.860304
```

Above hits is the matrix of binary data. The variable fldat contains the data array and parameters related to the sampling grid. The argument iso is given the value TRUE in order to specify that the MSE must be estimated under the isotropy assumption. Setting mse.only to FALSE, one gets both estimates.

The same approach extends to the case where the sampling figure does not reduce to a single point. It can be shown that short distance variations depend both on the perimeter (surface area) and the geometry of the sampling figure. Furthermore using observed short distance variations (empirical covariogram near the origin) and the sampling figure geometry, it is possible to estimate the perimeter (surface area). For the lattice of point patterns, one gets

```
> flat = PPRectLat2(1,1,5,0.4)
> fldat = FigLatData(flat,pp.counts)
> area.mse.est(fldat,iso=TRUE,
                mse.only=FALSE)
$B.est
[1] 41.47922

$mse.est
[1] 0.5150249
```

MSE ESTIMATION UNDER MILD ANISOTROPY

The MSE approximations (2–6) hold under the assumption that the boundary is isotropically oriented. Matheron (1971) proposed a remedy for the case where, by stretching the investigated region along a given direction, its boundary can be made isotropic. Matheron focused on sampling by a square point grid and the special case where the stretching direction coincides with one direction of the sampling grid. The stretching factor can be estimated by comparing the number of pairs 1-0 in the horizontal and vertical directions.

This approach has been generalized in pgs to the case where the stretching direction is not known a

priori. Considering pairs 1-0 in 4 directions, one can estimate simultaneously the stretching direction and factor. For the unit square point grid, one gets

```
> flat = PPRectLat2(1,1,1)
> fldat = FigLatData(flat,hits)
> area.mse.est(fldat,iso=FALSE,
                mse.only=FALSE)
$B.est
[1] 38.65081

$deformation
           [,1]      [,2]
[1,]  0.9262231 0.6027627
[2,] -0.4935774 0.7584457

$mse.est
[1] 2.713006
```

The component list deformation returned by area.mse.est defines the stretching matrix.

For the lattice of point patterns, one gets

```
> flat = PPRectLat2(1,1,5,0.4)
> fldat = FigLatData(flat,pp.counts)
> area.mse.est(fldat,iso=FALSE,
                mse.only=FALSE)
$B.est
[1] 41.02666

$deformation
           [,1]      [,2]
[1,]  0.9363621 0.6710069
[2,] -0.5056478 0.7056104

$mse.est
[1] 0.4688714
```

The results constrast depending on whether one considers the perimeter or the MSE estimates. Concerning perimeter, results are slightly worse than those based on the isotropy assumption both for the point grid and the lattice of point patterns. Concerning the MSE estimation, there is a slight improvement for the lattice of point pattern. An interpretation is that Italy border cannot really be made isotropic by a simple stretching. This is supported by an analysis of the numbers of pairs 1-0 in the four directions:

East	South	North-East	South-East
13	12	18	15

One expects to find the largest number of differing pairs in the North-East direction since Italy lies mainly along the perpendicular direction. But one should also notice that the number of differing pairs in the South-East direction is not small compared to the other directions.

The method is better exemplified if one considers Sardinia instead of the whole Italy. Sardinia area is equal to $2.2\,\ell^2$ and its perimeter is equal to $6.23\,\ell$. Consider sampling by a lattice of point patterns as shown in Figure 6.

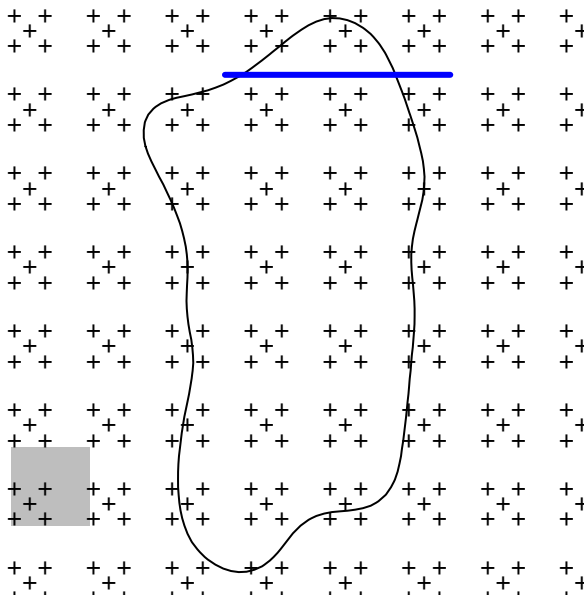

Fig. 6. *Sardinia sampled by a lattice of point patterns. Blue: scaling bar.*

MSE estimates provided by pgs are

```
> flat = PPRectLat2(0.35,0.35,5,
                    0.136)
> fldata = FigLatData(flat,
                      counts.sard)
> area.mse.est(fldat.sard,iso=TRUE,
               mse.only=FALSE)
$B.est
[1] 4.362244

$mse.est
[1] 0.002464126

> area.mse.est(fldata,iso=FALSE,
               mse.only=FALSE)
$B.est
[1] 5.957941

$deformation
            [,1]       [,2]
[1,] 1.33677766 -0.1782195
```

```
[2,] 0.09799083  0.7350034

$mse.est
[1] 0.002984931
```

Here, taking into account the anisotropy of Sardinia boundary and estimating the stretching making it isotropic really improves the perimeter estimation.

Figure shows Sardinia made isotropic by stretching.

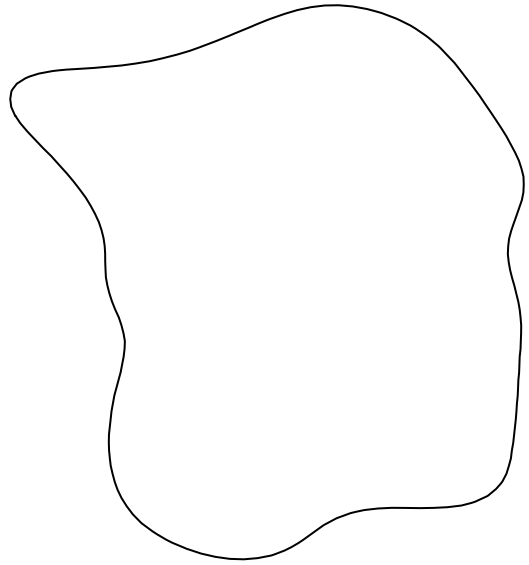

Fig. 7. *Isotropic Sardinia*

CONCLUSION

Concerning the precision of planar area and volume stereological estimators, there are now simple MSE approximations and estimation methods for a large range of sampling schemes used in practice. Theoretical fundations for the MSE approximation formulae are provided by Kiêu and Mora (2004; 2005), Kiêu and Mora (2006), Janáček (2006). MSE approximations are particularly useful when designing a sampling scheme. These approximations involve a multidimensional zeta function which can be computed numerically using the R package pgs. MSE estimation requires either to implement a specific sampling scheme for the estimation of perimeter or surface area or can be performed by analyzing the data collected for planar area or volume estimation.

In the future, the package pgs may evolve as a C or C++ library. This would make pgs available for developers of computer assisted stereological tools.

An important issue for future research is the extension of the approaches used for planar area and volume estimation to the estimation of lower dimensional features like length, surface area and number.

Recent investigations on the precision of systematic sampling have also concerned other aspects which may be of interest in practice. Some advances focus on the case where sampling is along a single axis, typically Cavalieri sampling. According to formulae (4) and (5), the MSE decreases to zero as fast as l^3 (in 2D) or l^4 (in 3D) to 0, where l is the distance between neighbour parallel probes. This behaviour is expected for regions with piecewise smooth boundary and general orientation with respect to the probes. Garcia-Fiñana and Cruz-Orive (2000) has shown that other behaviours are possible. Indeed any other exponent of l is theoretically possible. Examples with particular bodies with specific orientation with respect to the section planes have been provided. This theory may also be quite useful when considering regions with complex boundaries, that is boundaries with non-integer dimensions. Considering boundaries with non-integer dimensions is not so uncommon: one often expect boundary length to vary with magnification. Garcia-Fiñana and Cruz-Orive (2000) theory could be useful for obtaining MSE approximations at very different scales.

Another important issue concerns location errors. The MSE approximations given in this paper assume that the sampling figures are exactly (or rather precisely) repeated along a lattice. Location errors occur when the sampling involve physical operations like physical sectioning. A first exploration of consequences of irregular spacing on the MSE can be found in Baddeley *et al.* (2006).

Stereological estimation of lower dimensional features like length and surface area involves sampling in directional spaces, i.e. on the circle or on the sphere. Systematic sampling on spherical spaces has been the object of several papers (Gual-Arnau and Cruz-Orive, 2000; Cruz-Orive and Gual-Arnau, 2002). The special case of sampling by line grids has also been treated (Moran, 1966; Sandau, 1987; Hahn and Sandau, 1989; Sandau and Hahn, 1994).

As announced in the introduction of this paper, focus has been put on a design-based approach with few assumptions on the investigated region. An example of model-based approach has been provided by Hobolth and Jensen (2002) where planar area estimation based on sampling by rays through a reference point (2D nucleator principle) is considered.

REFERENCES

Baddeley A, Dorph-Petersen K, Jensen EV (2006). A note on the stereological implications of irregular spacing of sections. Journal of Microscopy 222:177–81.

Baddeley AJ, Cruz-Orive LM (1995). The Rao-Blackwell theorem in stereology and some counterexamples. Advances in Applied Probability 27:2–19.

Cruz-Orive LM (1993). Systematic sampling in stereology. In: Proceedings of the 49th Session of the International Statistical Institute, vol. 2.

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

Garcia-Fiñana M, Cruz-Orive LM (2000). Fractional trend of the variance in cavalieri sampling. Acta Stereologica 19:71–9.

Gual Arnau X, Cruz-Orive LM (1998). Variance prediction under systematic sampling with geometric probes. Advances in Applied Probability 28:982–92.

Gual-Arnau X, Cruz-Orive LM (2000). Systematic sampling on the circle and on the sphere. Advances in Applied Probability 32:628–47.

Gundersen HJG, Jensen EB (1987). The efficiency of systematic sampling in stereology and its prediction. Journal of Microscopy 147:229–63.

Hahn U, Sandau K (1989). Precision of surface area estimation using spatial grids. Acta Stereologica 8:425–30.

Hobolth A, Jensen E (2002). A note on design-based versus model-based variance estimation in stereology. Advances in Applied Probability 34:484–90.

Janáček J (2006). Variance of periodic measure of bounded set with random position. Comment Math Univ Carolinae 47:473–82.

Jensen EB, Gundersen HJG (1981). Stereological ratio estimation based on counts from integral test systems. Journal of Microscopy 125:51–66.

Kendall DG (1948). On the number of lattice points inside a random oval. Quarterly Journal of Mathematics Oxford Second Series 19:1–26.

Kiêu K, Mora M (2004). Asymptotics for geometric spectral densities and a stochastic approach of the lattice-point problem. Mathematicae Notae AÑO XLII:77–93.

Kiêu K, Mora M (2005). Stereological estimation of mean volume: precision of three simple sampling designs. Tech. Rep. 2005–1, Unité de Mathématiques et informatique appliquées, INRA, Domaine de Vilvert, F-78350 Jouy-en-Josas.

Kiêu K, Mora M (2006). Precision of stereological planar predictors. Journal of Microscopy 222:201–11.

Lantuéjoul C (1988). Some stereological and statistical consequences of Cartier's formula. Journal of Microscopy 151:265–76.

Matheron G (1965). Les variables régionalisées et leur estimation. Paris: Masson.

Matheron G (1970). La théorie des variables régionalisées et ses applications. Tech. rep., Ecole Nationale Supérieure des Mines de Paris.

Matheron G (1971). The theory of regionalized variables and its applications. Tech. rep., Centre de morphologie mathématique, Ecole des mines de Paris.

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

Sandau K (1987). How to estimate the area of a surface using the spatial grid. Acta Stereologica 6:31–6.

Sandau K, Hahn U (1994). Some remarks on the accuracy of surface area estimation using the spatial grid. Journal of Microscopy 173:67–72.

Voss F, Cruz-Orive LM (2009). Second moment formulae for geometric sampling with test probes. Statistics To appear.