Automatic sampling for unbiased and efficient stereological estimation using the proportionator in biological studies

J.E. GARDI, J.R. NYENGAARD & H.J.G. GUNDERSEN

Stereology and Electron Microscopy Research Laboratory and MIND Center, University of Aarhus, Aarhus, Denmark

Summary

Quantification of tissue properties is improved using the general proportionator sampling and estimation procedure: automatic image analysis and non-uniform sampling with probability proportional to size (PPS). The complete region of interest is partitioned into fields of view, and every field of view is given a weight (the size) proportional to the total amount of requested image analysis features in it. The fields of view sampled with known probabilities proportional to individual weight are the only ones seen by the observer who provides the correct count. Even though the image analysis and feature detection is clearly biased, the estimator is strictly unbiased. The proportionator is compared to the commonly applied sampling technique (systematic uniform random sampling in 2D space or so-called meander sampling) using three biological examples: estimating total number of granule cells in rat cerebellum, total number of orexin positive neurons in transgenic mice brain, and estimating the absolute area and the areal fraction of β islet cells in dog pancreas. The proportionator was at least eight times more efficient (precision and time combined) than traditional computer controlled sampling.

Introduction

The proportionator combination of biased image analysis and non-uniform sampling leading to unbiased estimation was previously studied using simulation (Gardi *et al.*, 2007). The proportionator combines computerized image analysis and non-uniform sampling, providing a more precise estimates using less workload than any other sampling method. The proportionator is based on automatic weight assignment to every field of view using image analysis, followed by systematic uniform random sampling (SURS) on the cumulative weights. An unconditionally unbiased estimate is ensured using well-known general statistical techniques (Hansen & Hurwitz, 1943; Horvitz & Thompson, 1952). The so-called Horvitz-Thompson estimator provides an unbiased estimate when the actual counts in the sampled fields of view are 'correct' (the actual counts are done by an expert user and not by image analysis) and the exact sampling probability of every field of view is known.

The weight of each field of view is automatically assigned by image analysis. The image analysis assigns weight to a field of view according to the amount of a requested image analysis feature. For estimating the number of green GFP-expressing neurons, for example, the weight of each field of view may be its amount of green colour observed under fluorescence illumination.

As shown in Fig. 1, the fields of view are first arranged according to the smooth fractionator (Gundersen, 2002) based on their weights, and then the cumulative weight **Z** is computed for this ordering. With a random start, a sample of the specified size **n** is sampled systematically on the ordinate of cumulative weight, using a sampling period of **Z/n**. These fields of view with known co-ordinates are then presented to the user.

The expert user assigns the unbiased count x_i for each sampled field of view with a weight z_i, using any relevant stereological probe (points, lines, frames, or disectors, optical or physical). The unbiased estimate X of the total content in the section is then simply

$$X := \frac{Z}{n} \sum_{i}^{n} \frac{x_i}{z_i}. \tag{1}$$

The correlation between the biased weight of a field of view and the correct count in the field may be positive or negative.

Correspondence to: Jonathan E. Gardi. Tel: (+45) 8942 2945; Fax: (+45) 8942 2952; e-mail: Jonathan.Gardi@ki.au.dk
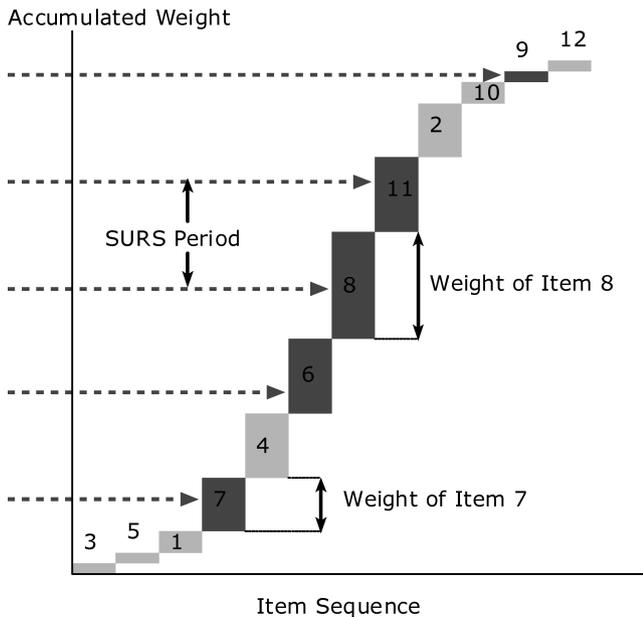
Fig. 1. Proportionator sampling. The ordinate shows the accumulated weights. Sampling on the ordinate is systematic uniform random sampling, after a smooth fractionator arrangement of the fields of view according to their weight. The sampled fields of view are marked with darker colour. Figure modified from (Gardi *et al.*, 2007).

Regardless of that, the estimate is always unbiased. The precision (*CE*) is, however, much dependent on the relationship between weight and count: the more positive the better the precision, if absent or negative the precision may be rather poor. Also, all kinds of noise in the relationship between weight and count reduce precision (for example, a low weight with a high count).

This study compares the actual performance of the proportionator to the traditional SURS (Gundersen & Jensen, 1987; Gundersen *et al.*, 1999) by applying it in three biological examples: estimating total number of granule cells in rat cerebellum, total number of orexin neurons in transgenic mice brain, and absolute and relative area of β islet cells in dog pancreas.

Methods and materials

In the preceding papers (Gardi *et al.*, 2006, 2007), smooth fractionator and proportionator sampling were described, tested and compared using simulations. The simulation framework was built on top of the existing stereological software CAST (VisioPharm, Hørsholm, Denmark). As mentioned in Appendix A in Gardi *et al.* (2006), the weight assignment procedure was designed and implemented from the start as an external component to CAST as a dynamically loaded library (dll). Currently, this weight assignment gets an image and one requested colour voxel as input, and gives back a weight as output.

The weight assignment used in this study is very basic but robust. The input image consists of voxels, and the requested colour, chosen by the user, is also a voxel. Each voxel has a colour, which is a mixture of red, green and blue. The voxels are observed in 3D colour space with these three fundamental colours as axes (Fig. 2). The distance *D* from every colour voxel in the image to the requested colour voxel is measured in this 3D colour space. Since the fundamental colour values are in the range of 0–255, the maximum possible distance in this cube of 3D space is $\sqrt{(255^2 + 255^2 + 255^2)} = 441.67$. For each voxel, the proximity to the requested voxel as a percentage of the maximum distance in the 3D colour space calculated:

$$Proximity = 100 \cdot \frac{441.67 - D}{441.67}. \qquad (2)$$

An additional feature in the weight assignment is that the user may indicate the minimal colour proximity beyond which voxels will be disregarded. Voxels contributing to the weight thereby are enclosed in a sphere around the requested voxel, cf. Fig. 2. The weight assigned to each field of view is the sum of the proximity percentages from all voxels in the field.

The proportionator is compared with the traditional SURS using the aforementioned three biological examples, using a microscope system modified for stereology (detailed set-up is presented in Appendix A). In all examples, four independent estimates are obtained for each slide: two estimates using the proportionator and two using SURS. The relative variance between the two estimates in each repetition of the same method is used to provide an (coarse) indication of how accurate the method is. The number of fields of view observed, as well as the time spent on delineating the
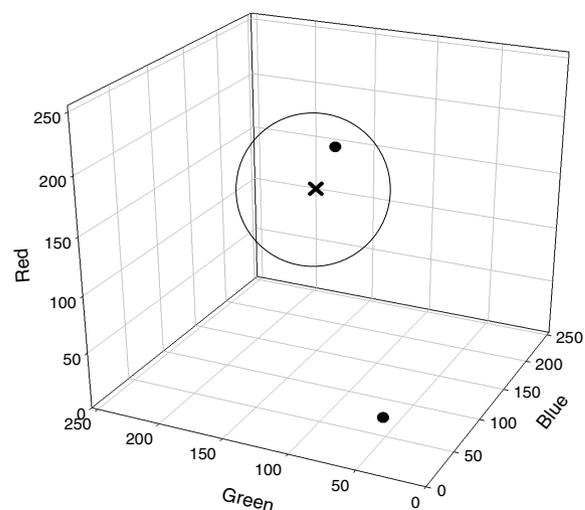

Fig. 2. The weight assignment. Voxels are mapped in 3D colour space of red, green and blue. X indicates the requested colour. ● indicates two examples of image voxels, the one inside the sphere contributes to the weight, and the other one is neglected.

region and assigning weights is recorded. A pilot study is performed for each example to adjust the sampling fractions necessary for obtaining approximately the same total counts using the proportionator and SURS. Those fractions remain constant throughout a biological example and the slides within an example, regardless of the total number of fields of view in each slide. In each slide, the region of interest is delineated independently four times and colour requests and weight assignments are performed independently for the two proportionator estimates. The ranges of weights differed between slides due to the difference in colour, staining and section artefacts.

Total number of granule cells in rat cerebellum

The estimation of total number of granule cells in rat cerebellum using the optical fractionator (West *et al.*, 1991) with a varying sampling fraction (Horvitz & Thompson, 1952; Dorph-Petersen *et al.*, 2000) was done on a systematic, uniformly random sample of sections from two normal rats. Following immersion fixation in 4% phosphate-buffered formaldehyde, the cerebellum was isolated and divided into halves. One random half was embedded isotropically in 5% agar using the isector (Nyengaard & Gundersen, 1992), embedded in glycolmethacrylate (Technovit 7100, Kulzer, Wehrheim, Germany), and cut exhaustively with a block advance of 40 μm. Every 24th section was taken by SURS and stained with a modified Giemsa stain (Larsen & Braendgaard, 1995), providing six sections for the first rat and eight sections for the second. The final screen magnification was 2800X using a 100X objective. The diameter or radius of the colour inclusion sphere was 20% (c.f. Fig. 2). The areas of the 2D unbiased counting frame (Gundersen, 1977) and the field of view were 418 μm^2 and 14000 μm^2, respectively. Step lengths in the x- and y-direction for SURS were 1864 μm and 1332 μm (providing a field of view sampling fraction of $5.64 \cdot 10^{-3}$) resulting in a total disector areal sampling fraction of $1.68 \cdot 10^{-4}$. With a sampling fraction of $1.09 \cdot 10^{-3}$ for the proportionator, the disector areal sampling fraction was $0.32 \cdot 10^{-4}$ (19% of SURS). The Q$^-$-weighted section thickness was 35 μm and the height of the optical fractionator was 25 μm. The estimator of the total number of cerebellar granule cells is

$$N(cells) := \frac{1}{SSF} \cdot \frac{1}{ASF} \cdot \frac{1}{HSF} \cdot 2 \cdot \sum Q^-, \qquad (3)$$

where the factor 2 is the inverse hemisphere sampling fraction, SSF is the section sampling fraction, ASF is the areal sampling fraction and HSF is height sampling fraction using the Q$^-$-weighted section thickness (Dorph-Petersen *et al.*, 2001):

$$\bar{t}_{Q^-} := \frac{\sum (t_i \cdot Q_i^-)}{\sum Q_i^-}. \qquad (4)$$

Generally speaking, this is an example of a semi-clustered distribution where the very irregular granule cell layer

constitutes roughly a $\frac{1}{4}$ to 1/3 of the organ, cf. Fig. 3. See Table 1 for results summary.

Total number of GFP orexin neurons in mice brain

Two brains were studied from mature transgenic mice, where orexin neurons in lateral hypothalamus and adjacent perifornical area could be visualized *in situ* by expression of enhanced green fluorescent protein (Burdakov *et al.*, 2006). Brains had been immersion fixed in 4% phosphate-buffered formaldehyde for a few hours, cryo-protected and frozen in liquid nitrogen. The brains were cut exhaustively using a cryomicrotome with a microtome advance of 80 μm and every second section was chosen by SURS. Eight and six sections were observed from the two brains. The total number of orexin neurons was estimated using fluorescence light and the optical fractionator. The final screen magnification was 1680× using a 60× objective. The colour inclusion sphere was 5% (c.f. Fig. 2). The area of the 2D unbiased counting frame was 18100 μm and the field of view area was 43200 μm^2. Step lengths in the x- and y-direction for SURS were 298 μm and 223 μm (field of view sampling fraction of 0.65) resulting in total area sampling fraction of 0.272. With a fields sampling fraction of 0.28 for the proportionator, the result was a total area sampling fraction of 0.117 (43% of SURS). The Q$^-$-weighted section thickness was 45 μm and the height of the optical fractionator was 35 μm. Total number was estimated as in the previous example. The orexin neurons have a mildly clustered distribution in the reference space. The example was selected in order to test the performance of the proportionator in a situation with a stain with a high and very varying unspecific 'staining' of the background, cf. Fig. 4.

Area of β cells in dog pancreas

Two arbitrarily sampled paraffin blocks from a dog pancreas were used for studying the estimation of absolute and relative area of insulin producing β cells. The pancreas had been perfusion fixed with 1% paraformaldehyde and 1% glutaraldehyde, it was cut into 3-mm thick complete cross-sections and embedded in paraffin (Kroustrup & Gundersen, 1983). A 3-μm-thick section was cut from each block and mounted on a Superfrost+ glass slide. Using an automatic stainer (Benchmark XT, Ventana, Illkirch Cedex, France), the β cells were stained with an insulin antibody (1:50 guinea pig anti-swine insulin, code A0564, DAKO, Glostrup, Denmark) and XT UltaView DAB. All cell nuclei were stained with Haematoxylin, cf. Fig. 5.

The final screen magnification was 1680X using a 60X objective. The colour inclusion sphere was 15% (c.f. Fig. 2). The total area of a field of view was 36300 μm^2. Step lengths in the x- and y-direction for SURS were 1580 μm and 1180 μm (field of view sampling fraction of 0.0196). The proportionator had an areal sampling fraction of 0.00383 (20% of SURS). The area per point (a/p) for the β cells was 386 μm and for the
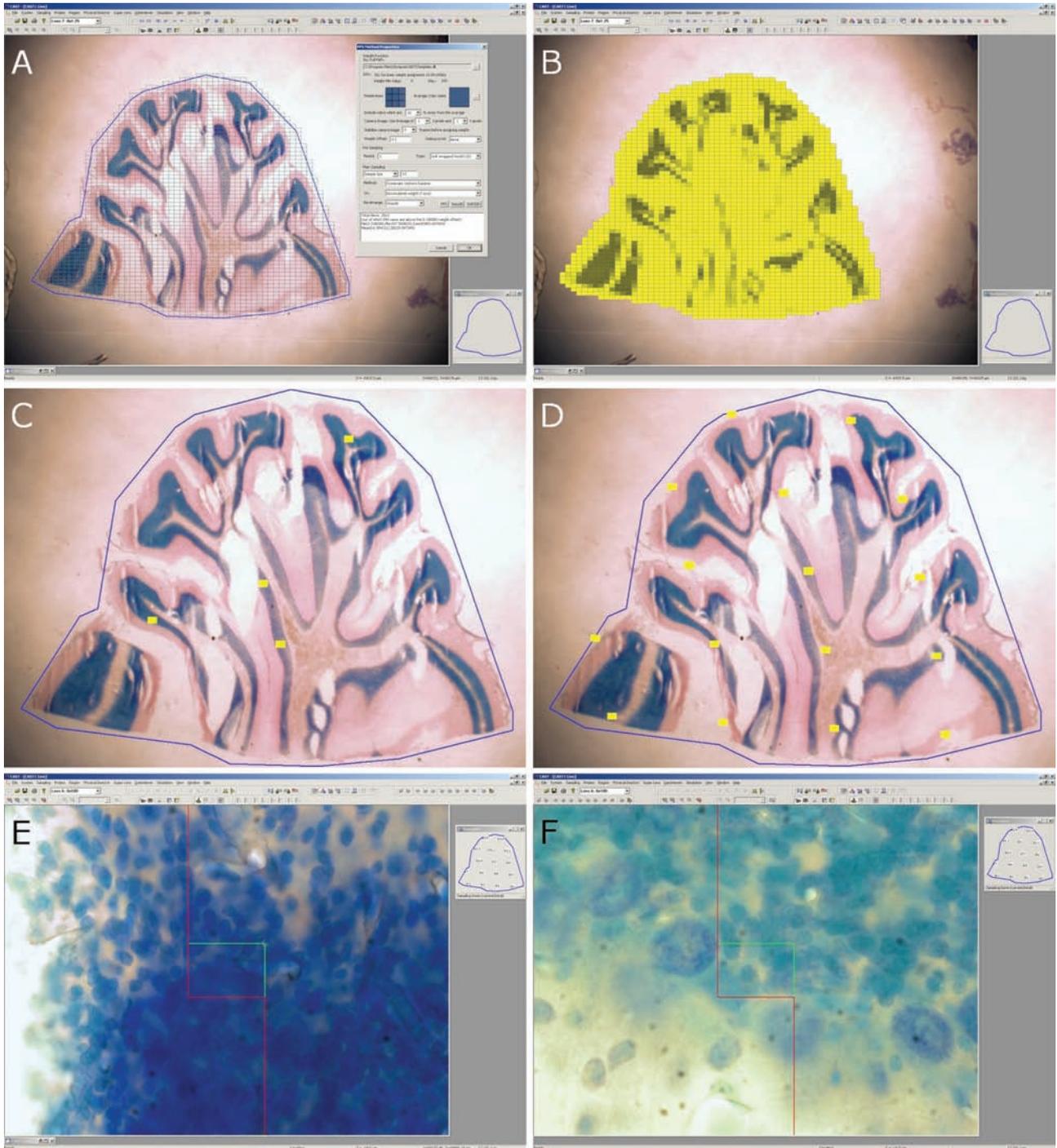
Fig. 3. Estimating total number of granule cells in rat cerebellum. The blue granule cell layer is clearly visible at 1.25X (panel A). The area of interest is delineated coarsely and partitioned into fields of view. Panel B shows the fields of view with their assigned weight on a grey-scale. Panel C shows the distribution of sampled fields (yellow rectangles) for the proportionator, the selected fields of view are almost surely in the granule cell layer. As shown in the panel D – sampling with the traditional SURS – such fields of view may or may not hit the blue region. Panels E and F are examples of counting at 100X magnification (oil lens).

Table 1. Summary of estimates and precision for the three biological examples. Values are means per animal (or per slide for pancreas). Values in brackets are proportionator percentage of SURS. 'Estimate *CE*' are the *CV*s of the replications. Proportionator with direct *CE* was run by splitting the proportionator sample into two independent samples, cf. text.

| SURS	Estimate	Estimate CE	Direct CE	Total count	Observed FOVs
Pancreas, tissue	53.4 mm^2	0.073		686	54
Pancreas, β cells	1.48 mm^2	0.078		76	54
GFP orexin neurons	1100	0.57		97	114
Granule cells	$2.01 \cdot 10^8$	0.63		255	115
Proportionator					
Pancreas, tissue	51.2 mm^2	0.038 (52%)	0.081 (111%)	188 (27%)	11 (20%)
Pancreas, β cells	1.52 mm^2	0.023 (30%)	0.066 (84%)	88 (116%)	11 (20%)
GFP orexin neurons	1152	0.14 (25%)	0.14 (24%)	98 (101%)	43 (38%)
Granule cells	$1.62 \cdot 10^8$	0.18 (29%)	0.14 (22%)	303 (109%)	31 (27%)
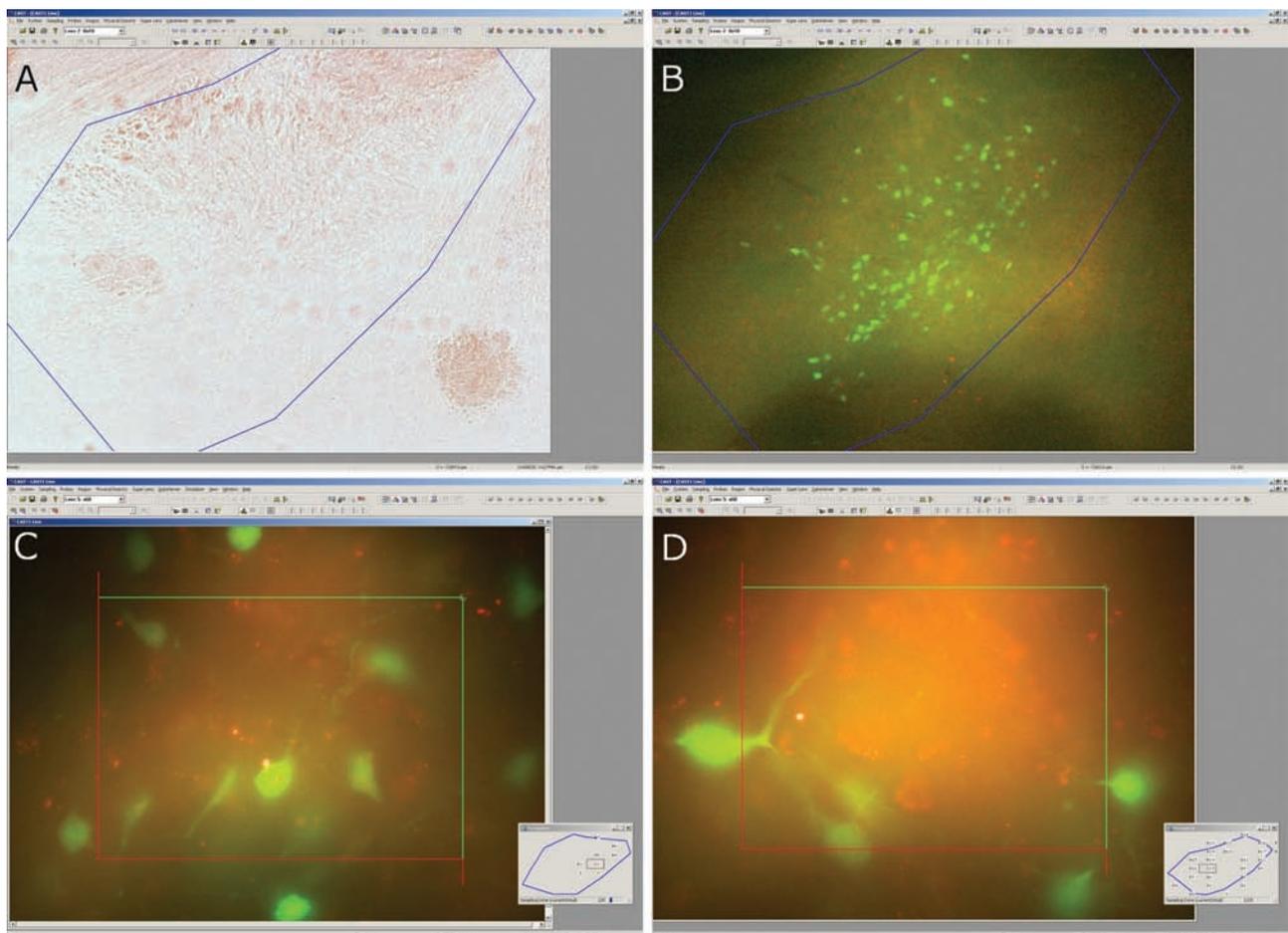
Fig. 4. Estimating total number of GFP-expressing orexin neurons in transgenic mice brain. Panels A and B show the same region of interest at 10× objective magnification in bright field (panel A) and during colour identification using fluorescence light (panel B). Note the greenish background noise. Counting is performed using a 60× oil objective using the optical disector, as shown in panels C and D. The small inserts indicate the positions of the sampled fields.
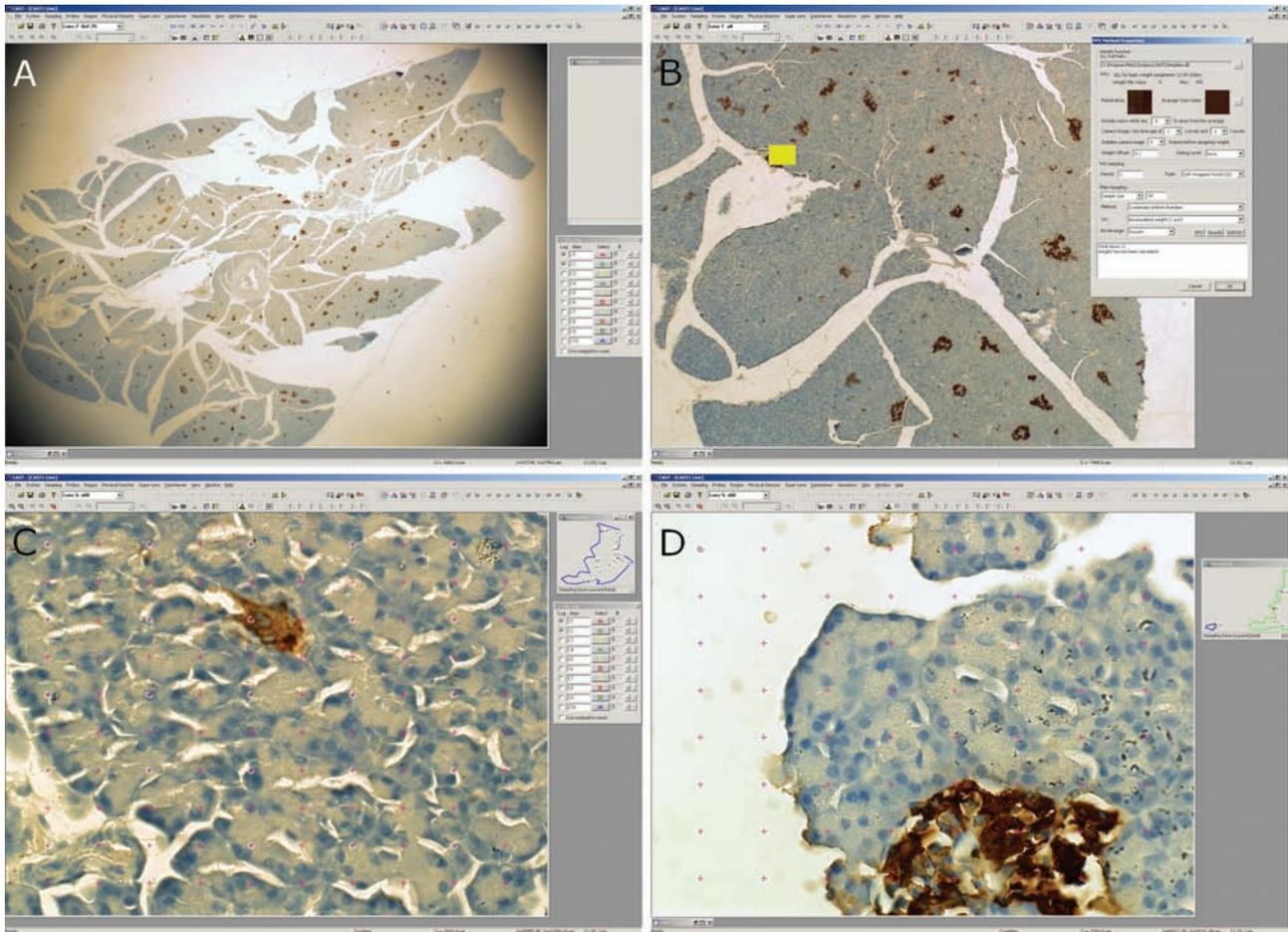
Fig. 5. Estimating area of β cells and total tissue in dog pancreas. Panel A shows the delineated area using a $1.25\times$ objective. Note the sparse but quite uniformly distributed islands of β cells. In panel B, the brown β cell colour is identified at $4\times$ objective magnification and weights are assigned (note the size of a selected field of view marked as a yellow rectangle). Images from the sampled fields of view with a $60\times$ objective and the point grid probes are shown at panels C and D.

containing tissue 1540 μm; both counts were performed in the same fields, which, for the proportionator, were selected based on the amount of insulin-stain. The estimator equation for the total area of β cell in each section is:

$$A(\beta \text{ cell}) := \text{Total}\,[P(\beta \text{ cell})] \cdot (a/p). \qquad (5)$$

These estimates are all what is needed if (a sufficiently large sample of) parallel sections are sampled uniformly with a constant separation, T. The total volume of β cells in the pancreas is then obtained by the Cavalieri-estimator:

$$V(\beta \text{ cell}) := T \cdot \Sigma A(\beta \text{ cell}). \qquad (6)$$

In the (unlikely) situation that the sections are uniformly sampled with unknown or varying distances, one would have to use the classical volume fraction estimator

$$V_V(\beta \text{ cell/tissue}) := \text{Total}\,[P(\beta \text{ cell})]/\text{Total}\,[P(\text{tissue})], (7)$$

which requires the additional counting of points hitting the reference space and an independent estimate of the total

pancreatic volume, V(tissue), to obtain the total volume of β cells in the pancreas:

$$V(\beta \text{ cell}) := V_V(\beta \text{ cell/tissue}) \cdot V(\text{tissue}). \qquad (8)$$

Both estimators require determination of various dimensional aspects of shrinkage for the total volume of β cells to be unbiased. The pancreatic β cells are an example of a roughly homogeneous distribution of small and sparse events, their volume fraction is only ~0.027.

Results

The first example studied was the estimation of total number of granule cells in rat cerebellum. The distinct blue stain of the granule cell layer was clearly visible with bright field 1.25X objective and made the colour identification and weight assignment fast and reliable (Fig. 3). The small area sampling fraction of the counting frame (3%) with a 100X objective made the identification of an empty
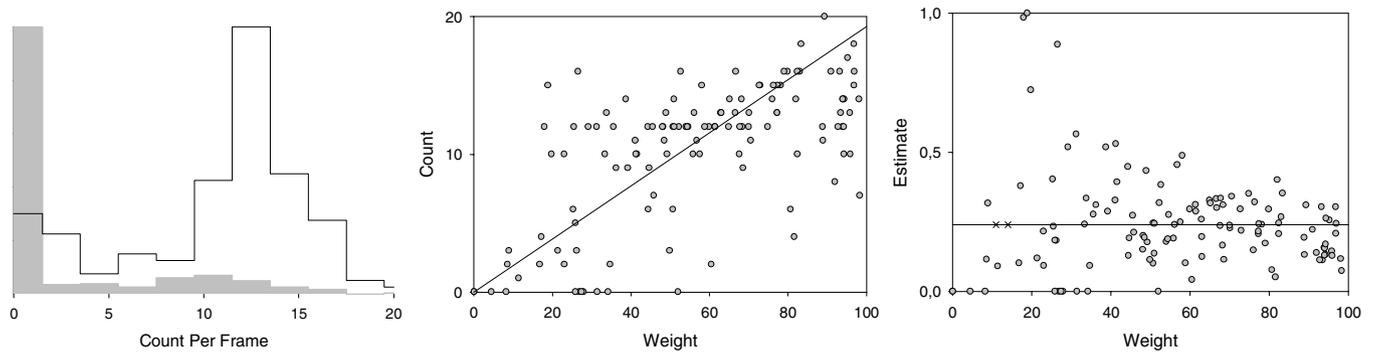
Fig. 6. The distribution of individual samples and the bivariate sampling distributions when estimating the total number of granule cells. Panel A shows the distributions of the correct counts per disector. The grey histogram is the cell counts in SURS samples, whereas the full drawn histogram is for the proportionator (the distributions are normalized to the same mode, to allow viewing the moderate variation within the counts per frame in the proportionator). Panel B shows the correct count and weight for all fields sampled with the proportionator. The contribution from each field to the total estimate is proportional to the slope of a line from origin to the data point. Panel C is similar to B, showing the estimates instead of counts (the ordinate is fraction of maximal estimate), the horizontal line is the average estimate. The slope corresponding to that average estimate is the slope of the line shown in panel B. The *CE* of the proportionator is the *CE* of the slopes around this slope. For the proportionator, the variability of the counts themselves is therefore irrelevant (it is insensitive to field-to-field variation).

field a fast process. Figure 6 shows the count distribution, as well as the correlation between weights, counts and estimates.

As illustrated in Fig. 6, panel A, the proportionator samples fields with a much higher average count than SURS (9.8 vs. 2.2), and one therefore needs only to study about $\frac{1}{4}$ of the number of fields necessary for SURS. Moreover, the *CV* of the proportionator estimates from each field (Fig. 6, panel C) is much lower than that among SURS fields (the grey distribution in Fig. 6, panel A): 0.24 vs. 0.63. Despite the lower number of fields studied, the statistical efficiency of the proportionator (roughly $1/CE^2$) is therefore much greater than that of SURS: ~17 vs. 2.5, cf. also Table 1, which shows the summary of results with regards to estimates and precision.

As described in Gardi *et al.* (2007), it is possible to estimate the *CE* of the proportionator estimate for each section by taking two independent samples of size *n/2* instead of one sample of size *n*. In ordinary practice, it is more useful to think of a direct estimation of the variance $Var_i(n) = (CE_i(n).n_i)^2$ of the estimate of the total number of particles n_i in the *i*th section. For the estimate of total number in all *m* independent sections, Σn, one may then compute the overall **direct CE**:

$$CE_m \left(\sum_i^m n_i \right) = \frac{\sqrt{\sum_i^m Var(n_i)}}{\sum_i^m n_i}. \qquad (9)$$

The same strategy does not work for SURS because SURS sampled FOVs are dependent.

The poor precision of the SURS estimator of total number of both granule cells and GFP orexin neurons is indicative of the section inhomogeneity or field-to-field variation. Note, however, that there are only a total of two replications to

provide the estimated *CE*s and also the direct *CE* estimates are the result of two comparisons of two independent samples – and of half-size. Consequently, all *CE* estimates in Table 1 are rather imprecise. In ordinary practice, one would average the direct *CE* estimates over all *m* animals in each group,

$$\overline{CE} := \sqrt{\frac{\sum_i^m CE_i^2}{m}} \qquad (10)$$

and thereby obtain much more useful estimates.

It is necessary to emphasize that the sampling design for each of the three examples is made to ensure that the results of SURS and proportionator sampling and estimation are as *comparable* as possible. Since the counting noise is roughly proportional to ΣQ^-, the number of fields studied are adjusted (in the pilot study) to provide roughly the same total count. The frame size and the disector height are the same for both sampling strategies. It follows that *none of the sampling designs are optimized* for the corresponding strategy.

As an example, the statistical efficiency of SURS might be improved by sampling about three times more fields. The counting frame could then be reduced to two maximally separated frames per field with a combined area of 1/6 of the present frame. The 6-fold higher number of frames examined would considerably reduce the impact of tissue inhomogeneity on SURS precision (the various ways of optimizing the proportionator are discussed below).

The example of optimizing a strategy at the expense of studying more fields underlines the importance of taking the time spent per animal into account when trying to make realistic comparison of sampling strategies. If *T* is the time spent per animal, the relative efficiency (time and precision) of

Table 2. Average time spent for counting one animal (or one slide in the pancreas example). Values in brackets are for the proportionator in percentage of SURS. The last column is computed using Eq. 11, below. The **CE²** used is a weighted average of **EstCE²** and **DirCE²**: **CE² := (2∗EstCE² + DirCE²)/3**, taking into account the lower number of observations for the Direct **CE** estimates.

SURS	Total counting time (min)	No of slides	Overhead time per slide (min)	Total time (min), T	Relative efficiency of proportionator compared to SURS
Pancreas, tissue	19:41	1	1:00	20:41	
Pancreas, β cells	19:41	1	1:00	20:41	
GFP orexin neurons	16:45	7	1:00	23:45	
Cerebellar granule cells	41:37	7	1:00	48:38	
Proportionator					
Pancreas, tissue	4:19	1	5:00	9:19 (45%)	8×
Pancreas, β cells	4:19	1	5:00	9:19 (45%)	25×
GFP orexin neurons	8:44	7	5:00	43:44 (184%)	9×
Cerebellar granule cells	10:06	7	2:00	24:06 (50%)	24×

the proportionator compared to SURS is

Relative Proportionator Efficiency

$$= \frac{CE_{SURS}^2 \times T_{SURS}}{CE_{Proportionator}^2 \times T_{Proportionator}}. \qquad (11)$$

Table 2 shows the total time spent on each of the examples. Since the poor SURS estimator already took twice as much time as the proportionator, SURS is clearly never going to be as efficient as the proportionator for cerebellar granule cell counting.

The second example was the estimation of the total number of orexin neurons in transgenic mice brain. Weight assignment for the GFP-expressing orexin neurons was only possible with a 10X objective under fluorescence light (Fig. 4). The noisy and greenish image contributed to non-trivial colour identification and weight assignment. The requested colour had to be fine tuned, and the maximum distance in 3D space (Fig. 2) had to be carefully adjusted to avoid picking up the background noise. The region of interest was small, and the number of total fields of view was not more than 40 per slide. The large sample size for both the proportionator

and SURS (28% and 65%, respectively) made the colour identification and weight assignment (for each slide) the most time-consuming operation in the proportionator (Table 2) and made the proportionator non-beneficial with regards to time as compared to the traditional SURS, cf. Table 2. The individual samples and weights are shown in Fig. 7.

The pronounced inhomogeneity did, however, make SURS both very inefficient and quite time consuming, so in the comparison in this really difficult example the proportionator came out about eight times more efficient, solely because of a much better statistical efficiency. The genuine efficiency of the combination of sampling proportional to weight and then estimating *inversely* proportional to it is highlighted in this example. Even if it takes twice the time to accomplish this, it is much more efficient thereby to avoid the inhomogeneity w. r. t. numerical density than the brute force counting of really many fields.

The third and last example is the estimation of the area of β cells in dog pancreas using point counting. The β cells were clearly visible as dark brown colour with a 1.25X objective. Due to camera artefacts at the tissue edges, which were also visible as a shade of brown, the region was delineated with a 1.25X
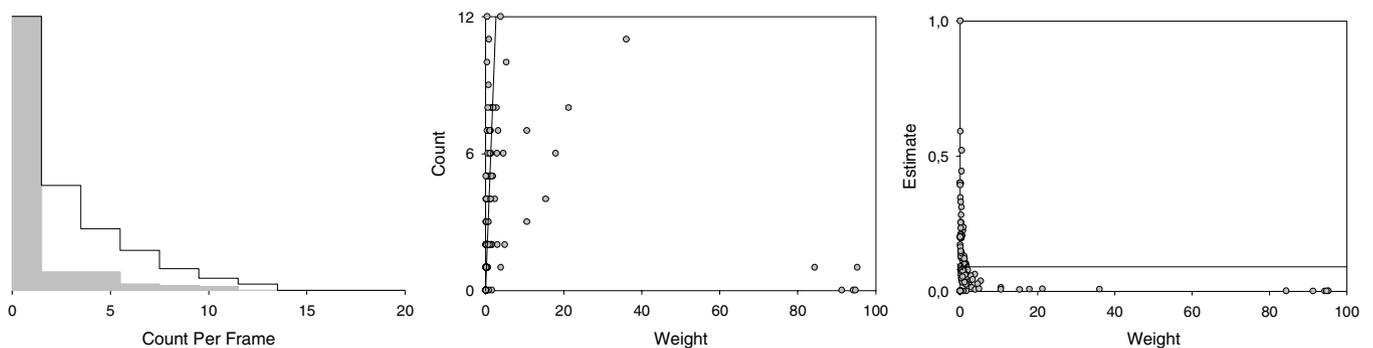

Fig. 7. Individual samples and weights when estimating total number of GFP-expressing orexin neurons. As shown in panel B, the example is clearly characterized both by fields with spuriously very large weights (with very low counts) and large counts in fields of low weight. Panel descriptions are as in Fig. 6.
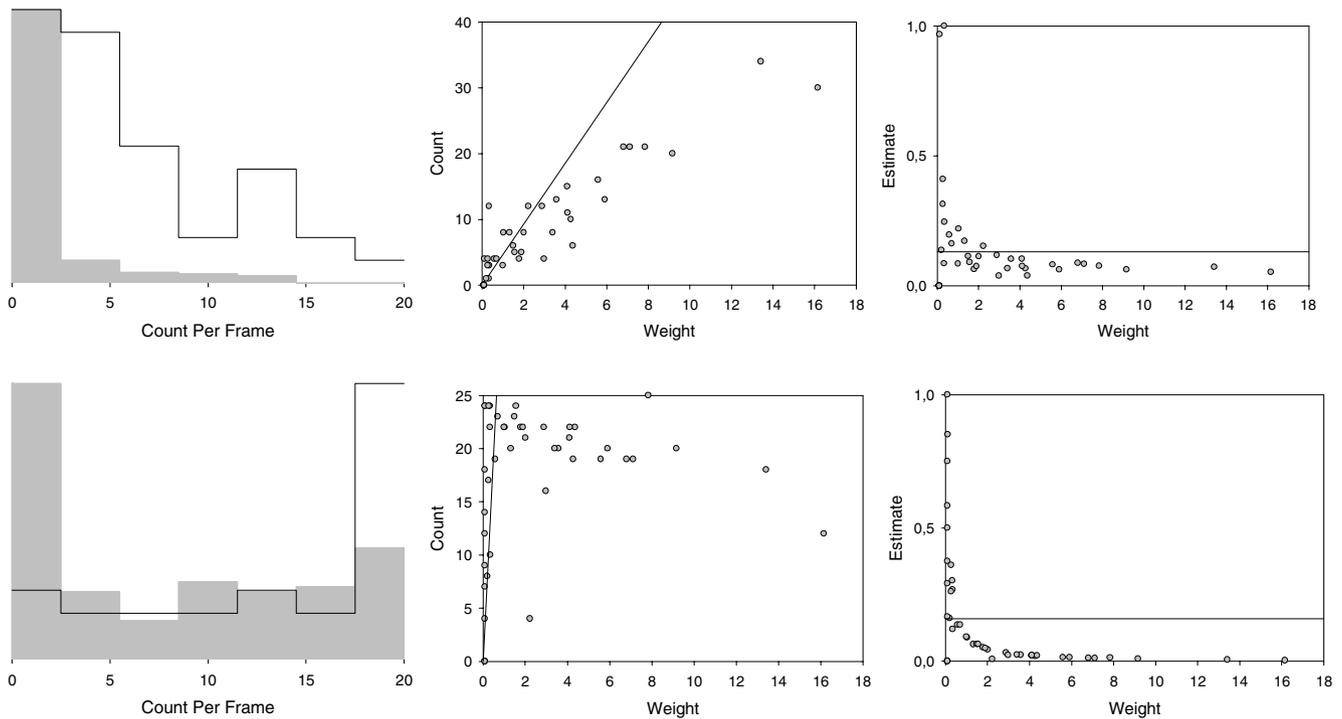
Fig. 8. Individual samples and the bivariate sampling distributions when estimating area of β cells (panels A1, B1, C1) and containing tissue in dog pancreas (panels A2, B2, C2). Panel C1 clearly indicates a special problem in sampling small and sparse events with the proportionator: non-zero counts may occur in fields of low weight and they provide very high estimates (upper two data points to the left), decreasing the precision. The sampling of total tissue (panels A2, B2, C2) is performed using the weights of the insulin-stain and the count-weight correlation is therefore very poor. Panels descriptions as in Fig. 6.

objective, but the colour identification and weight assignment was done with a 4X objective (Fig. 5). The number of fields of view was approximately 3000 (calculated to fill up the entire view at 60X), which led to a precise but time-consuming weight assignment operation at 4X. Identifying the colour was fast, but the actual stage movement for observing the total number of fields of view led to approximately 3 minutes of stage movement (Table 2).

Figure 8 shows the statistical characteristics of the counts, weights and estimates. The β cells are a typical case of proportionator performance in detecting relatively sparse events: it avoids very well the fields with low counts and focus on fields of a high count. The occasional positive count in a field of very low weight, providing extreme estimates and reducing precision, were too rare to really offset the efficiency which was roughly 25 times better than that of SURS, about equally due to better precision and faster performance.

The estimation of total tissue area was based on sampling of the β cell stain (and it is unnecessary for estimating total β cell volume, as outlined above). Because of the rather homogeneous distribution of islets in the sections, a large total amount of β cell stain, i.e. a large weight, provides proportionator sampled fields with large tissue areas as indicated in the first panel in the lower row of Fig. 8. The count-weight association is very weak, however, and the estimate

is not very precise. The procedure is fast, however, and the combined efficiency is clearly better that of SURS.

Discussion

This is the first study of the performance of the proportionator in real examples, and an obviously preliminary one. The main purpose was to get some experience with implementing the novel sampling mechanism; the imprecise estimates of efficiency were only of secondary importance (hence the low number of animals studied). The efficiency estimates did, however, provide encouragement for the continued work with this radically different sampling and estimation paradigm for quantitative microscopy.

With respect to the estimated efficiencies all examples indicate that the proportionator is much more efficient than SURS. However, the estimates of efficiency are not very precise and the examples are all inhomogeneous at various scales, so the above conclusion may not be valid in many cases of general interest. The combined efficiency nevertheless turned out, somewhat surprisingly, to be very robust against poor count-weight associations as witnessed by the GFP and pancreatic tissue examples.

The proportionator is unique among the efficient sampling strategies in that it allows the real precision, the direct *CE* in

Table 1, to be estimated unbiasedly—and at no extra cost to or effort of the user (Gardi *et al.*, 2007). There are several reasons why this is a very large advantage:

1. The estimator imprecision due to field-to-field variation (the tissue inhomogeneity) is not predictable using the current statistical predictors (Kieu *et al.*, 1999) of the *CE* (because fields are sampled systematically, and predicting the precision of that in 2D sections is mathematically difficult). As amply illustrated in Gardi *et al.* (2007), in inhomogeneous tissue the real *CE* may be several-fold larger than the (incompletely) predicted one.

2. For number estimation in very homogeneous tissue, the *CE* of the proportionator is *lower* than the counting noise, which is generally $CE_{noise} = 1/\sqrt{count}$, implying that one has to count 100 cells for this part of the *CE* to be ∼0.1. To take advantage of the higher precision of the proportionator, the dedicated researcher would like to know precisely what the precision *is* in the sections before counting 70 cells in 16 fields instead of 100 cells in 24 SURS fields (both with a *CE* of 0.1 in large sections of homogeneous tissue).

3. The proportionator, correctly performed, is guaranteed to be unbiased, irrespective of the count-weight correlation. This correlation may, however, be unexpectedly weak or even negative (or something may go wrong with the automatic weight assignment) and it is a comfortable safeguard that this will then be reflected in a (unexpectedly) high *CE*, which is shown on the monitor right after counting in the last field.

The many practical problems encountered in the wide range of examples selected for this study allow us to identify a number of features that may improve the efficiency of proportionator sampling and estimation. A number of these were anticipated, but we wanted to keep the technical set-up and the software as simple as possible at this stage, considering that the starting point was the software developed for the primitive simulation study.

It is characteristic of the proportionator that it is insensitive to ordinary field-to-field variation; the section inhomogeneity with respect to the feature under study becomes a signal rather than a noise. On the other hand, any noise in the count-weight correlation, shown in Figs 6–8, decreases precision. The specific characteristic of the tissue and the stain, including antibodies, give rise to many kinds of 'noise', discussed extensively in Gardi *et al.* (2007). However, many of the sources of noise are of a technical character and are potentially preventable or may be overcome.

Microscope and optics

The microscope and optics are much more intimately integrated in the proportionator procedure than in ordinary microscopy. For optimal performance a number of features are important:

1. At low magnification, the illumination of the section is very uneven, cf. Fig. 3. This is relatively easily removed by incorporating simple image analysis algorithms (Bischof *et al.*, 2004; Osadchy & Keren, 2004) in the procedure.

2. Various diffraction phenomena may occur at the edges of the section at very low magnification. To minimize such problems, it is necessary to have a range of low magnification objectives to choose from. Depending on the manufacturer of the microscope, 1X, 1.25X, 1.6X, 2X, 4X, 6X, 10X and 15X are often available.

3. Almost all inhomogeneous organs also show section-to-section variation. The obvious way of turning also this noise into a signal is to make the weight assignment and sampling on the whole set of sections in one run. That requires that the microscope is equipped with a multi-slide stage, usually these accommodate eight sections, which in most cases would be enough for one animal (if not, one should analyze every second section in one series and the others in another one, that essentially eliminates their variability (Gundersen, 2002). When applicable, this stratagem alone may increase the efficiency of the proportionator manifold. For optimal efficiency, it is necessary that the staining intensity and the section thickness are roughly constant among sections.

4. If the total tissue area becomes large at the scale of the final magnification, one may use SURS subsampling of FOVs before weight assignment and proportionator sampling. Once the number of FOVs becomes larger than a few thousand, efficiency is unlikely to improve if many more FOVs are sampled, the obvious exception being the analysis of rare events.

Low magnification scanning

The low magnification scanning of all FOVs is also critical for proportionator efficiency and may be optimized in several ways:

1. The initial indication of the position of the section on the slide may be performed by the fast dragging of a rectangle.

2. In many cases, the area of interest is the whole section and no further delineation is necessary since all empty FOVs between the section and the above, outer rectangle will not produce the requested specific signal (obtaining a weight of 0) and they are therefore automatically eliminated in the sampling proportional to weight.

3. Automatic detection of the section boundary (Sahoo *et al.*, 1988; Skarbek & Koschan, 1994; Wang *et al.*, 2006; Jonasson *et al.*, 2007; Gomez *et al.*, 2007) may in some cases be an alternative to the above.

4. Some modern stereological software systems like NewCast® (VisioPharm Hørsholm, Denmark) already make a fast scan at low magnification and present the composite image of the whole section on the monitor. When the area of interest

is just a (small) part of the section, the necessary manual delineation by the expert user may now often be performed on this so-called 'SuperLens' image, generally much faster than the interactive delineation used by most software. If the initial scan was performed at a sufficient resolution, the information for the weight assignment is already available, which will further reduce the time spent on setting up each section.

Field of views portioning

The partitioning of the section into FOVs should also be considered. The 'catchment area' for collecting the weight information may in many cases be different from the precise FOV at high magnification. It should be as large as the stereological test system and often also include a guard area around it, both defined by the user in the pilot study. Additionally, it may be necessary to increase this area to allow for imprecision in the translation of section co-ordinates between very different magnifications (sensitive to the so-called parcentering of the lenses in question). As illustrated in Fig. 3, this feature would have improved the estimator considerably in the analysis of the rat cerebellum. The frame area was only 3% of the FOV and most of the data points below the line in Fig. 6, panel B, owe their relatively low count to the fact that the weight for the total FOV poorly matched the count in a very small area of the FOV.

Colour pickup

The indication of the requested colour may easily be optimized in several ways, since we just used primitive pointing at a characteristic pixel:

1. The range of colours actually represented in the section by the many instances of the structure under study should be taken into account.
2. The box enclosing these colours in colour space may automatically be enlarged by a preset amount to make the weight assignment sufficiently sensitive and still specific.
3. The colours indicated in the first section may often be used on all following ones, particularly if the above box is large enough.
4. The weight assignment is faster if all image pixels that happen to be inside the box are given a weight of 1, all others are disregarded (implicitly given a weight of 0). The weight of the FOV is then simply the sum of weights of all pixels in it.
5. The fine-tuning of the colour selection should be interactive using a stored image with indication of all pixels of weight 1.
6. The pancreas tissue example is particular in that is does not have a single characteristic colour. Inspection of the high magnification images in Fig. 5 clearly indicates that pancreas tissue, due to counter staining, possesses a large range of colours – including shades of brown similar to

the insulin antibody stain. In such cases, the requested 'colour' should be pointed out by dragging rectangles over the tissue, proving a rather large box enclosing all requested colours. The stain, specific for a particular phase, the insulin stain in the example, is then indicated separately (it is likely already stored in a file which is simply reused). The weight assignment for the pancreas tissue may now be a fast Boolean procedure: A weight of 1 is assigned to all pixels with a colour in the large box, unless they are in the 'insulin-box' and therefore disregarded.

7. The assignment of weight based on pixel colours may most likely be more sensitive and specific if colours are represented in a space of hue, saturation and intensity (Pydipati *et al.*, 2006) instead of the primitive red-green-blue representation used here.

Detecting requested features

The use of other automatically detected features than colour is a very promising area of research. The proportionator principle has just three requirements:

1. the weight assignment must be automatic,
2. it must produce a number in the range 0 to a suitable maximum for each FOV,
3. and the weight should have a positive correlation to the correct signal to be provided in sampled FOVs.

We have just used colour of individual pixels because it is very simple to implement (and fits into the general strategy that stains are used to positively enhance the structure under study). However, even now there is an enormous amount of features of an image that may be extracted automatically and quantitated (Gonzalez & Woods, 2002) – and they may all be used for proportionator sampling and estimation, as discussed in Gardi *et al.* (2007). As an example, if the distinct *edges* of the granular cell layer in the cerebellum are detected and given a weight according to their amount in each FOV, the sampling of Purkinje cells would be greatly facilitated. However, if all FOVs are at the edges of the granular cell layer, every second will still not contain Purkinje cells. Purkinje cells are only at the edge, which in the direction of the blue-not blue gradient is neighbouring the cerebellar surface, the natural section edge – and that neighbour-relation is evidently also an automatically detectable feature. Although outside the scope of this paper, also the correct signal from sampled fields is in no way restricted to just a stereological count, it may be *any correct signal* from the FOV, including calibrated photometry or structure mass provided by acoustic microscopy, just to mention a few examples.

Weight values

Editing the weights is in practice often an advantage. The map of weights as shown in Fig. 3, panel B, makes it rather easy to detect problems. Editing may take many forms:

1. The user may simply indicate that certain fields must be ignored because of technical problems like folds or localized staining artefacts, for example (this is not, however, compatible with an unbiased estimate).
2. The weights are just numbers, and almost surely biased ones w. r. t. the structure under study. All kinds of transformations that preserve the positive correlation to the count (and do not exclude any field with structure) are allowed.
3. One may add a constant to all weights to avoid that FOVs with a very small weights provide a positive count. If this happens too often it is likely to decrease precision markedly because these estimates are very high (the contribution to the total from each field is proportional to the count *divided* by the weight).
4. If the count-weight correlation, illustrated in Figs 6–8, is best represented by a monotonous curve and not a straight line one may transform all weights to $\sqrt{weight}$ or use any other similar mathematical transform that rectifies the correlation.
5. If just colour is used for weight assignment, as in this paper, the count-weight correlation for Purkinje cells, at the *edge* of the blue, detected granular cell layer, may be biphasic: after a maximum at a moderate weight, **Wm**, it decreases towards higher weights. This might be remedied by penalizing all weights larger than **Wm** by assigning final weights as **Weight' := weight − Wm**.
6. Unspecific staining of the physical edges of the section is sometimes encountered even with highly specific antibodies. Automatic detection of section edges, discussed above, may be followed by reducing the weight of all pixels in a rim near the edge to some small weight (it must be larger than 0 to preserve unbiasedness, the rim could of course also contain the specific structure under study).

To summarize all of the above, this is the first report of the real performance of the proportionator, but we do not anticipate it to be the last. Biological research has for decades profited enormously from the availability of very specific markers for proteins or peptides or gene sequences or products of specific expressions etc. (as it has from less specific chemical stains for a century). With the proportionator we strongly believe that this quantization will be much more efficient and thereby in itself promote the widespread use of reliable stereological procedures.

Appendix A – Hardware set-up

The system used was an Olympus BX50F-3 microscope (Japan), with a motorized stage manufactured by Prior Scientific Instruments model H101BX and Joystick Prior model H152EF both connected to Prior controller box H128V3 (Cambridge, England), which connects via a serial port to a computer. A Heidenhaim microcator model ND 281 (Traunreut, Germany) was connected via a serial port as well. An Olympus 100W high-pressure mercury burner USIHIO BH2-RFL-T3 with lamp USH-102D model U-ULS100HG was transmitting the fluorescence light. The fluorescence filter used (if applicable) was 'pe U-N51006 F/TR C59531'. When normal bright field light was needed, the Olympus halogen lamp JC12V 100W HAL-L U-LH100 was applied. An Olympus DP70 digital camera with a 1.45 million pixel CCD coupled with pixel-shifting technology resulting in images with a resolution of 4080 × 3072 pixels was connected to the computer via a dedicated PCI bus card. The following Olympus lenses were used: UPlanApo 1.25x/NA 0.04, UPlanFl 4x/NA 0.13, UPlanApo 10x/NA 0.40, UPlanFL 40x/NA 0.75, UPlanApo 60x/NA 1.4 oil, UPlanApo 100x/NA 1.35 oil. The computer used was a mixed brand, Intel based, running a specially upgraded Computer Aided Stereology Tool (CAST software - Visiopharm, Hørsholm, Denmark) based on a code branch that was made on 10 November 2003 from the original Olympus CAST code version 2.3.0.2.

Acknowledgements

References

Bischof, H., Wildenauer, H. & Leonardis, A. (2004) Illumination insensitive recognition using eigenspaces. *Comput. Vision Image Und.* **95**, 86–104.

Burdakov, D., Jensen, L. T., Alexopoulos, H. *et al.* (2006) Tandem-pore K+ channels mediate inhibition of orexin neurons by glucose. *Neuron* **50**, 711–722.

Dorph-Petersen, K. A., Gundersen, H. J. G. & Jensen, E. B. (2000) Non-uniform systematic sampling in stereology. *J. Microsc.* **200**, 148–157.

Dorph-Petersen, K. A., Nyengaard, J. R. & Gundersen, H. J. G. (2001) Tissue shrinkage and unbiased stereological estimation of particle number and size. *J. Microsc.* **204**, 232–246.

Gardi, J. E., Nyengaard, J. R. & Gundersen, H. J. G. (2006) Using biased image analysis for improving unbiased stereological number estimation - a pilot simulation study of the smooth fractionator. *J. Microsc.* **222**, 242–250.

Gardi, J. E., Nyengaard, J. R. & Gundersen, H. J. G. (2007) The proportionator: unbiased stereological estimation using biased automatic image analysis and non-uniform probability proportional to size sampling. *Comp. Bio. Med.* **38**, 313–328.

Gomez, D., Montero, J., Yanez, J. & Poidomani, C. (2007) A graph coloring approach for image segmentation. *Omega* **35**, 173–183.

Gonzalez, R. C. & Woods, R. E. (2002) *Digital Image Processing*. Prentice-Hall, Upper Saddle River, New Jersey.

Gundersen, H. J. G. (1977) Notes on the estimation of the numerical density of arbitrary particles: the edge effect. *J. Microsc.* **111**, 219–233.

Gundersen, H. J. G. (2002) The smooth fractionator. *J. Microsc.* **207**, 191–210.

Gundersen, H. J. G. & Jensen, E. B. (1987) The efficiency of systematic sampling in stereology and its prediction. *J. Microsc.* **147**, 229–263.

Gundersen, H. J. G., Jensen, E. B., Kieu, K. & Nielsen, J. (1999) The efficiency of systematic sampling in stereology - reconsidered. *J. Microsc.* **193**, 199–211.

Hansen, M. M. & Hurwitz, W. N. (1943) On the theory of sampling from finite populations. *Ann. Math. Statist.* **14**, 333–362.

Horvitz, D. G. & Thompson, D. J. (1952) A generalization of sampling without replacement from a finite universe. *J. Am. Statist. Assoc.* **47**, 663–685.

Jonasson, L., Hagmann, P., Pollo, C., Bresson, X., Richero W. C., Meuli, R. & Thiran, J. P. (2007) A level set method for segmentation of the thalamus and its nuclei in DT-MRI. *Signal Proc.* **87**, 309–321.

Kieu, K., Souchet, S. & Istas, J. (1999) Precision of systematic sampling and transitive methods. *J. Statist. Plan. Infer.* **77**, 263–279.

Kroustrup, J. P. & Gundersen, H. J. G. (1983) Sampling problems in an heterogeneous organ: quantitation of relative and total volume of pancreatic islets by light microscopy. *J. Microsc.* **132**, 43–55.

Larsen, J. O. & Braendgaard, H. (1995) Structural preservation of cerebellar granule cells following neurointoxication with methyl mercury: a stereological study of the rat cerebellum. *Acta Neuropathol. (Berl)* **90**, 251–256.

Nyengaard, J. R. & Gundersen, H. J. G. (1992) The isector: a simple and direct method for generating isotropic, uniform random sections from small specimens. *J. Microsc.* **165**, 427–431.

Osadchy, M. & Keren, D. (2004) Efficient detection under varying illumination conditions and image plane rotations. *Comput. Vision Image Und.* **93**, 245–259.

Pydipati, R., Burks, T. F. & Lee, W. S. (2006) Identification of citrus disease using color texture features and discriminant analysis. *Comp. Elec. Agr.* **52**, 49–59.

Sahoo, P. K., Soltani, S. & Wong, A. K. C. (1988) A survey of thresholding techniques. *Comp. Vision Graph. Image Proc.* **41**, 233–260.

Skarbek, W. & Koschan, A. (1994) Colour image segmentation - a survey. *Tech. Bericht Tech. Uni. Berlin* 32–94.

Wang, Y. G., Yang, J. & Chang, Y. C. (2006) Color-texture image segmentation by integrating directional operators into JSEG method. *Pat. Rec. Let.* **27**, 1983–1990.

West, M. J., Slomianka, L. & Gundersen, H. J. G. (1991) Unbiased stereological estimation of the total number of neurons in thesubdivisions of the rat hippocampus using the optical fractionator. *Anat. Rec.* **231**, 482–497.