Image Analysis of Representative Food Structures: Application of the Bootstrap Method

Cristian Ramírez, Juan C. Germain, and José M. Aguilera

ABSTRACT: Images (for example, photomicrographs) are routinely used as qualitative evidence of the microstructure of foods. In quantitative image analysis it is important to estimate the area (or volume) to be sampled, the field of view, and the resolution. The bootstrap method is proposed to estimate the size of the sampling area as a function of the coefficient of variation (CV_{Rn}) and standard error (SE_{Rn}) of the bootstrap taking sub-areas of different sizes. The bootstrap method was applied to simulated and real structures (apple tissue). For simulated structures, 10 computer-generated images were constructed containing 225 black circles (elements) and different coefficient of variation (CV_{image}). For apple tissue, 8 images of apple tissue containing cellular cavities with different CV_{image} were analyzed. Results confirmed that for simulated and real structures, increasing the size of the sampling area decreased the CV_{Rn} and SE_{Rn}. Furthermore, there was a linear relationship between the CV_{image} and CV_{Rn}. For example, to obtain a CV_{Rn} = 0.10 in an image with CV_{image} = 0.60, a sampling area of 400 × 400 pixels (11% of whole image) was required, whereas if CV_{image} = 1.46, a sampling area of 1000 × 100 pixels (69% of whole image) became necessary. This suggests that a large-size dispersion of element sizes in an image requires increasingly larger sampling areas or a larger number of images.

Keywords: image analysis, microscopy, microstructure, representative elementary area, sampling, size distribution

Introduction

It is now well accepted that some relevant food properties are related to their structure and that key elements defining the architecture of this structure are not discernible by the naked eye. Thus, the use of several microscopy techniques and other imaging methods has become common in the study of food microstructures (Kaláb and others 1995; Aguilera 2005; Martin and Sagalowicz 2008). It follows that images (photomicrographs) are extensively used in the scientific literature as evidence of the structural features of foods (Aguilera 2005). These images are usually used as support for a qualitative description of the microstructure and supposedly complement data generated by rigorous protocols of sampling and preparation steps of chemical analyses and physical experiments. Although the techniques used in the preparation of samples for microscopy are usually well described (notwithstanding that several of them may lead to the formation of “artifacts”) the representativeness of images with regard to the whole food specimen, not to mention the amount of “processing” an image may have undergone (see subsequently), is rarely (if ever) discussed. Thus, images have become a sui generis type of data in scientific research.

Several studies suggest relationships between bulk properties of foods (for example, mechanical moduli, thermal properties, and so on) and their microstructural characteristics (based on images). For instance, Brunschwiler and others (2006) studied the microstructure–texture relationships of fresh pastes of yam,insinating that the firmness of yam pastes was related to the extent of cell disintegration. Lee and others (2003) examined the relationships between microstructure and rheological properties in processed cheese indicating that the formation of a protein network determined the viscosity. Mellema and others (2000) explained the development of the storage modulus in casein gels with time under different conditions and related the values to the “density” of the gel network. Yet, the question often arises as to whether the examined microstructure is representative of the “bulk” of the product.

The objective of this hypothesis article was to use the bootstrap method to determine a representative sampling area size and the error involved in measuring elements by image analysis, using images of simulated and real structures (apple tissue).

Selecting a scale

Obtaining quantitative information of food structures is a complex task since the spatial scales involved vary from the nanodomain (1 to 100 nm) to the macroscale (order of millimeter), encompassing over 6 decades of length. Most real foods (for example, an apple slice, the crumb of loaf of bread, a portion of yoghurt) have dimensions in the order of centimeters while key microstructural elements (for example, cell walls, air cells, protein network) may be 1 to several orders of magnitude smaller. Natural and processed foods may look quite uniform when examined at the macroscopic scale or viewed with the naked eye (or at least “regularly” disordered, as the porous crumb of bread), thus, measured bulk physical properties (moduli, viscosity, diffusivity, and so on) are intrinsically assumed to correspond to a “homogeneous” material. However, most of the macroscopic properties depend on microscopic features, as is the case of texture of hard cooked beans (and the adhesion between cell walls) or the stability of an oil-in-water (O/W) emulsion (and amount of emulsifier located at the interface of fine droplets) and at this scale foods may look quite heterogeneous and anisotropic. This frequently means that to obtain meaningful structural data, a large number of samples at various places within the total volume of the food and at different spatial scales must be obtained (Graham and Yang 2003). Usually, a "relevant" structural scale or size becomes apparent to explain a macroscopic
behavior; in the previous example, this corresponds to that of the cell wall and the droplet interface, respectively. Mebatian and others (2008) explain that to study gas transport through intercellular spaces during respiration of whole fruits, imaging at very high magnifications will place us “inside the pores” by increasing the field of view the individual pores, which are the keys elements in the phenomenon, will become apparent.

Quantification of the microstructure

Imaging foods with sophisticated microscopes is not a final objective of microstructural studies. Once digital images are obtained, a goal of food materials science is to derive numerical models that describe the relationship between the structure and properties (Aguilera and Lillford 2008). Image analysis or acquiring numerical data from images is becoming a standard procedure due to the proliferation of software that allow measuring some structural characteristics of foods from digital images (Lim and Barigou 2004; Aguilera and Germain 2007). In simple terms, the process of image analysis starts by identifying key elements or features in the image, isolating them from the background, and finally taking measurements from the binary image (for example, where elements are black and background is white). Manipulation of a digital image is performed at the level of pixels and proper calibration transforms pixel values into actual dimensions (for example, micrometers). Most of the direct measurements on individual objects are related to their size and shape, in other words, to their morphometry (Table 1). It is also possible to acquire quantitative information of the overall “pattern” of an image, this is, the visual appearance of irregularities and variations in distribution of elements throughout the entire image (Zheng and others 2006; Aguilera and Germain 2007). Quantitative analysis may extend to the chromatic intensity of elements or zones, which, depending on the method of illumination, can imply changes in composition, temperature, or density.

Original compared with “processed” images

Very often, captured digital images are subject to defects or limitations that preclude the direct use of software for quantitative analysis. Overlapping objects and poorly defined edges of elements are among the several nuisances that complicate subsequent binarization, segmentation, and measurement. Thus, the operator is confronted with the option of acquiring quantitative information directly from the original image or to improve the quality of the initial image by processing (processed image) before proceeding with segmentation and measurement (Aguilera and Germain 2007). Both approaches present advantages and disadvantages. Manual image processing (sometimes done at the pixel level) has the advantage of being more “realistic” than processing with a machine or software (which depends on the spatial resolution of the image) because a human operator is far better trained to discriminate objects; however, this is time consuming and subject to a strong bias (that is, the “processed” image depends on criteria used by each operator). Software-based routines for processing whole images are bias-free (independent of the operator), give reproducible results, and are also faster if a large amount of images has to be analyzed. However, these processed images may create features or remove important information (Aguilera and Germain 2007).

The bootstrap method

The main advantage of the bootstrap method in image analysis is that no assumptions have to be previously stated about the distribution of elements within the population. It is an empirical nonparametric technique commonly used to obtain several measures of a given statistic when the underlying statistical model is not known (Pajevic and Basser 2003). Bootstrap theory entails collecting a random set of data points with replacement from the original population. The bootstrap data set may then be formed by members of the original population appearing repeatedly (Efron and Tibshirani 1993). Data collection is performed for a previously defined number of times so that several bootstrap data sets are generated. Statistical analysis of the values contained in the bootstrap data sets are used to estimate the standard error, bias, confidence intervals, probability distribution, and other measures of uncertainty for a given statistic (Pajevic and Basser 2003). For the determination of the standard error, a typical number of collected data sets for bootstrapping ranges from 50 to 200. The bootstrap standard error (SE\text{Boot}) is the standard deviation of the bootstrap replication (Efron and Tibshirani 1993):

\[
\text{SE}_{\text{Boot}} = \sqrt{\frac{1}{B-1} \sum_{b=1}^{B} (\theta(b) - \bar{\theta})^2}
\]

where \(\theta(b)\) is the measured statistic (for example mean, median, and so on), \(B\) is the bootstrap number (number of random data sets), \(n\) is the subsampling used in analysis, and

\[
\bar{\theta} = \frac{1}{B} \sum_{b=1}^{B} \theta(b)
\]

Table 1 – Most common morphological measurements used in quantitative food microscopy.

<table>
<thead>
<tr>
<th>Measurement</th>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size</td>
<td>Area</td>
<td>Determined by counting the number of pixels that are part of the objects</td>
</tr>
<tr>
<td></td>
<td>Perimeter</td>
<td>Determined by counting the number of pixels that are part of the border of the objects</td>
</tr>
<tr>
<td></td>
<td>Diameter</td>
<td>Determined precisely for “round” objects. For other shapes equivalent diameters may be calculated (see below)</td>
</tr>
<tr>
<td></td>
<td>Volume fraction</td>
<td>Determined by dividing the total area of region of interest by the total area of the image</td>
</tr>
<tr>
<td>Shape</td>
<td>Shape factor</td>
<td>(\frac{4 \times \pi \times \text{area}}{\text{perimeter}^2}) major axis length</td>
</tr>
<tr>
<td></td>
<td>Elongation</td>
<td>(\frac{4 \times \pi \times \text{area}}{\text{minor axis length}^2})</td>
</tr>
<tr>
<td></td>
<td>Equivalent diameter</td>
<td>(\sqrt{\frac{4 \times \pi \times \text{area}}{\text{major axis length}^2}})</td>
</tr>
<tr>
<td></td>
<td>Compactness</td>
<td>(\frac{4 \times \pi \times \text{area}}{4 \times \text{area}})</td>
</tr>
<tr>
<td></td>
<td>Roundness</td>
<td>(\frac{\pi \times \text{max diameter}^2}{\pi \times \text{max diameter}^2})</td>
</tr>
</tbody>
</table>
Image analysis of food structures...

The bootstrap coefficient of variation ($CV_{Bo}$) is estimated as follows (Wang and others 2008):

$$CV_{Bo} = \frac{SE_{Bo}}{\theta * (\cdot)}$$

The absolute error ($d$) and the relative error (RE) of the measured statistics are calculated as follows:

$$d = |\mu - \theta * (\cdot)|$$

$$RE = \frac{d}{\mu}$$

where $\mu$ is the mean value of data for the population.

In our demonstration of the bootstrap method to obtain representative sampling areas, it will be assumed that the area of a whole image is actually the object to be analyzed, thus, it contains the total population of relevant elements, while the random data sets correspond to the sampling areas (in this case of equal sizes). Each bootstrap data set then is formed the measured morphological parameters of each element (for example, the diameter, perimeter, area, and so on) from the sampling area and the calculated statistical parameters (for example, the mean). This results in a large number of estimations of the statistics of interest (one for each bootstrap data set), which will then be used to determine the coefficient of variation ($CV_{Bo}$) and standard error ($SE_{Bo}$) for the particular sampling area (Figure 1). By increasing the sampling area size it is possible to determine the representative sampling area size that defines a minimum (or acceptable) value for the coefficient of variation or the standard error. For the application of the bootstrap method, the reader is referred to the study by Wang and others (2008) who studied a sampling strategy for measuring the distribution of soil moisture from digital images.

**Relationship between representative sampling area size and size distribution curves**

The elements of food structures such as pores, cracks, cells, networks, bubbles, crystals, droplets, and so on, have size distributions that are generally nonparametric. However, parametric statistics often suffice to evaluate size distributions of elements in an image and the parameter to determine is the coefficient of variation ($CV_{image}$). This coefficient gives a measure of how disperse is

---

**Figure 1** — Scheme of the method applied to an image to determine the bootstrap coefficient of variation and the standard error for the measured statistics.

---

**Figure 2** — Computer-simulated structures and their respective $CV_{image}$ values characterizing the size distribution of circular elements. Simulated structures are sorted from lowest to highest $CV_{image}$ value.
Image analysis of food structures...

the distribution relative to its mean value ($\mu$) and is calculated as:

$$CV_{image} = \frac{SD}{\mu}$$  \hspace{1cm} (6)

where SD is the standard deviation of the elements in the population.

Recapitulating, measuring key features from images is a necessary but not sufficient condition to develop meaningful structure–property relationships. Acquired structural data also has to be representative of the entire sample, as the bulk property being measured is usually a lumped value. This requirement has been difficult to accomplish in foods where structural evidence is often presented in the form of photomicrographs taken from a minute portion of the original area or volume. Hence, a balance has to be achieved between the cost and effort of acquiring microstructural images from the original area or volume. This requirement has been difficult to accommodate in foods where structural evidence is often presented in the form of photomicrographs taken from a minute portion of the original area or volume. Hence, a balance has to be achieved between the cost and effort of acquiring microstructural images from a food of macroscopic dimensions and the need to have representative structural data valid for the whole food. As stated before, in this study, the structures of real foods have been first replaced by computer-simulated images and then represented by apple tissue samples, which has been done to assess the quality of the information acquired by analyzing randomly selected areas of different sizes.

Application of Bootstrap to Estimate the Sampling Area Size of Simulated Structures

Simulated structures

To apply the bootstrap method to simulated structures 10 images (S1, S2, …, S10) consisting of sets of black circles with different diameters (D) and different size distributions were generated using a routine developed in Matlab version 7.0.1 release 14 (Mathworks Inc., Natick, Mass., U.S.A.). All simulated structures images had a size of 1000 × 1000 pixels and contained a total of 225 elements (Figure 2).

Bootstrap method for image analysis in simulated structures

The bootstrap method is based in the study by Efron and Tibshirani (1993). The procedure defines a certain sampling area of a size smaller than the original image, which is then moved randomly throughout the image and in each step ($b = 1$ to $B_b$), the mean diameter of the elements inside (not on the edges) the sampling area is determined. The process is repeated for $n$ different sampling areas sizes. Finally, the different estimations of the mean $\theta*(b)$ are used to calculate $SE_{Bn}$ and $CV_{Bn}$ according the Equations (1) and (3), as shown schematically in Figure 1. A total of 9 sizes of sampling areas ($n$) were used for each of the simulated structures shown in Figure 2, namely, squares of $100 \times 100$, $200 \times 200$, $300 \times 300$, $400 \times 400$, $500 \times 500$, $600 \times 600$, $700 \times 700$, $800 \times 800$, and $900 \times 900$ pixels corresponding to 1%, 4%, 9%, 16%, 25%, 36%, 49%, 64%, and 81% of the total image area, respectively. The number of iterations for each sample area was $B = 300$. This value was selected based on a recommendation given by Efron and Tibshirani (1993) to estimate the standard error. The algorithm was developed in Matlab version 7.0.1 release 14.

Application of bootstrap method to estimate the sampling area size in simulated structures

The representativeness of a sampling area depends on the goodness to match statistical parameters obtained for the entire original population.

Table 2 — Summary of statistical parameters for the 10 simulated structures presented in Figure 2.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>S1</th>
<th>S2</th>
<th>S3</th>
<th>S4</th>
<th>S5</th>
<th>S6</th>
<th>S7</th>
<th>S8</th>
<th>S9</th>
<th>S10</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>39.6</td>
<td>49.3</td>
<td>40.2</td>
<td>39.1</td>
<td>16.3</td>
<td>33.3</td>
<td>6.2</td>
<td>9.1</td>
<td>10.2</td>
<td>4.8</td>
</tr>
<tr>
<td>SD</td>
<td>5.6</td>
<td>9.6</td>
<td>8.8</td>
<td>10.9</td>
<td>8.3</td>
<td>20.4</td>
<td>5.6</td>
<td>8.5</td>
<td>10.2</td>
<td>5.6</td>
</tr>
<tr>
<td>$CV_{image}$</td>
<td>0.14</td>
<td>0.19</td>
<td>0.22</td>
<td>0.28</td>
<td>0.51</td>
<td>0.61</td>
<td>0.91</td>
<td>0.93</td>
<td>1.01</td>
<td>1.16</td>
</tr>
<tr>
<td>Max D</td>
<td>56.1</td>
<td>75.5</td>
<td>65.8</td>
<td>71.8</td>
<td>48.6</td>
<td>69.7</td>
<td>27.1</td>
<td>56.9</td>
<td>76.3</td>
<td>33.4</td>
</tr>
<tr>
<td>Min D</td>
<td>24.3</td>
<td>25.7</td>
<td>21.8</td>
<td>9.4</td>
<td>4.1</td>
<td>1.1</td>
<td>1.1</td>
<td>1.1</td>
<td>1.1</td>
<td>1.1</td>
</tr>
</tbody>
</table>

$\mu$ = mean diameter; SD = standard deviation; $CV_{image}$ = coefficient of variation of elements on image; D = diameter. All values are in pixels except $CV_{image}$, which is dimensionless.

Figure 3 — Effect of sampling area size on the bootstrap coefficient of variation for the 10 simulated structures. Simulated structures were sorted from the lowest (S1) to the highest (S10) $CV_{image}$ values. ◯ S1; ▲ S2; ▼ S3; □ S4; ▼ S5; + S6; ◆ S7; ○ S8; ◆ S9; + S10.
Image analysis of food structures . . .

images. Table 2 summarizes the statistical parameters calculated for simulated structures represented by images S1 to S10 (Figure 2). $CV_{image}$ varied between 0.14 for S1 and 1.16 for S10. Figure 3 and 4 show the effect of increasing the sampling area size from $100 \times 100$ to $900 \times 900$ pixels on the values of $CV_{Bn}$ and $SE_{Bn}$, respectively. Increasing the size of the sampling area reduced the values of $CV_{Bn}$ and $SE_{Bn}$ in all cases, independent of the size distribution of elements. This is due to a decrease in the variability of the calculated mean since a larger sampling area includes more elements.

However, since $CV_{Bn}$ is a normalized value with respect to the mean, it presents an advantage over the $SE_{Bn}$ when comparing different images. One would surmise that the order of $CV_{image}$ values (for example, from lowest to highest, S1 to S10) would be maintained in Figure 3 and 4 when $CV_{Bn}$ or $SE_{Bn}$ are plotted. Nevertheless, this only occurred for $CV_{Bn}$ (Figure 3) and for this reason it was decided to continue the rest of the analysis based on this parameter.

Figure 5 suggests a linear relationship between $CV_{Bn}$ and $CV_{image}$ with decreasing slopes as the sampling area size increases. This graph helps in choosing a sample area size that will give a desired $CV_{Bn}$. For example, to achieve a $CV_{Bn}$ not higher than 0.10 (see horizontal dotted line in Figure 5), the sampling size area for an image with low dispersion of elements (for example, S1 to S4) would be only 9% of the total area ($300 \times 300$ pixels) while for an image of highly dispersed elements (for example, S10), the sampling area would be 49% of the whole image ($700 \times 700$ pixels). Data for the smallest sampling area size of $100 \times 100$ pixels was not incorporated in Figure 5 because it did not follow a linear trend. As the size of the sampling area becomes of the order of magnitude of the largest element present in the image, the possibility of having an empty set during bootstrapping increases, thus, biasing the calculation. From a practical viewpoint, Figure 5 suggests that if the size distribution of structural elements in the food has a low coefficient of variation (that is, low $CV_{image}$), then the size of the sampling area
Image analysis of food structures...

Figure 6 — Relationship between the coefficient of variation of elements present in simulated structures and the required sampling area size to achieve a relative error of 10%. ♦ S1; ✶ S2; △ S3; □ S4; ■ S5; ♦ S6; ⊙ S7; ○ S8; ▲ S9; + S10.

Figure 7 — Schematic representation of processing of images of apple tissue to determine cellular cavities. The black marker represents 200 μm.

Figure 8 — Photomicrographs of apple tissue and their respective CV_image values characterizing the size distribution of the areas of cellular cavities. Images are sorted from the lowest (I1) to the highest (I8) CV_image value. The black marker represents 200 μm.

becomes irrelevant (for example, a small sampling area would be enough).

The estimation of the size of the sampling area must be supported by quantitative information obtained from images. A possibility is to consider the relative error (RE), representing how much the mean of the bootstrap ($\bar{\theta}^*$) deviates from the mean of the whole sample ($\mu$). Figure 6 shows a linear relation ($R^2 = 0.83$) between $CV_{image}$ and the sampling area size required to obtain a mean value of D with a relative error of 10% or less. For distributions with lower dispersions (for example, image S3 with $CV_{image} = 0.22$) a sampling area size of $200 \times 200$ pixels is required, whereas in the case of higher dispersions (for example, image S9 with $CV_{image} = 1.01$) sampling of an area 6.25 times larger ($500 \times 500$ pixels) is needed.

Graham and Yang (2003) estimated a representative area size to analyze calcium sulfate (CS) inclusions in a steel specimen of 50 mm² from images acquired with an optical microscope. To
Image analysis of food structures...

estimate the representative area size, they started with a sampling area of 0.091 mm² (0.2% of original specimen) and successively increased the size to evaluate changes in particle statistics (area fraction, average particle area, and number density). Their results suggest that data converged to a stable value of particle statistics for a sampling area larger than 20 mm² (40% of original specimen). The main advantage of applying bootstrap as presented in this study is that measurements are taken at randomly selected zones of the image rather than by enlarging an originally selected sampling area.

Application of Bootstrap to Estimate the Representative Sampling Area Size for Real Structure

Real structures

To apply the bootstrap method to real food structures, images of thin sections of apple tissue were analyzed for cellular cavities defined as areas within the image with a continuous boundary (Lewicki and Pawlak 2003). In this case, the boundaries were identified as cell walls or their remnants. Images were taken with a digital camera CoolSnap-Pro (Media Cybernetics Inc., Bethesda, Md., U.S.A.) attached to an optical microscope (Olympus BX50, Tokyo, Japan) and the area of cellular cavities \(A\) was determined using an algorithm developed in Matlab version 7.0.1 release 14 that consisted in image acquisition, conversion to gray scale, filter processing, binarization, and segmentation of cell cavities (Figure 7). To prevent noise in the analysis, only cellular cavities with an area larger than 1000 pixels were considered (Table 3).

Table 3 – Summary of statistical parameters for the 8 images of apple tissue presented in Figure 8.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>I1</th>
<th>I2</th>
<th>I3</th>
<th>I4</th>
<th>I5</th>
<th>I6</th>
<th>I7</th>
<th>I8</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\mu)</td>
<td>0.0031</td>
<td>0.0034</td>
<td>0.0038</td>
<td>0.0047</td>
<td>0.0043</td>
<td>0.0051</td>
<td>0.0050</td>
<td>0.0054</td>
</tr>
<tr>
<td>SD</td>
<td>0.0019</td>
<td>0.0023</td>
<td>0.0029</td>
<td>0.0037</td>
<td>0.0036</td>
<td>0.0036</td>
<td>0.0057</td>
<td>0.0059</td>
</tr>
<tr>
<td>CV(_{image})</td>
<td>0.60</td>
<td>0.67</td>
<td>0.76</td>
<td>0.79</td>
<td>0.85</td>
<td>1.11</td>
<td>1.20</td>
<td>1.41</td>
</tr>
<tr>
<td>Max A</td>
<td>0.0135</td>
<td>0.0157</td>
<td>0.0203</td>
<td>0.0202</td>
<td>0.0197</td>
<td>0.0375</td>
<td>0.0376</td>
<td>0.0662</td>
</tr>
<tr>
<td>Min A</td>
<td>0.0013</td>
<td>0.0014</td>
<td>0.0013</td>
<td>0.0013</td>
<td>0.0013</td>
<td>0.0014</td>
<td>0.0013</td>
<td>0.0013</td>
</tr>
</tbody>
</table>

\(\mu\) = mean of the area of cellular cavities; SD = standard deviation; CV\(_{image}\) = coefficient of variation of elements on image; \(A\) = area of cellular cavities. All values are in millimeters except CV\(_{image}\), which is dimensionless.

Figure 9 – Effect of sampling area size on the bootstrap coefficient of variation for the 8 apple tissue images analyzed. The images were sorted from the lowest (I1) to the highest (I10) CV\(_{image}\) value. △ I1; □ I2; ■ I3; ● I4; • I5; ○ I6; ▲ I7; + I8.
relationship was also obtained between $CV_{image}$ and size (or percentage) of the sampling area but with a higher regression coefficient ($R^2 = 0.987$) than for simulated structures. In case of images with small dispersion values (for example, I1 with $CV_{image} = 0.60$), a sampling area size of $200 \times 200$ pixels is required to obtain an average area of cellular cavities with RE = 10%, whereas in images with higher dispersion (for example, I8 with $CV_{image} = 1.41$), sampling of an area of $1000 \times 1000$ pixels was needed.

When images of apple tissue (for example, I1) and simulated structures (for example, S6) having approximately similar dispersion values to obtain a RE = 10% ($CV_{image}$ approximately equal to 0.60) are compared it is observed that in the 1st case, a sampling area of $200 \times 200$ pixels is enough while for simulated structures it required $300 \times 300$ pixels. This difference could be attributed to the fact that the size distribution of elements in images may not be parametric (for example, not represented by a normal distribution), thus, the $CV_{image}$ is only a limited estimator of element dispersion. Alternatively the difference could also be ascribed to the way in which the elements are presented in the image. In the simulated structures, the circles were aligned at a specific distance between them, while in the apple tissue, cellular cavities were randomly located and with different distance separations between them (or their centroids). This suggests that the number of elements considered during each bootstrapping could influence the calculated mean value ($\theta^{-1}$), thus, RE. However, these differences could be reduced by increasing the B value and in this way, incorporate more elements during bootstrapping.

Conclusions

Since structure–property relationships in foods usually link information at the microlevel (microstructure) with lumped properties at the macrolevel (product), structural data should be representative of the whole sample under consideration. The bootstrap method has proven to be a suitable tool to determine a representative sampling area size given that the overall structure has been properly characterized (for example, the CV of elements in the whole sample to be analyzed has been estimated). The inverse relationship between sampling area size and $CV_{info}$ and $SE_{info}$ suggests that a small sampling area results in high $CV_{info}$ and $SE_{info}$. Samples with a large widespread in size distributions ($CV_{image}$ in this article) will demand larger sampling area sizes to achieve low REs or $CV_{info}$.

In microscopy practice, this means that a larger number of images should be examined so that the structural information derived becomes reliable.

Acknowledgments

The authors acknowledge the discussions with Ph.D. students Wenceslao Medina and Rommy Zúñiga and the financial support from CONICYT to Cristian Ramírez.

References