

## Smooth estimation of size distributions in an oriented cylinder model

Jongbloed, Geurt; McGarrity, Kimberly S.; Sietsma, Jilt

**DOI**

[10.1007/s11749-020-00729-9](https://doi.org/10.1007/s11749-020-00729-9)

**Publication date**

2020

**Document Version**

Final published version

**Published in**

Test

**Citation (APA)**

Jongbloed, G., McGarrity, K. S., & Sietsma, J. (2020). Smooth estimation of size distributions in an oriented cylinder model. *Test*, 30 (2021)(2), 505-526. <https://doi.org/10.1007/s11749-020-00729-9>

**Important note**

To cite this publication, please use the final published version (if applicable). Please check the document version above.

**Copyright**

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

**Takedown policy**

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.



# Smooth estimation of size distributions in an oriented cylinder model

Geurt Jongbloed<sup>1</sup> · Kimberly S. McGarrity<sup>2</sup> · Jilt Sietsma<sup>1</sup>

Received: 10 March 2017 / Accepted: 5 August 2020 / Published online: 14 August 2020  
© The Author(s) 2020

## Abstract

Kernel estimators are proposed for estimating the cumulative distribution functions and the probability density functions of several quantities of interest in a stereological oriented cylinder model. This oriented cylinder model was developed to represent anisotropic microstructural features in materials. The asymptotic properties of these estimators are studied, and the estimators are applied to two banded dual phase steel microstructures. The estimation method is quite general and can also be applied to distributions of other univariate quantities of interest.

**Keywords** Nonparametric estimation · Asymptotics · Materials science · Inverse problem · Stereology

**Mathematics Subject Classification** 62G05 · 62G20 · 62P30

## 1 Introduction

The so-called Wicksell problem introduced in Wicksell (1925) is a classical inverse problem in statistics. The original motivation was medical. A postmortem examination of spleens containing approximately spherical tumors was performed. Based on cross sections of the spleens (showing circular profiles of the tumors), the aim was to estimate the distribution of tumor sizes based on the observed circle radii. Wicksell's problem is a typical example of a stereological problem, where one aims to infer 'three-dimensional properties' from 'two-dimensional information'. Not only within the field of anatomy, but also in materials science and astronomy, this type of problem

---

✉ Geurt Jongbloed  
G.Jongbloed@tudelft.nl  
Kimberly S. McGarrity  
mcgarrityk@gmail.com

<sup>1</sup> Delft University of Technology, Delft, The Netherlands

<sup>2</sup> Delft University of Technology and Materials innovation institute, Delft, The Netherlands

is frequently encountered. See, e.g., Sen and Woodroffe (2011) for an astronomical application of the model. Over the years, quite some stereological problems related to Wicksell's problem have been introduced and studied; see, e.g., Ohser and Mücklich (2000) for problems related to different shapes of the three-dimensional objects and Feuerverger and Hall (2000) for a problem where the data are obtained slightly differently.

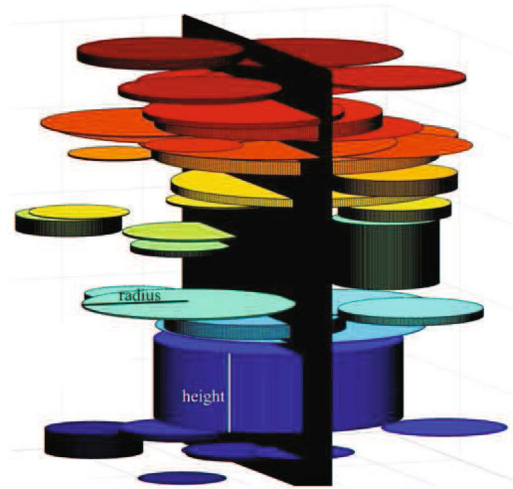
In this paper, we study another related model, specifically designed for a materials science problem. In this model, circular cylinders (all with the same orientation, say vertical axes) are distributed within an opaque medium which is cut vertically (parallel to the axes). The problem then is to estimate distributional properties of various three-dimensional quantities related to size (volume, surface area, e.g.,) only based on data obtained from the two-dimensional section. This model was introduced in McGarrity et al. (2014), where also the relations between the distributions of (unobservable) three-dimensional quantities and (observable) two-dimensional quantities are derived. These will be reviewed in Sect. 2. In that paper, estimators of the distribution functions are defined and studied asymptotically. These estimators are step functions. Especially in the metallurgical context, such cumulative distribution functions are considered undesirable, as they are harder to interpret for practitioners than density functions that give more direct visual information on the relative occurrence of the various sizes in the material. Section 3 discusses smooth estimators for cumulative functions in the oriented cylinder model. A particular feature of these smooth estimators is that their pointwise asymptotic behavior does depend on the rate at which the bandwidth vanishes (rate  $n^{-1/4}$  is optimal), but *not* on the constant in front of this rate.

Estimates of density functions can be obtained from these via differentiation. In Sect. 4, these density-like functions are defined and studied asymptotically. It turns out that for estimating these, the bandwidth should vanish at rate  $n^{-1/6}$  in order to let the MISE vanish at rate  $n^{-2/3}$ . The choice of the constant in front of  $n^{-1/6}$  is important, and we describe a reference family method to obtain data driven bandwidths based on the expressions for the asymptotically MISE-optimal bandwidths. Finally, in Sect. 5 we apply the proposed estimators to a real microstructural dataset obtained at TATA Steel.

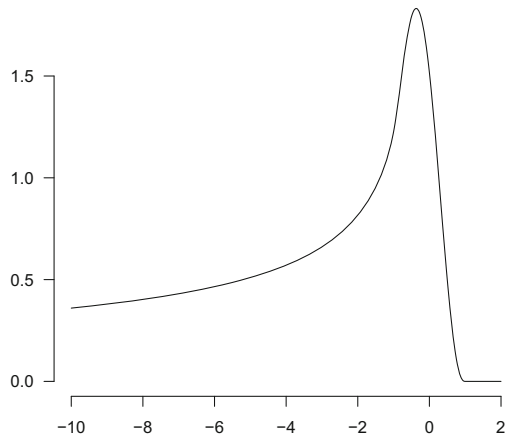
## 2 An oriented cylinder model

In the process of representing microstructural features of interest like those mentioned in Sect. 1, a first (simple) model was proposed in McGarrity et al. (2014). We describe this model briefly here. Consider a large box in 3D, that is cut by a vertical plane. Throughout the box, points are distributed according to a low-intensity Poisson process that is homogeneous in the direction perpendicular to the cutting plane. At these points, circular cylinders are placed, all oriented in the same way, with vertical axes of symmetry. See Fig. 1 for an illustration of the situation. The squared radius  $X$  (which we consider rather than the radius itself, following the example of Hall and Smith (1988) in Wicksell's problem) and height  $H$  of the cylinders are generated as i.i.d. bivariate random vectors  $(X, H)$  drawn from the bivariate density  $f$ , corresponding to the 3D microstructural features of interest. Note that  $f$  is a joint density and  $X$  and

**Fig. 1** Impression of the cylinders randomly distributed in a box, with the cutting plane where rectangular intersections can be observed



**Fig. 2** Function  $\bar{K}$  given in (11) based on the biweight kernel

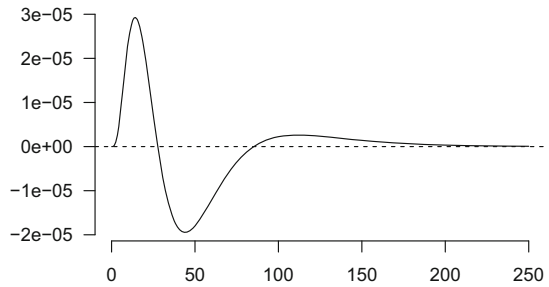


$H$  are not assumed independent. The data consist of the rectangular profiles of the cylinders cut by the plane.

The height of the rectangle is equal to the height of the cut cylinder, and the width of the rectangle is a fraction of the diameter of the cylinder. Taking into account that the probability of cutting a cylinder by the plane depends on the radius of the cylinder, the relationship between the joint density  $g$  of the observed rectangle pairs  $(Z, H)$ , the squared half-width  $Z$  and height  $H$ , and the joint density  $f$  of  $(X, H)$  is derived in McGarrity et al. (2014):

$$g(z, h) = \frac{\int_{x=z}^{\infty} (x - z)^{-\frac{1}{2}} f(x, h) dx}{2 \int_{x=0}^{\infty} \sqrt{x} f_X(x) dx} = \frac{1}{2 m_F^+} \int_{x=z}^{\infty} (x - z)^{-\frac{1}{2}} f(x, h) dx. \quad (1)$$

**Fig. 3** Numerical approximation of the function  $\phi_{11,7}$



Here,  $m_F^+ = E_f[\sqrt{X}] < \infty$ . This equation can be inverted to express the joint density  $f$  in terms of  $g$ :

$$f(x, h) = -\frac{1}{m_G^-} \frac{\partial}{\partial x} \int_{z=x}^{\infty} (z - x)^{-\frac{1}{2}} g(z, h) dz. \tag{2}$$

Here  $m_G^- = E_g[Z^{-1/2}] < \infty$ . For the derivation of these relations, see Sect. 2 in McGarrity et al. (2014).

Based on these relations, it is possible to estimate the distribution of univariate quantities related to the distribution of  $(X, H)$ . In this paper, we restrict ourselves to the squared radius  $(X)$  and the volume  $V = \pi XH$ .

In order to save space in notation, define for  $h, t > 0$  the function

$$q(h; t) = \begin{cases} t & \text{squared radius: } T = X, \\ \frac{t}{\pi h} & \text{volume: } T = \pi XH. \end{cases} \tag{3}$$

These quantities are chosen such that the random variable of interest,  $T$ , satisfies  $T > t$  if and only if  $X > q(H; t)$ . Hence, using (2) we obtain the general form of the distribution functions for these quantities

$$F_T(t) = 1 - \int_{h=0}^{\infty} \int_{x=q(h;t)}^{\infty} f(x, h) dx dh = 1 - \frac{N(t)}{N(0)} \tag{4}$$

where

$$N(t) = \int_{h=0}^{\infty} \int_{z=q(h;t)}^{\infty} [z - q(h; t)]^{-\frac{1}{2}} g(z, h) dz dh. \tag{5}$$

Note that  $N(t) \leq N(0) = m_G^- = E_g[Z^{-1/2}]$ . The distribution functions of the unobservable cylinder quantities are expressed now in terms of  $N$  which can in turn be derived from the joint density  $g$  of the observable pairs  $(Z, H)$ .

In practice, one is often also interested in an estimate of the probability density functions of the various univariate quantities mentioned in (3). Taking the derivative of (4) yields

$$f_T(t) = \frac{d}{dt} F_T(t) = \frac{d}{dt} \left( 1 - \frac{N(t)}{N(0)} \right) = -\frac{\frac{d}{dt} N(t)}{N(0)}. \tag{6}$$

Now  $N$  can be estimated empirically, replacing the integral with respect to the joint density  $g$  in (5) by the integral with respect to the empirical distribution of the observed pairs  $(X_i, H_i)$ ,  $1 \leq i \leq n$ . This leads to

$$N_n(t) = \frac{1}{n} \sum_{i=1}^n [Z_i - q(H_i; t)]^{-\frac{1}{2}} 1_{[Z_i > q(H_i; t)]}. \tag{7}$$

While this estimator will provide an estimate for the function  $N$  given in (5), it can have some undesirable properties. One is that it is not a decreasing function, in fact it has discontinuities of infinite size. Indeed, taking the squared radius as example, such discontinuity appears at  $t = Z_i$ , for any value of  $i$ . In McGarrity et al. (2014), an isotonization procedure is used to obtain an estimator that is a decreasing step function.

Another, related issue is that the estimate  $N_n$  will not be smooth. Often one wants to assume that  $N$  is smooth and estimate its derivative. This function gives insight in the relative probabilities with which certain values of the quantities of interest occur. In Sects. 3 and 4, we introduce kernel estimators for the functions  $N$  and their derivatives.

### 3 Estimators for the function $N$

In view of the function  $N$  defined in (5) and its empirical estimator  $N_n$  given in (7), there are various approaches one can take to obtain smooth estimators for  $N$ . One approach is to substitute a smoothed empirical distribution of the observed pairs  $(X_i, H_i)$  for  $g$  in (5) rather than the empirical distribution itself. In the (univariate) context of Wicksell’s problem, this approach was originally proposed in Taylor (1983). A related estimator (based on squared radii rather than radii) was introduced in Hall and Smith (1988).

Still in Wicksell’s problem, van Es and Hoogendoorn (1990) suggest an alternative smooth estimator, obtained that by kernel smoothing of the function  $N_n$ .

Following the latter approach, we define a smooth estimator of  $N$ , based on smoothing the empirical plug-in estimator  $N_n$  defined in (7). For this, we use a smoothing kernel  $K$ , having the usual properties (symmetric continuously differentiable probability density supported on  $[-1, 1]$ ). To make it more concrete, we use the biweight kernel  $K$ , defined as

$$K(u) = \frac{15}{16} (1 - u^2)^2 1_{[-1, 1]}(u) \tag{8}$$

although this particular choice is not essential. Take a sequence of bandwidths  $(b_n)$  with  $b_n \downarrow 0$  as  $n \rightarrow \infty$  and define for  $t > 0$

$$\begin{aligned} \tilde{N}_n(t) &= \frac{1}{b_n} \int_{s=t-b_n}^{t+b_n} K\left(\frac{t-s}{b_n}\right) N_n(s) ds \\ &= \frac{1}{nb_n} \sum_{i=1}^n \int_{s=t-b_n}^{t+b_n} K\left(\frac{t-s}{b_n}\right) [Z_i - q(H_i; s)]^{-\frac{1}{2}} 1_{[Z_i > q(H_i; s)]} ds. \end{aligned} \tag{9}$$

For this estimator, the mean squared error for estimating  $N(t)$  (for fixed  $t$ ) is defined by

$$\text{MSE}(\tilde{N}_n(t)) = (E[\tilde{N}_n(t)] - N(t))^2 + \text{Var}(\tilde{N}_n(t)).$$

In order to study the asymptotic behavior of this MSE at fixed location  $t > 0$ , we impose the following condition:

**Condition 3.1** The function  $N$  is twice continuously differentiable at  $t$ .

Now, under Condition 3.1 and fixing  $t > 0$ , for  $n$  tending to infinity the expectation of  $\tilde{N}_n(t)$  is given by

$$\begin{aligned} E[\tilde{N}_n(t)] &= \frac{1}{b_n} \int_{s=t-b_n}^{t+b_n} K\left(\frac{t-s}{b_n}\right) E[N_n(s)] ds \\ &= \frac{1}{b_n} \int_{s=t-b_n}^{t+b_n} K\left(\frac{t-s}{b_n}\right) N(s) ds = \int_{u=-1}^1 K(u)N(t-ub_n) du \\ &= \int_{u=-1}^1 K(u) \left[ N(t) - ub_n N'(t) + \frac{1}{2}(ub_n)^2 N''(\xi_{u,n}) \right] du \\ &= N(t) + \frac{1}{2} b_n^2 N''(t) \int_{u=-1}^1 u^2 K(u) du + o(b_n^2) \quad \text{for } n \rightarrow \infty, \end{aligned} \tag{10}$$

where  $\xi_{u,n}$  denotes a point between  $t$  and  $t - ub_n$ . For the squared bias part of the MSE this yields, for  $n \rightarrow \infty$ ,

$$(E[\tilde{N}_n(t)] - N(t))^2 = \frac{1}{4} b_n^4 N''(t)^2 \left( \int_{u=-1}^1 u^2 K(u) du \right)^2 + o(b_n^4).$$

Note that this asymptotic bias can be derived for both choices of  $q$  listed in (3) simultaneously. For the asymptotic variance of  $\tilde{N}(t)$ , both choices for  $q$  need to be dealt with separately. We follow the approach adopted in Hall and Smith (1988) for Wicksell’s problem and express  $\tilde{N}_n$  as a convolution of two functions.

First for the volume, we get (using that  $t > b_n$  for sufficiently large  $n$  because  $t > 0$ )

$$\begin{aligned} \tilde{N}_n^{vol}(t) &= \frac{1}{b_n n} \sum_{i=1}^n \int_{s=-\infty}^{\infty} K\left(\frac{t-s}{b_n}\right) \left[ Z_i - \frac{s}{\pi H_i} \right]^{-\frac{1}{2}} 1_{[Z_i > \frac{s}{\pi H_i}]} ds \\ &= \frac{1}{b_n n} \sum_{i=1}^n \int_{u=0}^{\infty} K\left(\frac{t - \pi H_i(Z_i - u)}{b_n}\right) u^{-\frac{1}{2}} \pi H_i du \\ &= \frac{1}{n} \sum_{i=1}^n \int_{u=0}^{\infty} K\left(\frac{\pi H_i u}{b_n} + \frac{t - \pi H_i Z_i}{b_n}\right) \left(\frac{\pi H_i u}{b_n}\right)^{-\frac{1}{2}} \left(\frac{\pi H_i}{b_n}\right)^{\frac{1}{2}} d\left(\frac{\pi H_i u}{b_n}\right) \end{aligned}$$

$$= \frac{1}{\sqrt{b_n n}} \sum_{i=1}^n \sqrt{\pi H_i} \int_{u=0}^{\infty} K \left( u + \frac{t - \pi H_i Z_i}{b_n} \right) u^{-\frac{1}{2}} du.$$

For the squared radius, also using that  $t > 0$  and so  $t > b_n$  for sufficiently large  $n$ , we obtain

$$\begin{aligned} \tilde{N}_n^{sr}(t) &= \frac{1}{b_n n} \sum_{i=1}^n \int_{s=-\infty}^{\infty} K \left( \frac{t-s}{b_n} \right) [Z_i - s]^{-\frac{1}{2}} 1_{[Z_i > s]} ds \\ &= \frac{1}{b_n n} \sum_{i=1}^n \int_{u=0}^{\infty} K \left( \frac{t - (Z_i - u)}{b_n} \right) u^{-\frac{1}{2}} du \\ &= \frac{1}{n} \sum_{i=1}^n \int_{u=0}^{\infty} K \left( \frac{u}{b_n} + \frac{t - Z_i}{b_n} \right) \left( \frac{u}{b_n} \right)^{-\frac{1}{2}} b_n^{-\frac{1}{2}} d \left( \frac{u}{b_n} \right) \\ &= \frac{1}{\sqrt{b_n n}} \sum_{i=1}^n \int_{u=0}^{\infty} K \left( u + \frac{t - Z_i}{b_n} \right) u^{-\frac{1}{2}} du. \end{aligned}$$

For smooth kernel functions supported on  $[-1, 1]$ , such as the biweight function given in (8), define the function

$$\bar{K}(v) = \int_{u=0}^{\infty} u^{-\frac{1}{2}} K(u+v) du = \begin{cases} 0 & \text{for } v \geq 1, \\ \int_0^{-v+1} u^{-\frac{1}{2}} K(u+v) du & \text{for } -1 < v < 1, \\ \int_{-v-1}^{-v+1} u^{-\frac{1}{2}} K(u+v) du & \text{for } v \leq -1. \end{cases} \tag{11}$$

See (30) for the explicit expression and Fig. 2 for a visualization of the function  $\bar{K}$  on the biweight Kernel function defined in (8). Then, the function  $\tilde{N}_n$  corresponding to the volume can be expressed as

$$\tilde{N}_n^{vol}(t) = \frac{1}{\sqrt{b_n n}} \sum_{i=1}^n \sqrt{\pi H_i} \bar{K} \left( \frac{t - \pi H_i Z_i}{b_n} \right). \tag{12}$$

In a similar fashion, we get for the function  $\tilde{N}_n$  corresponding to the squared radius distribution

$$\tilde{N}_n^{sr}(t) = \frac{1}{\sqrt{b_n n}} \sum_{i=1}^n \bar{K} \left( \frac{t - Z_i}{b_n} \right). \tag{13}$$

Now note that for  $v < -1$ ,

$$(-v + 1)^{-\frac{1}{2}} \leq \bar{K}(v) \leq (-v - 1)^{-\frac{1}{2}}. \tag{14}$$



This leads to the following asymptotic behavior of  $\bar{K}(v)$  for  $v \rightarrow -\infty$

$$\sqrt{\frac{-v}{-v+1}} \leq \sqrt{-v}\bar{K}(v) \leq \sqrt{\frac{-v}{-v-1}} \Rightarrow \sqrt{-v}\bar{K}(v) \rightarrow 1 \text{ for } v \rightarrow -\infty. \tag{15}$$

To pin down the asymptotic variance of  $\tilde{N}_n(t)$ , we also need the function

$$\tau_q(z) = \int_{h=0}^{\infty} g(q(h; z), h) dh \tag{16}$$

on  $[0, \infty)$

**Lemma 3.2** *Let  $K$  be a symmetric and continuously differentiable probability density on  $\mathbb{R}$ , supported on  $[-1, 1]$ . Let  $N$  satisfy Condition 3.1. and  $E_g[H] < \infty$ . Let  $t > 0$  and choose  $0 < b_n \rightarrow 0$  as  $n \rightarrow \infty$ . Assume that the function  $\tau_q$  defined in (16) is bounded and right continuous at  $t$ . Then,*

$$\text{Var}(\tilde{N}_n(t)) = \tau_q(t) n^{-1} \ln(b_n^{-1}) + O(n^{-1}), \tag{17}$$

for both the squared radius and volume.

**Proof** Consider  $\tilde{N}_n^{vol}(t)$ . Using representation (12) and the definition of  $\tau_q(z)$  at  $z = t$ , for the volume we have

$$\begin{aligned} n\text{Var}(\tilde{N}_n^{vol}(t)) &= \frac{\pi}{b_n} \text{Var}\left(\sqrt{H_1}\bar{K}\left(\frac{t - \pi H_1 Z_1}{b_n}\right)\right) \\ &= \frac{\pi}{b_n} \left\{ E\left[H_1\bar{K}\left(\frac{t - \pi H_1 Z_1}{b_n}\right)^2\right] - \left(E\left[\sqrt{H_1}\bar{K}\left(\frac{t - \pi H_1 Z_1}{b_n}\right)\right]\right)^2 \right\}. \end{aligned}$$

Using continuity of  $N^{vol}$  at  $t$  yields

$$E\left[\sqrt{H_1}\bar{K}\left(\frac{t - \pi H_1 Z_1}{b_n}\right)\right] = \sqrt{\frac{b_n}{\pi}} E[\tilde{N}_n^{vol}(t)] = \sqrt{\frac{b_n}{\pi}} N^{vol}(t) + o(\sqrt{b_n}),$$

giving, for  $n \rightarrow \infty$ ,

$$n\text{Var}(\tilde{N}_n^{vol}(t)) = \frac{\pi}{b_n} E\left[H_1\bar{K}\left(\frac{t - Z_1}{b_n}\right)^2\right] - N^{vol}(t)^2 + o(1).$$

Now, for  $\epsilon > 0$  and  $n$  sufficiently large such that  $b_n < \epsilon$ ,

$$\frac{\pi}{b_n} E\left[H_1\bar{K}\left(\frac{t - \pi H_1 Z_1}{b_n}\right)^2\right]$$

$$= \frac{\pi}{b_n} \int_{h=0}^\infty h \left( \int_{z=\frac{t-b_n}{\pi h}}^{\frac{t+\epsilon}{\pi h}} + \int_{\frac{t+\epsilon}{\pi h}}^\infty \right) \bar{K} \left( \frac{t - \pi h z}{b_n} \right)^2 g(z, h) dz dh = I_1 + I_2.$$

For  $I_2$ , squaring the upper bound on  $\bar{K}$  given in (14) and using that for  $\pi h z > t + \epsilon > t + b_n$ , we have  $(t - \pi h z)/b_n < -1$

$$\begin{aligned} I_2 &\leq b_n^{-1} \int_{h=0}^\infty \pi h \int_{z=\frac{t+\epsilon}{\pi h}}^\infty b_n(z h \pi - t - b_n)^{-1} g(z, h) dz dh \\ &\leq \int_{h=0}^\infty \pi h (t + \epsilon - t - b_n)^{-1} \int_{z=\frac{t+\epsilon}{\pi h}}^\infty g(z, h) dz dh \\ &= \frac{1}{\epsilon - b_n} \int_{h=0}^\infty \pi h \int_{z=\frac{t+\epsilon}{\pi h}}^\infty g(z, h) dz dh \leq \frac{2}{\epsilon} \pi E_g[H] \end{aligned}$$

for all  $n$  sufficiently large. Since  $E_g[H] < \infty$ ,  $I_2$  is bounded as  $n \rightarrow \infty$ . For  $I_1$ , we have for any  $c < -1$  and  $n$  sufficiently large

$$\begin{aligned} I_1 &= b_n^{-1} \int_{h=0}^\infty \pi h \int_{z=\frac{t-b_n}{\pi h}}^{\frac{t+\epsilon}{\pi h}} \bar{K} \left( \frac{t - z h \pi}{b_n} \right)^2 g(z, h) dz dh \\ &= \int_{h=0}^\infty \int_{v=-\frac{\epsilon}{b_n}}^1 \bar{K}(v)^2 g \left( \frac{t - b_n v}{\pi h}, h \right) dv dh \\ &= \int_{h=0}^\infty \left[ \int_{v=-\frac{\epsilon}{b_n}}^c + \int_{v=c}^1 \right] \bar{K}(v)^2 g \left( \frac{t - b_n v}{\pi h}, h \right) dv dh. \end{aligned}$$

For any fixed  $c$ , the second term is clearly bounded by a constant. Taking  $c < -1$  sufficiently small, by right continuity of  $\tau_q$  at  $t$ , the first term becomes

$$\begin{aligned} \int_{h=0}^\infty \int_{-\epsilon/b_n}^c \bar{K}(v)^2 g \left( \frac{t - b_n v}{\pi h}, h \right) dv dh &= \int_{-\epsilon/b_n}^c \bar{K}(v)^2 \tau_q(t - b_n v) dv \\ &= \int_{h=0}^\infty g \left( \frac{t}{\pi h}, h \right) dh \int_{-\epsilon/b_n}^c \frac{1}{-v} dv + O(1) = \tau_q(t) \ln \left( b_n^{-1} \right) + O(1) \end{aligned}$$

using (15), the fact that  $\epsilon$  can be chosen arbitrarily small in this argument and dominated convergence. The exact same method can be used for the squared radius. The result for  $I_2$  is  $2/\epsilon$  since there is no  $\pi h$  term in the integral. The result for  $I_1$  will be exactly the same as the result for the volume where  $\tau_q(t)$  is  $g_Z(t)$ . Together, these again lead to (17). □

We can now prove the following theorem.

**Theorem 3.3** *Under Condition 3.1 and the assumptions of Lemma 3.2, for  $b = b_n \downarrow 0$  as  $n \rightarrow \infty$ , for  $t > 0$*

$$\text{MSE}(\tilde{N}_n(t)) = \frac{1}{4} b_n^4 N''(t)^2 \left( \int u^2 K(u) du \right)^2 + \frac{\tau_q(t) \ln(b_n^{-1})}{n} + O(n^{-1}) + o(b_n^4).$$

Moreover,

$$\sqrt{\frac{n}{\ln n}} (\tilde{N}_n(t) - N(t)) \rightarrow^D N(0, \tau_q(t)/4) \text{ as } n \rightarrow \infty.$$

This holds for both the squared radius and the volume distribution.

**Proof** The MSE part immediately follows from the asymptotic bias derived in (10) combined with Lemma 3.2. The asymptotic distribution result follows by also using the central limit theorem for i.i.d. random variables with infinite variance; see Chow and Teicher (1988) Theorem 4 on p. 305). □

As a consequence, the asymptotically MSE optimal bandwidth is given by

$$b_n = n^{-\frac{1}{4}} \tau_q(t)^{\frac{1}{4}} \left( |N''(t)| \int u^2 K(u) du \right)^{-\frac{1}{2}},$$

yielding

$$\lim_{n \rightarrow \infty} \frac{n}{\ln n} \text{MSE}(\tilde{N}_n(t)) = \frac{1}{4} \tau_q(t).$$

The MSE of the initial plug-in estimator  $N_n$  defined in (7) is infinite, because its variance is infinite. A notable property of the estimator  $\tilde{N}_n(t)$  is that as long as the bandwidth tends to zero at rate  $n^{-1/4}$ , the asymptotic MSE does not depend on the constant that is chosen in the bandwidth. In Sect. 5.4 of McGarrity (2013), a numerical simulation illustrates the difference between the initial plug-in estimator  $N_n$  and the smoothed estimator  $\tilde{N}_n$  and how the bandwidth impacts the estimates of the distribution functions of the squared radius and volume. In other contexts, including the estimation of the density function that will be considered in Sect. 4, choosing this constant optimally is often a delicate matter.

Another notable fact is the value of the asymptotic MSE in relation to asymptotic distribution results of the empirical (non-smoothed) estimator  $N_n$  and the isotonic inverse estimator studied by McGarrity et al. (2014). Both estimators are asymptotically unbiased, and normal with variance  $\tau_q(t)$  and  $\tau_q(t)/2$ , respectively (both rescaled with rate  $\sqrt{n/\ln n}$ ).

In view of Theorem 3.3, these estimators are comparable to smoothed estimators with bandwidths of order  $n^{-1}$  and  $n^{-1/2}$ , respectively. Taking these small bandwidths results in asymptotically unbiased estimators. Smoothing a bit more (using  $b_n \sim n^{-1/4}$ ) decreases the variance but still results in an (asymptotically) unbiased estimator. Taking a larger bandwidth will make the bias term in the MSE the dominating one, and increase the asymptotic MSE.

The derivative of this smooth estimator may be taken to give an estimate of the density. This will be explored further in the next section.

### 4 Smooth density estimators

In order to obtain estimators for the densities  $f$ , we study the derivative of  $\tilde{N}_n$  as given in (9). Contrary to the non-smoothed estimator  $N_n$ , the estimator  $\tilde{N}_n$  can be differentiated to obtain an estimator for the derivative  $\nu = N'$ . We define the estimator of this derivative as

$$\tilde{\nu}_n(t) = \frac{d}{dt} \tilde{N}_n(t) = \frac{1}{nb_n^2} \sum_{i=1}^n \int K' \left( \frac{t-s}{b_n} \right) [Z_i - q(H_i; s)]^{-\frac{1}{2}} 1_{[Z_i > q(H_i; s)]} ds. \tag{18}$$

Note that just as in the setting of estimating  $N(t)$ , the expectation of the estimators for the function  $\nu$  related to the two choices of  $q$  in (3) can be dealt with at once. To this end, we need

**Condition 4.1** The function  $N$  is three times continuously differentiable at  $t$ .

Under Condition 4.1, we can write:

$$E [\tilde{\nu}_n(t)] = \frac{1}{b_n} \int N(t - b_n u) K'(u) du = \nu(t) + \frac{1}{2} b_n^2 \nu''(t) \int u^2 K(u) du + o(b_n^2) \tag{19}$$

for  $n \rightarrow \infty$ . In order to obtain the asymptotic variance of the estimators for the squared radius and volume, we use representations (12) and (13) to write

$$\tilde{\nu}_n^{sr}(t) = \frac{1}{nb_n^{\frac{3}{2}}} \sum_{i=1}^n \bar{K}' \left( \frac{t - Z_i}{b_n} \right), \quad \tilde{\nu}_n^{vol}(t) = \frac{1}{nb_n^{\frac{3}{2}}} \sum_{i=1}^n \sqrt{\pi H_i} \bar{K}' \left( \frac{t - \pi H_i Z_i}{b_n} \right). \tag{20}$$

For the variances of the estimators, we have the following lemma.

**Lemma 4.2** Fix  $t > 0$  and suppose Condition 4.1 holds and function  $\tau_q$  defined in (16) is right continuous at  $t$ . Let  $K$  be a continuously differentiable symmetric probability density with support  $[-1, 1]$ . Then, as  $b_n \downarrow 0$ ,

$$\text{Var}(\tilde{\nu}_n(t)) = \tau_q(t) \frac{\int \bar{K}'(u)^2 du}{nb_n^2} + O\left((nb_n)^{-1}\right). \tag{21}$$

This result holds for the squared radius as well as the volume.

**Proof** Considering the volume, by (20),  $nb_n^2 \text{Var}(\tilde{\nu}_n^{vol}(t))$  equals

$$\frac{\pi}{b_n} \left\{ E \left[ H_1 \bar{K}' \left( \frac{t - \pi H_1 Z_1}{b_n} \right)^2 \right] - \left( E \left[ \sqrt{H_1} \bar{K}' \left( \frac{t - \pi H_1 Z_1}{b_n} \right) \right] \right)^2 \right\}.$$

Using the asymptotic bias (19) and Condition 4.1, it follows that the second term in the above expression is  $o(1)$  for  $n \rightarrow \infty$ . Because  $K(\pm 1) = K'(\pm 1) = 0$  for the

kernel function we consider, for  $v < -1$

$$\bar{K}'(v) = \frac{1}{2} \int_{-v-1}^{-v+1} K(u+v)u^{-3/2} du,$$

implying that for  $v < -1$

$$\frac{1}{4}(-v+1)^{-3} \leq \bar{K}'(v)^2 \leq \frac{1}{4}(-v-1)^{-3}.$$

This bound, with boundedness of  $\bar{K}'$ , imply that  $\bar{K}'$  is square integrable. Now note that

$$\begin{aligned} \frac{\pi}{b_n} E \left[ H_1 \bar{K}' \left( \frac{t - \pi H_1 Z_1}{b_n} \right)^2 \right] &= \frac{\pi}{b_n} \int_{h=0}^{\infty} h \int_{z=0}^{\infty} \bar{K}' \left( \frac{t - \pi h z}{b_n} \right)^2 g(z, h) dz dh \\ &= \int_{h=0}^{\infty} \int \bar{K}'(u)^2 g \left( \frac{t - b_n u}{\pi h}, h \right) du dh = \int \bar{K}'(u)^2 \tau_q(t - b_n u) du \\ &= \tau_q(t) \int \bar{K}'(u)^2 du + o(1), \end{aligned}$$

where we use dominated convergence and right continuity of  $\tau_q$  at  $t$ . □

As in Sect. 3, we can define the mean squared error of the estimator by

$$\text{MSE}(\tilde{v}_n(t)) = (E_g [\tilde{v}_n(t)] - v(t))^2 + \text{Var}(\tilde{v}_n(t)).$$

As the global behavior of the density estimator as a function is maybe even more of interest than its local behavior (more so than for the estimator of the distribution function), the mean integrated squared error,

$$\text{MISE}(\tilde{v}_n) = \int \text{MSE}(\tilde{v}_n(t)) dt$$

is also interesting for  $\tilde{v}_n$ . We have the following result.

**Theorem 4.3** Fix  $t > 0$ . Under the assumptions of Lemma 4.2 and Condition 4.1, as  $n \rightarrow \infty$  and  $b_n \downarrow 0$ ,

$$\text{MSE}(\tilde{v}_n(t)) = \tau_q(t) \frac{\int \bar{K}'(u)^2 du}{nb_n^2} + \frac{1}{4} b_n^4 v''(t)^2 \left( \int u^2 K(u) du \right)^2 + o \left( \frac{1}{nb_n^2} \right) + o(b_n^4)$$

for the squared radius and volume. If  $v$  has a uniformly bounded third derivative and  $f$  bounded support in  $[0, \infty)^2$ , then

$$\begin{aligned} \text{MISE}(\tilde{v}_n) &= \int \tau_q(t) dt \cdot \frac{\int \bar{K}'(u)^2 du}{nb_n^2} + \frac{1}{4}b_n^4 \int v''(t)^2 dt \cdot \left( \int u^2 K(u) du \right)^2 \\ &\quad + o\left(\frac{1}{nb_n^2}\right) + o(b_n^4). \end{aligned}$$

**Proof** The asymptotics of the MSE immediately follows from (19) and Lemma 4.2. For the MISE, note that bounded support of  $f$  implies bounded support of  $g$  which in turn implies bounded support of  $\tau_q$  and  $v$ . □

From Theorem 4.3, we infer that the asymptotic M(I)SE- optimal bandwidth corresponds to a balance of the two terms, leading to  $b_n \sim n^{-1/6}$ . Taking  $b_n = \alpha n^{-1/6}$ , the asymptotically MISE-optimal choice for  $\alpha$  is given by

$$\alpha_{opt} = \left[ \frac{2 \int \tau_q(t) dt \int \bar{K}'(u)^2 du}{\int v''(t)^2 dt \left( \int u^2 K(u) du \right)^2} \right]^{\frac{1}{6}}. \tag{22}$$

Taking the asymptotically optimal bandwidth leads to

$$\begin{aligned} &\lim_{n \rightarrow \infty} n^{\frac{2}{3}} \text{MISE}(\tilde{v}_n) \\ &= 3 \left[ \frac{1}{4} \int \tau_q(t) dt \left( \int (v''(t))^2 dt \right)^{1/2} \int \bar{K}'(u)^2 du \int u^2 K(u) du \right]^{\frac{2}{3}}. \end{aligned}$$

Unlike the asymptotically MSE-optimal bandwidth for estimating  $N(t)$ , which is dependent only on the sample size, finding the MISE-optimal bandwidth for estimating the pdf must be done more carefully. This bandwidth also depends on the second derivative of the function being estimated, as well as on integrals related to the kernel.

For the kernel-dependent constants in (22) based on the biweight kernel,

$$\int u^2 K(u) du = \frac{1}{7} \text{ and } \int \bar{K}'(u)^2 du = \frac{25}{8}. \tag{23}$$

Details on the latter are given in the appendix. Furthermore, note that for the squared radius (to which we restrict ourselves for the moment)

$$\int_0^\infty \tau_q(t) dt = \int_0^\infty \int_0^\infty g(t, h) dh dt = 1,$$

so that considering the squared radius and using the biweight kernel, the asymptotically MISE optimal bandwidth is given by

$$b_n = \left[ \frac{1225}{4 \int v''(t)^2 dt} \right]^{\frac{1}{6}} n^{-1/6}.$$

We propose a reference family method to come to a concrete choice of the bandwidth parameter. This method actually imposes a *parametric* model for the observations to estimate the unknown quantity  $\int (v''(t))^2 dt$ . As a rather natural (and as will be seen convenient) choice for the underlying density  $f$ , we take

$$f(x, h) = \frac{1}{\sigma_X \sigma_H} e^{-h/\sigma_H - x/\sigma_X} 1_{[0, \infty)^2}(x, h).$$

This means that  $X$  and  $H$  are independent and exponentially distributed. It is straightforward to show that then

$$g(z, h) = \frac{1}{\sigma_X \sigma_H} e^{-h/\sigma_H - z/\sigma_X} 1_{[0, \infty)^2}(z, h)$$

so  $(X, H) =^D (Z, H)$ . This is a very specific property of the underlying density  $f$ . It also immediately gives MLE's for  $\sigma_X$  and  $\sigma_H$ : the respective sample means of the  $Z_i$ 's and  $H_i$ 's. Using the joint exponential distribution, the function  $N$  and its derivatives can be derived (for details see the appendix), leading to

$$\int v''(t)^2 dt = \frac{\pi}{2} \sigma_X^{-6}. \tag{24}$$

Estimating  $\sigma_X$  by  $\bar{Z}_n$ , the final automatic bandwidth choice becomes

$$b_n = \left[ \frac{1225}{4\pi/2} \right]^{\frac{1}{6}} \sigma_X n^{-1/6} \approx 2.4 \bar{Z}_n n^{-1/6}. \tag{25}$$

For the volume density, we take a similar approach, taking as reference family again a set of product densities with scale-family marginals to be chosen later, i.e.,

$$f_{\sigma_X, \sigma_H}(x, h) = \frac{1}{\sigma_X \sigma_H} f_X(x/\sigma_X) f_H(h/\sigma_H).$$

Using that  $m_f^+ = \sigma_X^{-1} \int \sqrt{x} f_X(x/\sigma_X) dx = \sqrt{\sigma_X} \int \sqrt{x} f_X(x) dx$ , we have

$$\begin{aligned} g_{\sigma_X \sigma_H}(z, h) &= \frac{1}{2\sigma_X^{3/2} \sigma_H \int \sqrt{x} f_X(x) dx} \int_{x=z}^{\infty} \frac{f_X(x/\sigma_X) f_H(h/\sigma_H)}{\sqrt{x-z}} dx \\ &= \frac{1}{\sigma_H} f_H(h/\sigma_H) \frac{1}{2 \int \sqrt{x} f_X(x) dx \sqrt{\sigma_X}} \int_{x=z/\sigma_X}^{\infty} \frac{f_X(x)}{\sqrt{\sigma_X} \sqrt{x-z/\sigma_X}} dx \\ &= \frac{1}{\sigma_H} f_H(h/\sigma_H) \frac{1}{\sigma_X} f_Z(z/\sigma_X) \end{aligned}$$

where

$$f_Z(z) = \frac{1}{2 \int \sqrt{x} f_X(x) dx} \int_{x=z}^{\infty} \frac{f_X(x)}{\sqrt{x-z}} dx.$$

This yields  $E_{g_{\sigma_X\sigma_H}}[H] = \sigma_H \int h f_H(h) dh$  and, using Fubini,

$$E_{g_{\sigma_X\sigma_H}}[Z] = \sigma_X \int_z z f_Z(z) dz = \sigma_X \text{Beta}(2, 1/2) \frac{\int x^{3/2} f_X(x) dx}{2 \int \sqrt{x} f_X(x) dx} = \sigma_X \frac{2 \int x^{3/2} f_X(x) dx}{3 \int \sqrt{x} f_X(x) dx},$$

immediately leading to moment estimators for  $\sigma_H$  and  $\sigma_X$ :

$$\hat{\sigma}_H = \frac{\bar{H}_n}{\int h f_H(h) dh} \text{ and } \hat{\sigma}_X = \bar{Z}_n \frac{3 \int \sqrt{x} f_X(x) dx}{2 \int x^{3/2} f_X(x) dx}.$$

Also note that

$$\int_t \tau_q(t) dt = \int_t \int_h g_{\sigma_X\sigma_H}(t/(\pi h), h) dh dt = \int_h \pi h g_{\sigma_H}(h) dh = \pi \sigma_H \int_h h f_H(h) dh.$$

In the appendix [see (32) and (33)], it is shown that

$$\int_0^\infty v''(t)^2 dt = \frac{\int_u (\int_x x^{-3} f_X(x) f_H''(\frac{u}{x}) dx)^2 du}{4\pi^3 \sigma_X^6 (\int \sqrt{x} f_X(x) dx)^2 \sigma_H^5} \tag{26}$$

From this expression, it is clear that the exponential marginal densities as used for the squared radius density will not lead to a useful method, because the integral in the numerator will be infinite. The densities should tend to zero sufficiently fast near zero. For practical reasons, we take Gamma distributions for both marginals:

$$f_X(x) = \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-x} \text{ and } f_H(h) = \frac{1}{\Gamma(\beta)} x^{\beta-1} e^{-x}.$$

This choice leads to concrete expressions for the various quantities derived. Indeed,

$$\int_t \tau_q(t) dt = \pi \beta \sigma_H, \hat{\sigma}_H = \frac{\bar{H}_n}{\beta} \text{ and } \hat{\sigma}_X = \bar{Z}_n \frac{3\Gamma(\alpha + 1/2)}{2\Gamma(\alpha + 3/2)} = \frac{3\bar{Z}_n}{2\alpha + 1}. \tag{27}$$

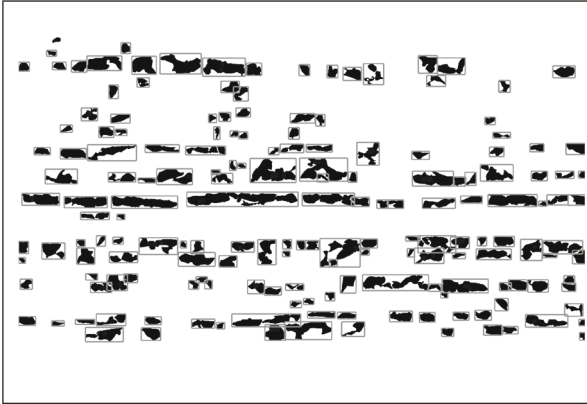
Also using (23),

$$b_n = \left[ \frac{1225\pi\beta\sigma_H}{4 \int v''_{\alpha,\beta}(t)^2 dt} \right]^{\frac{1}{6}} n^{-1/6}.$$

In the Appendix, it is shown that taking  $\alpha = 11$  and  $\beta = 7$  (these values will be chosen for the data example in the next section) leads to the numerical approximation

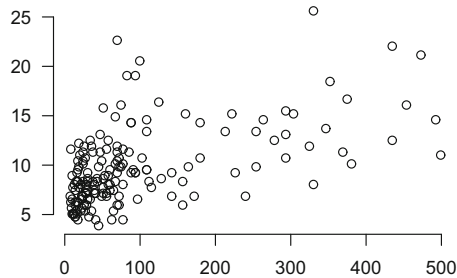
$$\int_0^\infty v''_{11,7}(t)^2 dt \approx 1.4\sigma_X^{-6}\sigma_H^{-5} \times 10^{-11}.$$





**Fig. 4** Bounding boxes taken to be the rectangles around the observable 2D features of interest in the microstructure. Sample size is  $n = 175$

**Fig. 5** Scatter plot of the 171 pairs  $(Z_i, H_i)$



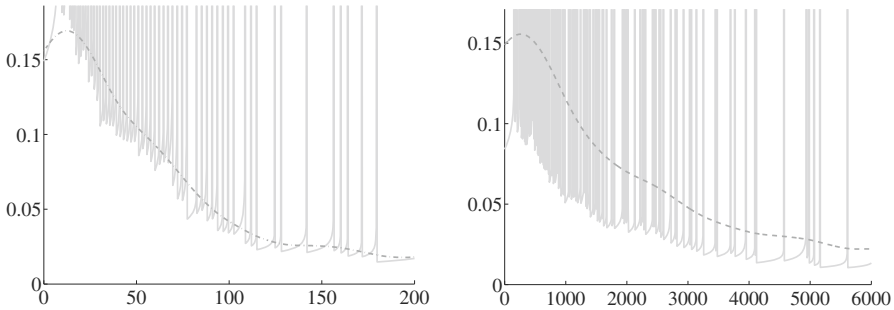
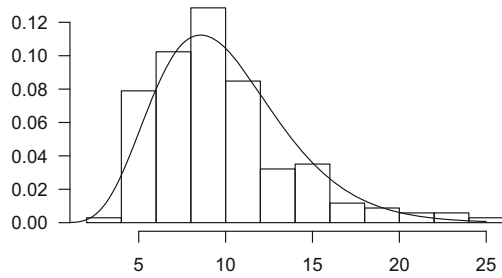
Substituting (27) and for  $\alpha = 11$  and  $\beta = 7$ , this leads to the following bandwidth choice:

$$b_n = \left[ \frac{1225\pi\beta\sigma_H}{4 \int v''_{11,7}(t)^2 dt} \right]^{\frac{1}{6}} n^{-1/6} \approx 280\sigma_X\sigma_H n^{-1/6} \approx 5.2\bar{Z}_n\bar{H}_n n^{-1/6}. \quad (28)$$

## 5 Application to a steel microstructure

Figure 4 shows the binary image of a steel microstructure obtained at TATA Steel. The grey rectangles correspond to the bounding boxes of the features of interest and are used as the observed rectangles on the cut plane within the oriented cylinder model. Bounding boxes were chosen because they are well-defined rectangular objects around the interesting structures in the image, which themselves are certainly not perfectly rectangular. From these bounding boxes, the pairs  $(Z_i, H_i)$  are distilled. As such, a total of 175 pairs are obtained, of which 4 are ignored because the corresponding  $Z$ -values are way out of range. Figure 5 shows the scatter plot of the data obtained. For a complete discussion on the choice and effects of the bounding box on the estimation results, see Sect. 5.4 in McGarrity (2013).

**Fig. 6** Probability histogram of the 171 measured heights with maximum likelihood fitted Gamma density



**Fig. 7** Estimates  $N_n$  and  $\tilde{N}_n$  of  $N$  for the squared radius (left,  $b_n = 30$ ) and volume (right,  $b_n = 900$ ). The data correspond to those depicted in Fig. 5

Figure 7 shows the empirical plug-in estimator  $N_n$  defined in (7) as well as the kernel estimator  $\tilde{N}_n$  for  $N$  based on the microstructure data. In Fig. 8, the isotonic estimators as well as the kernel estimators for the distribution functions  $F$  of the squared radius and the volume are given. For these estimates, we use relation (4), taking

$$\frac{1}{n} \sum_{i=1}^n Z_i^{-1/2} \tag{29}$$

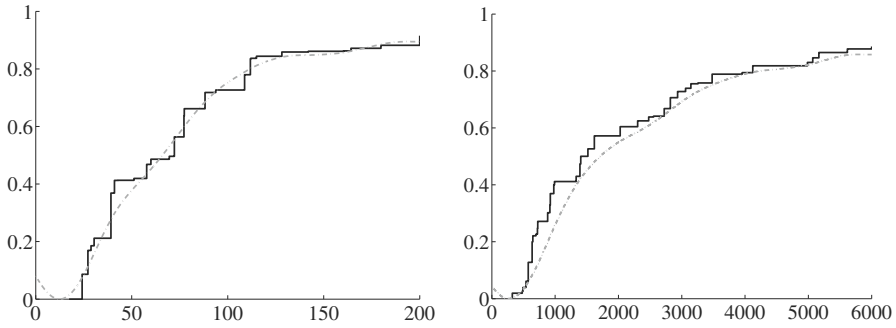
as (consistent) estimator of  $N(0)$ . From these pictures, it is clear that the isotonic and kernel smoothed estimators are quite close.

As indicated before, special interest is in the estimates of the functions  $\nu$  for both the squared radii and the volumes. For the squared radii,  $Z_n \approx 99.7$ , leading via (25) to bandwidth choice

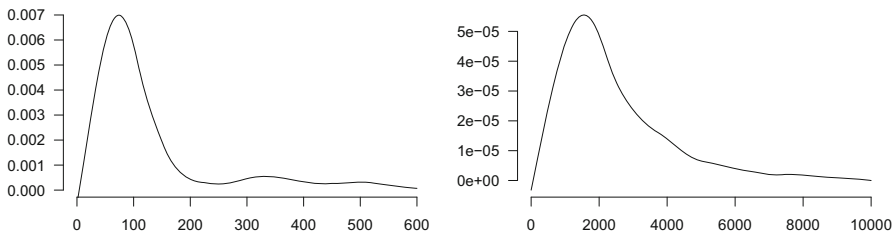
$$b_n = 2.4\bar{Z}_n n^{-1/6} \approx 102.$$

The left panel of Fig. 9 shows the resulting kernel estimate for the probability density  $f$  (related to  $\nu$  via (6), also using (29) as estimator for  $N(0)$ ).

For the volume density, the reference family with Gamma densities requires an a priori choice for the respective shape parameters  $\alpha$  and  $\beta$ . As the independence between  $H$  and  $Z$  in the reference family implies that the observed  $H$ -values could be viewed as sample from the Gamma distribution of interest, we computed the MLE



**Fig. 8** Smooth kernel estimate for the distribution function  $F$  of the squared radius (left,  $b_n = 30$ ) and volume (right,  $b_n = 900$ ). The step functions are the isotonic estimates studied in McGarrity et al. (2014). Data are those of Fig. 5



**Fig. 9** Estimates for the density  $f$  of the squared radius (left,  $b_n = 102$ ) and volume (right,  $b_n = 2130$ ). Data are those of Fig. 5

based on the observed values, resulting in  $\beta \approx 7$ . See Fig. 6, showing a surprisingly good Gamma fit to the observed heights. For the squared radii, we more arbitrarily choose a Gamma density that guarantees convergence of the needed integrals,  $\alpha = 11$ . Based on the measured data ( $\bar{H}_n = 9.7$ ,  $\bar{Z}_n = 99.7$ ), we obtain via (27)

$$\hat{\sigma}_H = \frac{\bar{H}}{\beta} \approx 1.4 \text{ and } \hat{\sigma}_X = \frac{3\bar{Z}}{2\alpha + 1} \approx 13,$$

leading via (28) to bandwidth choice

$$b_n = 5.2\bar{Z}_n\bar{H}_n n^{-1/6} \approx 2130.$$

Figure 9 shows the resulting kernel estimate of the probability density of the volume.

### 6 Discussion

We consider estimation of distributions within the oriented cylinder model. This is an extension of the classical Wicksell model. The estimators considered up till now in the oriented cylinder model do not include estimators that can be differentiated to obtain estimators of the probability densities. In this paper, we introduce smooth estimators

that can be used in that way. We restrict attention to the estimation of the distribution of the (squared) radius of the cylinder base and the volume of the cylinders. Other aspects can also be of interest and studied in the same way based on relation (2). For example, the distributions of the aspect ratio  $R = \sqrt{X}/H$  and the surface area  $S = 2\pi(X + \sqrt{X}H)$ . The corresponding functions  $q$  are given by

$$q(h; t) = \begin{cases} h^2 t^2 & \text{aspect ratio: } T = \frac{\sqrt{X}}{H}, \\ \left[ \sqrt{\frac{h^2}{4} + \frac{t}{2\pi}} - \frac{h}{2} \right]^2 & \text{surface area: } T = 2\pi(X + \sqrt{X}H). \end{cases}$$

Besides the ones considered in this paper, there are other natural choices to estimate  $N$ . One would be to use the isotonic estimator of Groeneboom and Jongbloed (1995) as initial estimator and smoothing this. The other is to isotonize the estimator  $\tilde{N}_n$ . The asymptotic theory for the first type of estimator will be much harder to develop. Moreover, the conjecture is that the resulting estimators will not be better asymptotically than  $\tilde{N}_n$ . See Groeneboom et al. (2010) for a study of smoothed isotonic estimators in the current status model, where such ‘smoothed isotonic’ and ‘isotonized smooth’ estimators are studied and shown to have similar asymptotic behavior.

**Acknowledgements** This research was carried out under the project number M41.10.09330 in the framework of the Research Program of the Materials innovation institute M2i ([www.m2i.nl](http://www.m2i.nl)). We thank Piet Kok, Koen Lammers and Karin de Moel (TATA steel) for fruitful discussions. René Swarttouw we thank for his help with Maple. We also thank the referees and (associate) editor for their beneficial comments.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

### A Computational details

In this section, we give some details on the computations needed to apply the described reference family method to choose the bandwidth. Based on the biweight kernel, the integral of the squared second derivative of the function  $\bar{K}$  defined in (11) is needed. Note that

$$\bar{K}(v) = \begin{cases} 0 & \text{for } v \geq 1, \\ \frac{4(4v^2 + 10v + 7)(1 - v)^{5/2}}{21} & \text{for } |v| < 1, \\ \frac{4(4v^2 + 10v + 7)(1 - v)^{5/2}}{21} - \frac{4(v + 1)^2(4v^2 - 10v + 7)\sqrt{-v - 1}}{21} & \text{for } v \leq -1. \end{cases} \quad (30)$$

Its derivative  $\bar{K}'$  is given by

$$\bar{K}'(v) = \begin{cases} 0 & \text{for } v \geq 1, \\ \frac{2(12v^3 + 6v^2 - 13v - 5)\sqrt{1-v}}{7} & \text{for } |v| < 1, \\ \frac{2(12v^3 + 6v^2 - 13v - 5)\sqrt{1-v}}{7} - \frac{2(12v^2 - 18v + 5)(v + 1)\sqrt{-v-1}}{7} & \text{for } v \leq -1. \end{cases} \tag{31}$$

Although it is not immediately clear from this expression, some algebra shows that for  $v \rightarrow -\infty$ ,  $\bar{K}'(v) \sim v^{-3/2}/2$ , implying square integrability. Moreover, using explicit expression (31), it follows that  $\int \bar{K}'(u)^2 du = 25/8$  as stated in (23).

For deriving (24), note that  $v'' = N'''$  and for the squared diameter we have for  $t > 0$

$$\begin{aligned} N(t) &= \frac{1}{\sigma_X \sigma_H} \int_{h=0}^{\infty} \int_{z=t}^{\infty} (z-t)^{-1/2} e^{-z/\sigma_X - h/\sigma_H} dz dh \\ &= \frac{1}{\sigma_X} \int_t^{\infty} (z-t)^{-1/2} e^{-z/\sigma_X} dz \\ &= \frac{1}{\sigma_X} e^{-t/\sigma_X} \int_0^{\infty} z^{-1/2} e^{-z/\sigma_X} dz = \sigma_X^{-1/2} \Gamma(1/2) e^{-t/\sigma_X} = \sigma_X^{-1/2} \sqrt{\pi} e^{-t/\sigma_X}. \end{aligned}$$

Therefore, for  $t > 0$

$$\begin{aligned} v''(t) &= -\sigma_X^{-7/2} \sqrt{\pi} e^{-t/\sigma_X}, \text{ yielding } \int_0^{\infty} v''(t)^2 dt \\ &= \pi \sigma_X^{-7} \int_0^{\infty} e^{-2t/\sigma_X} dt = \frac{\pi}{2} \sigma_X^{-6}. \end{aligned}$$

For the volume distribution, (26) follows by first noting that

$$\begin{aligned} \frac{N(t)}{N(0)} &= P_f(\pi XH > t) = \int_{x=0}^{\infty} \int_{h=t/(\pi x)}^{\infty} \frac{f_X(x/\sigma_X) f_H(h/\sigma_H)}{\sigma_X \sigma_H} dh dx \\ &= \int_{x=0}^{\infty} f_X(x) \int_{h=t/(\pi \sigma_H \sigma_X x)}^{\infty} f_H(h) dh dx = 1 - \phi\left(\frac{t}{\pi \sigma_H \sigma_X}\right) \end{aligned}$$

with

$$\phi(u) = \int_0^{\infty} f_X(x) F_H(u/x) dx \text{ so that } \phi'''(u) = \int_0^{\infty} x^{-3} f_X(x) f_H''(u/x) dx \tag{32}$$

This leads to

$$v''(t) = N'''(t) = -\frac{N(0)}{(\pi \sigma_H \sigma_X)^3} \phi''' \left( \frac{t}{\pi \sigma_H \sigma_X} \right)$$

Noting that

$$N(0) = \frac{\pi}{2 \int_0^\infty \sqrt{x} f_{\sigma_X}(x) dx} = \frac{\pi}{2\sqrt{\sigma_X} \int_0^\infty \sqrt{x} f_X(x) dx}$$

we obtain

$$\int_0^\infty v''(t)^2 dt = \frac{\int_0^\infty \phi''' \left( \frac{t}{\pi\sigma_H\sigma_X} \right)^2 dt}{4\pi^4\sigma_X^7 \left( \int \sqrt{x} f_X(x) dx \right)^2 \sigma_H^6} = \frac{\int_0^\infty \phi'''(t)^2 dt}{4\pi^3\sigma_X^6 \left( \int \sqrt{x} f_X(x) dx \right)^2 \sigma_H^5} \tag{33}$$

Choosing  $f_X$  and  $f_H$  Gamma with respective shape parameters  $\alpha$  and  $\beta$  (making  $\phi$  dependent on  $\alpha$  and  $\beta$ ) and unit scale parameters, this results in

$$\int_0^\infty v''_{\alpha,\beta}(t)^2 dt = \frac{\Gamma(\alpha)^2 \int_0^\infty \phi'''_{\alpha,\beta}(u)^2 du}{4\pi^3\sigma_X^6 \Gamma(\alpha + 1/2)^2 \sigma_H^5}$$

Note that

$$\begin{aligned} \phi'''_{\alpha,\beta}(u) = & \frac{1}{\Gamma(\alpha)\Gamma(\beta)} \int_0^\infty \left( (\beta - 1)(\beta - 2) - 2(\beta - 1) \frac{u}{x} \right. \\ & \left. + \left( \frac{u}{x} \right)^2 \right) \left( \frac{u}{x} \right)^{\beta-3} x^{\alpha-4} e^{-x-u/x} dx \end{aligned}$$

The integral in this expression can be further evaluated to make its dependence on  $\alpha$  and  $\beta$  more explicit, but as we will choose specific  $\alpha$  and  $\beta$ , we will approximate the integral numerically. Taking  $\beta = 7$  and  $\alpha = 11$ , Fig. 3 shows the numerical approximation of  $\phi_{11,7}$ , resulting in  $\int_0^\infty \phi'''_{11,7}(u)^2 du \approx 1.9 \times 10^{-8}$ . This leads to  $\int_0^\infty v''_{11,7}(t)^2 dt \approx 1.4\sigma_X^{-6}\sigma_H^{-5} \times 10^{-11}$ .

### References

Chow YS, Teicher H (1988) Probability theory: independence, interchangeability, martingales. Springer texts in statistics. Springer, Berlin

Feuerverger A, Hall P (2000) Methods for density estimation in thick-slice versions of Wicksell’s problem. J Am Stat Assoc 95(450):535–546

Groeneboom P, Jongbloed G (1995) Isotonic estimation and rates of convergence in Wicksell’s problem. Ann Stat 23(5):1518–1542

Groeneboom P, Jongbloed G, Witte B (2010) Maximum smoothed likelihood estimation and smoothed maximum likelihood estimation in the current status model. Ann Stat 38(1):352–387

Hall P, Smith RL (1988) The kernel method for unfolding sphere size distributions. J Comput Phys 74:409–421

McGarrity KS (2013) Stereological Estimation of Anisotropic Microstructural Features, Ph.D. thesis, Delft University of Technology. <https://doi.org/10.4233/uuid:c548f1eb-b44a-4192-a309-8271c4b36308>.

McGarrity KS, Jongbloed G, Sietsma J (2014) Nonparametric inference in a stereological model with oriented cylinders applied to dual phase steel. Ann Appl Stat 8(4):2538–2566

Joachim O, Frank M (2000) Statistical analysis of microstructures in materials science. Wiley, New Jersey

- Sen B, Woodroffe M (2011) Bootstrap confidence intervals for isotonic estimators in a stereological problem. *Bernoulli* 18:1249–1266
- Taylor CC (1983) A new method for unfolding sphere size distributions. *J Microsc* 132(1):57–66
- van Es B, Hoogendoorn A (1990) Kernel estimation in Wicksell's corpuscle problem. *Biometrika* 77(1):139–145
- Wicksell SD (1925) A mathematical study of a biometric problem. *Biometrika* 17(1/2):84–99

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.