# ORIENTATION ANALYSIS OF STOCHASTIC FIBRE SYSTEMS WITH AN APPLICATION TO PAPER RESEARCH

SALME KÄRKKÄINEN

# Abstract

Two methods for estimating the orientation distribution of fibres from greyscale or binary images have been developed. The approach is based on both stereological reasoning in the feature extraction from the image and on the interpretation of the results in terms of the underlying fibrous structure. As a stereological tool, we use line sampling in several directions. In each direction, the intersection between the image and the sampling line forms a grey-valued function whose variation is measured by the scaled variogram. The connections between the observed scaled variograms and the fibre orientation distribution have been approximated under several parametric image models. The methods are based on these connections. The first method is theoretically addressed to digital greyscale (or binary) images originating from dead leaves models, and the second method, being a refinement of the first one, is planned for binary images. Assuming a parametric model for the orientation distribution, the model parameters can be obtained by the numerical optimization of the weighted sum of squares for both methods. The new orientation analysis techniques have been compared with each other and with a commercially available gradient-based method widely applied in paper technology.

# Acknowledgements

I wish to express my sincere graditude to my supervisor, Professor Antti Penttinen, for introducing me into the research field of stochastic geometry, for his expert supervision and constant support during my work. My warm graditude also goes to Professor Eva B. Vedel Jensen, University of Aarhus, for her expert guidance in stereology. Special thanks are also due to Professor Dominique Jeulin, Ecole des Mines de Paris, Professor Nikolai G. Ushakov and Professor Alexandra P. Ushakova, Norwegian University of Science and Technology, for valuable suggestions and discussion on image analysis. I also appreciate the fruitful discussions with PhD Markku Kellomäki, VTT Processes, PhLic Timo Lappalainen, VTT Processes, and MSc Petri Jetsu, Metso Paper Inc., on problems related to paper physics. I am also grateful to the reviewers, Professor Dietrich Stoyan, TU Bergakademie Freiberg, and Professor Joël Chadœuf, INRA, for their encouraging comments on my work. Mrs. Tuula Blåfield deserves special thanks for her linguistic comments on the introductory part.

This research has been done at the Department of Mathematics and Statistics, Laboratory of Data Analysis, University of Jyväskylä, and at the Statistical Department, University of Aarhus. The financial support was provided by the COMAS Graduate School, the Nordic Foundation and the MaDaMe project (no. 50139) of the Academy of Finland. Additional financial support is due to the Ellen and Artturi Nyyssönen Foundation, Center for Mathematical and Computational Modeling (CMCM), Centre for Mathematical Physics and Stochastics (MaPhySto) funded by the Danish National Research Foundation, INTAS, and VTT Processes. I highly appreciate all of them.

My visits to University of Aarhus had great significance to my work. I am grateful to Professor Eva B. Vedel Jensen for inviting me to collaborate during the years 2000 and 2001. I also thank her, the staff, guests and students at the Statistical Department of University of Aarhus for making my visits so memorable.

I would also like to thank my teachers, colleagues and guests at the Department of Mathematics and Statistics and the Laboratory of Data Analysis of University of Jyväskylä for the inspiring and pleasant atmosphere.

My special thanks are due to Heikki Liimatainen for computer assistance and his support also otherwise - Hegu, thanks.

I would like to express my appreciation to my parents, Aino and Eino, and to my brother, Matti, for their encouragement and support from the very beginning.

Jyväskylä, June 2003

*Salme Kärkkäinen*

# List of Original Publications

This thesis consists of an introductory part, three publications and one technical report listed below.

**A** Kärkkäinen, S., Penttinen, A., Ushakov, N. G. and Ushakova, A. P. (2001). Estimation of orientation characteristic of fibrous material. *Adv. Appl. Prob.* **33**, 559–575.

**B** Kärkkäinen, S. and Jensen, E. B. V. (2001). Estimation of fibre orientation from digital images. *Image Anal. Stereol.* **20**, 199–202.

**C** Kärkkäinen, S., Jensen, E. B. V. and Jeulin, D. (2002). On the orientational analysis of planar fibre systems from digital images. *J. Microsc.* **207**, 69–77.

**D** Kellomäki, M., Kärkkäinen, S., Penttinen, A. and Lappalainen, T. (2003). Determination of fibre orientation distribution from images of fibre networks. Publications of the Laboratory of Data Analysis, No. 4, University of Jyväskylä.

# Contents

# Chapter 1

# Introduction

This thesis deals with statistical orientation analysis for stochastic fibre systems. The problem originates from industrial needs for determining the orientation distribution of paper fibres through digital images. The new analysis methods are essentially based on stochastic geometric modelling, stereology and image analysis. The performance of the methods has been investigated with extensive simulation experiments and with real-world industrial data originating from paper technology.

Fibrous structures appear commonly both in natural objects such as nerves, muscle fibres, wood fibres, and in human-made products such as building materials, textiles, and paper products. The fibrous structure is one object of the quality analysis in the production since it determines a great deal of the properties of the material. As a consequence, much effort has been directed to theoretical and experimental research of fibrous structures in material physics and paper technology.

In contrast to the seemingly regular fibrous structures of woven products, many materials are made up of random fibrous structures. In our approach, randomness consists of three hierarchical levels. The first level is the properties of fibres, such as length and shape. The second level is the spatial (2D and 3D) distribution of fibres: how the fibres are distributed and oriented in the material and with respect to each other. The third level is the observational mechanism which, however, depends on the application and objectives. Fibres are observed as a digital greyscale (or binary) image. In such cases, thickness, overlapping, high density of fibres, grey levels, and digitization degrade the observation of individual fibres, cf. Figures 1.1 and 1.2.

In the modelling of the first two levels, it is natural to consider fibrous structures as stochastic fibre processes. Those random models have been developed in stochastic geometry, which is a science of random structures of

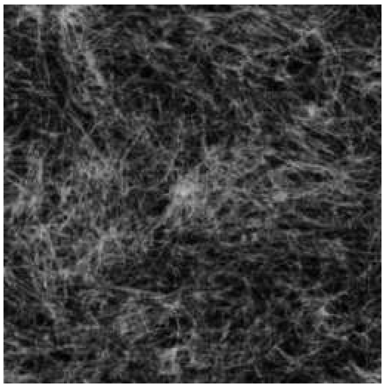

Figure 1.1: A greyscale image of $256 \times 256$ pixels with real size of 8.13 mm $\times$ 8.13 mm from a paper layer stripped by a tape technique (Article D).

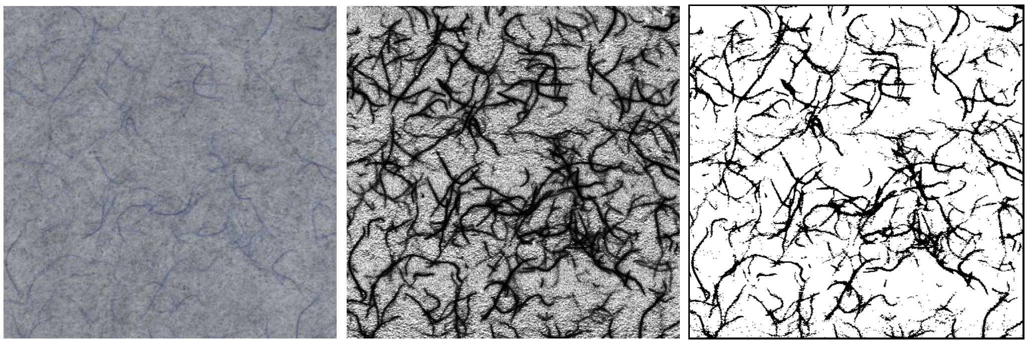

Figure 1.2: **On the left:** An image of the surface of a paper sample in which 0.1 percent of fibres are dyed. The image is of size $500 \times 500$ with real size 25.4 mm $\times$ 25.4 mm. **In the middle:** The left image after contrasting. **On the right:** The middle image after 20 % thresholding (Article D).

geometrical objects, see Stoyan et al. (1995). A convenient model for the orientation distribution is the rose of directions of the fibre process, which is the distribution of the tangent angle in a typical fibre point with respect to a fixed direction. On the third level, instead of stochastic geometric models, random fields induced by the underlying fibre system together with a possibly complex observational mechanism are more relevant. Therefore, tools from statistical image analysis are needed, cf. e.g. Serra (1982), and Glasbey and Horgan (1995). Three types of grey level models for fibre systems, Boolean models, dead leaves models and shot-noise models, have been applied, see Matheron (1975), Jeulin (1989, 1993) and Rice (1977), respectively. See the illustration of the orientation distribution and image models in Figure 4.2. An elliptic fibre orientation model is convenient for paper industry applications.

In the statistical inference on fibre orientation, the direct observation of fibre orientation from a fibre system is usually not possible. Instead, stereological methods have been introduced. Stereology studies the problems of recovering information on 3D or 2D structures when only information in lower dimension is available, cf. Weibel (1979). An integral part of our fibre orientation approach is that fibres are intersected by sampling lines in several directions. It is well-known from earlier studies, for example, by Mecke and Stoyan (1980) that the orientation distribution and the point intensity, the mean number of intersections per length unit, are related when those intersections can be observed. If the available data is a greyscale (or binary) image, the individual fibres and, consequently, the intersection points along sampling lines are not necessarily detectable, see Figures 1.3 and 1.4. Information on the fibre orientation is, however, preserved in the directional variation of grey levels.

Our objective is to develop a method for orientation analysis through digital images according to the following principles: *(i)* the method should be based on solid theory of probability and statistics, *(ii)* the method should be computationally fast, *(iii)* the results should be interpretable in terms of the underlying fibre system, that is, in terms of fibre orientation distribution, *(iv)* one should be able to control statistical uncertainty, *(v)* the method should be applicable in a range of real industrial problems, and *(vi)* the new approach should be competitive compared with the gradient-based orientation analysis method commonly applied in industry, see Erkkilä (1995); Erkkilä et al. (1998).

Two methods for estimating the fibre orientation distribution from digital images are implemented. We assume that the observed fibres are described as long and narrow sets around "central" fibres. The central fibres are typically regarded as line segments and their thickened versions, for example, as

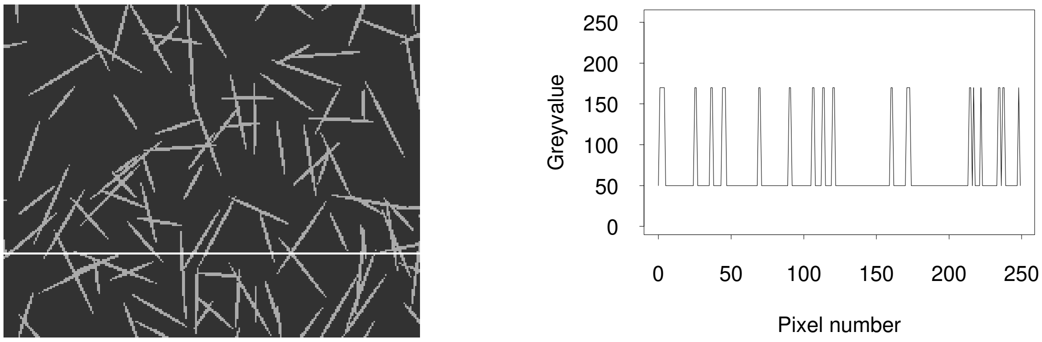


Figure 1.3: Illustration of thickness of fibres and digitization. **On the left**: A simulated realization of flat ellipses in the binary image with pixels $250 \times 200$ is intersected by a horizontal sampling line. **On the right:** A binary-valued function is generated by the intersections along the line (the images first published in Article A).

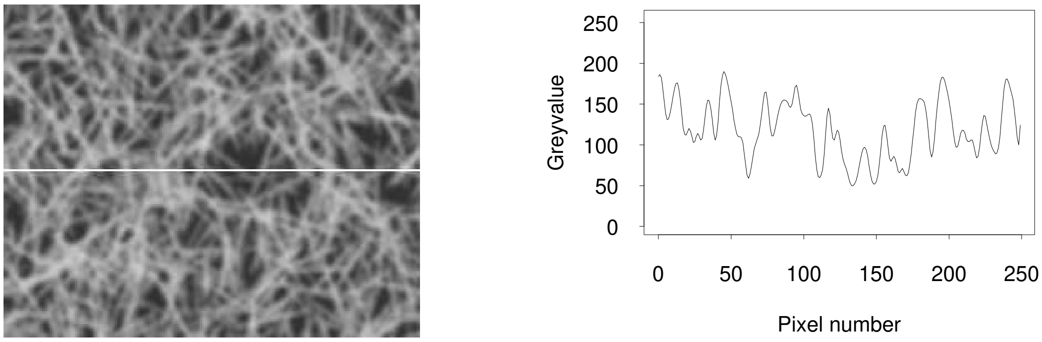


Figure 1.4: Illustration of grey levels and blurring. **On the left**: A simulated realization of flat ellipses in the blurred greyscale image with pixels $250 \times 200$ is intersected by a horizontal sampling line. **On the right**: The number of intersections is not observable from the grey-valued function induced on the line (the images first published in Article A).

ellipses (Figure 4.1). The objective is to estimate the orientation distribution of the central fibres from observations made on a digital image of thick fibres. In our approach, instead of observing the intersection between the central fibres and a sampling line, we observe a grey-valued function along the sampling line. The directional variation of grey levels is measured in terms of the scaled variogram which is similar to the variogram of order one by Matheron (1971, 1982). The scaled variogram is defined as the mean of absolute difference between grey values at two points scaled by the distance of the points (5.1). In practice, it is easy and fast to calculate. The basic idea of our methods is to use an approximative relation between the scaled variogram and the point intensity of central fibres. The method based on the proportional relationship is first suggested in Article A, cf. also Kärkkäinen (1999). Under the proportional model, the fibre orientation distribution is assumed to be in relationship to the ratio of two scaled variograms. This idea was based on the fact that the ratio is exact if the intersection points can be observed, cf. Forgacs and Strelis (1963). Empirical reasoning showed that the model performed quite well even for greyscale images generated by dead leaves models. Later, Jeulin (2000) considered the slope of the variogram of order one for these type of images. He found the approximatively proportional relation between the slope and the apparent intercepts along the sampling line. In Article C, this proportional link is used as an approximation between the scaled variogram and the point intensity of the intersections formed by the line and the boundary of the union of fibres. Furthermore, if fibres are convex and sufficiently thin, the scaled variogram is proportional to the point intensity of central fibres. For a binary image of Boolean fibres, an improvement of the relation, the so-called refined relation, is suggested in Article C, and the method based on that is employed in Article B. In both of the estimation methods, a parametric model for the orientation distribution is used in order to relate the ratios of the observed scaled variograms to the orientation parameters. Since the estimation equations are non-linear, the orientation parameters are obtained numerically from least-squares type procedures. As a parametric model, the elliptic density of two parameters, orientation angle and the strength of anisotropy, is used. Other orientation distribution can be applied as well.

In Article D, the performance of the new methods is compared with each other and with an industrial standard (Erkkilä et. al., 1998) by simulation experiments. As simulation models we apply Boolean and shot-noise models with varied length intensity, that is, the mean length of fibres per area unit and varied orientation angle and anisotropy of the elliptic model. As a result of the simulation study, the variogram-based methods seem to perform better in the anisotropy estimation in the binary case. The refined method performs in a larger range of the varied parameters than the proportional method, as was expected according to Article C. In the shot-noise case all methods perform quite well in the ranges of paper parameters. The industrial standard is, however, the most stable. Concerning a dead leaves model, we can say that the refined method is, experimentally, more appropriate in strong anisotropic cases.

The new methods have also been applied to the analysis of empirical data from paper technology research. The data contain greyscale images of the layered paper strips. In Article A, the standard errors of estimators have been calculated using the bootstrap, cf. Efron and Tibshirani (1993). In practice, it is worth evaluating the magnitude of the standard errors for each type of images but not for all analyses.

Our work combines stereology and image analysis, which traditionally have been quite separate and almost non-communicating approaches. From the point of view of stochastic geometry, stereology and spatial statistics, the new results consist of finding stereological connections between scaled variograms and the orientation density, solving the estimation equations based on that connection, and estimating the sampling uncertainty. This problem is solved in this thesis for various underlying fibre models. It turns out that the distributional theory of stochastic geometric as well as random field models is complex. This leads to the use of stochastic simulation both in controlling statistical uncertainty and in evaluating the methods.

The content of the introduction part of this thesis is the following: Chapter 2 reviews the fibrous structure in paper and outline the methods used in paper technology. Chapter 3 introduces stochastic geometric models for fibrous structures and the basic stereological formulas. In Chapter 4, the random fields of fibres (with non-zero area) are considered, whereas the new stereological connections for random fields are introduced in Chapter 5. Chapter 6 shows statistical orientation analysis for various fibre models, consideration of estimation of standard errors and evaluation of methods by simulation. Chapter 7 is the discussion. In Chapter 8 the summaries of the included original publications are presented.

# Chapter 2

# Fibre Systems and Paper Structure

Many innovations concerning statistics of fibre systems have been published in paper technology journals. For this reason, we give a brief overview upon those industrial results which are close to our objectives.

Paper is a mixture of pulp fibres, fines and filling materials. The interests of paper makers are in the production of paper quality with spatially homogeneous fibre mass distribution, which is related to the mass distribution and the geometrical structure of fibres. In a machine-made paper sheet, the fibrous structure is typically anisotropic, non-homogeneous and layered due to the paper making process. Anisotropy of fibres can be decreased by changes in the flow of suspension, such as turbulence, which may, however, increase flocculation of fibres. The fibrous structure usually varies considerably in the vertical direction since fibres being first filtered are in greater number in the machine direction than those in the middle of a paper sheet.

In this work, our focus is on the orientation analysis of fibres. Fibre orientation is an important factor to be controlled in the paper production line. Together with internal strength, it causes anisotropy to the strength and the elasticity of paper. This is important for an end-user of paper, because anisotropy may cause curling of paper in copying and printing. Therefore, in papers for copying machines and printers isotropic orientation is the objective. On the contrary, in printing press of newspapers, anisotropic fibre orientation in the production direction is preferable. The influence of fibre orientation on paper properties is considered in Loewen (1997).

Early 3D models for paper fibres are considered, for example, in Corte and Kallmes (1962). In a network of horizontal fibres, the centres of fibres are assumed to be randomly distributed throughout the volume of paper. The fibres lie in parallel planes in which they have a random orientation density.

In Kallmes and Corte (1960), the ideal planar random structure has been considered, whereas the planar system either with anisotropy or flocculation of fibres is examined in Corte and Kallmes (1962) and Dodson (1971). As an example, such a mechanical property of paper as strength depends on the number of fibre crossings per area unit, cf. Kallmes and Corte (1960). It is related to anisotropy and flocculation of fibres. A wide review of stochastic structure in paper is given by Deng and Dodson (1994).

In laboratory environments, fibre orientation distribution has conventionally been estimated indirectly by a ultrasound tester, cf. e.g. Hutten (1994). The tester measures the velocity of ultrasound through paper along lines at various polar angles with respect to the machine direction. The velocity is assumed to be related to fibre orientation density but it is also related to the internal strength of paper. This method is based on a set of one-dimensional observations. Heuristically, it can be regarded as a stereological method. The availability of an on-line sensor for fibre orientation has been promoted, cf. Chapman et al. (2001). It is based on the combination of the reflection of laser light, and the polarization between the laser source and detectors. The sensors measure the orientation characteristics on both sides of the sheet in such a way that the effect of the filling material is minimized. This is an advantage in comparison with the ultrasound tester.

Recently, the layered fibre orientation analysis has been based on the combination of sheet splitting and dyeing techniques in connection with image analysis. A mechanical tape stripping technique is used in Erkkilä et al. (1998). Paper is split until the squared mass of a layer is 5-15 g/m$^2$. The number of layers in a paper sheet will usually be ten or more. Each layer is scanned using an optical scanner, cf. Figure 1.1. The estimate of the orientation distribution is based on an approximation of the gradient direction in each pixel. Stripping and dyeing techniques are used in Thorpe (1999), whereas in Xu et al. (1999) a series of images is obtained by confocal laser scanning microscopy. In these two latter methods the fibre orientation analysis is based on the Hough transform.

Staining of paper fibres is an old method. In Danielsen and Steenberg (1947), stained fibres are approximated by line segments. Their angles to the machine direction are counted and used for an approximation of fibre orientation distribution. In Forgacs and Strelis (1963), stained fibres are intersected by two straight lines, in the machine and cross directions, and the numbers of intersections are counted. The colouring of the fibres during the sheet forming process makes it possible to extract stained fibres better from the background and from each other, cf. Figure 1.2. This design can, however, be applied in experimental situations only.

# Chapter 3

# Stochastic Fibre Systems

Any stochastic model for a planar fibrous system is always an approximation, but it is fundamental to choose a model which is able to capture the essential features from the data. Let us think at the moment that the fibres can be detected with a reasonable precision. This is the case in paper technology experimenting where a part of fibres are stained during the paper making process. Hence, the image of the paper sample consisting of stained fibres might be thresholded in such a way that the fibres can be distinguished from the background, cf. the right image in Figure 1.2.

## 3.1 Boolean Line Segment Model

In the simplest form, fibres can be approximated by a line segment or by a chain of consecutive line segments. This approach was originally used for paper fibres in the estimation of orientation distribution, to our knowledge, by Danielsen and Steenberg (1947).

For simplicity, let us consider the planar Boolean model of line segments, cf. Matheron (1972, 1975). Line segments $\Gamma_n$ have fixed length $l$, an orientation direction $\alpha_n$ with $x_1$-axis from the common segment orientation density $f_{\mathcal{A}}(\alpha)$ on $[0, \pi)$, and the centre point at the origin. Location points $x_n$ are from a stationary planar Poisson point process of intensity $\lambda$. Here, $\lambda$ stands for the mean number of points per area unit. Further, the line segments are independent of each other and independent of $\{x_n\}$. Then, the typical grain $\Gamma_0$ represents line segments. The Boolean line segment model $\Gamma$ is constructed as a union of transformed line segments

$$\Gamma = \cup_{n=1}^{\infty}(x_n + \Gamma_n) = (x_1 + \Gamma_1) \cup (x_2 + \Gamma_2) \cup \ldots \ . \tag{3.1}$$

Two main characteristics of $\Gamma$ are the length intensity, $L_A = \lambda l$, the total mean length of segment pieces in a unit area, and the segment orientation density $f_{\mathcal{A}}(\alpha)$. Note that if $f_{\mathcal{A}}(\alpha)$ is uniform, $\Gamma$ is isotropic. See a simulation of isotropic and anisotropic line segments in Figures 3.1 and 4.2, respectively.

Inference on the segment orientation density $f_{\mathcal{A}}(\alpha)$ can be based on stereological observation, cf. e.g. Weibel (1979). Here, a Boolean model $\Gamma$ is cutted by a line $L_\beta$ forming an angle $\beta \in [0, \pi)$ with respect to the $x_1$-axis. Intersection points generate a stationary Poisson point process on the line $L_\beta$, cf. Matheron (1975). The intensity $P_L(\beta)$ of the point process is defined as the mean number of points per length unit. It is related to the projection length of a typical segment, $l|\sin(\alpha-\beta)|$, through the well-known formula

$$\begin{aligned} P_L(\beta) &= \lambda \int_0^\pi l|\sin(\alpha-\beta)| f_{\mathcal{A}}(\alpha) d\alpha \qquad (3.2) \\ &= L_A \int_0^\pi |\sin(\alpha-\beta)| f_{\mathcal{A}}(\alpha) d\alpha. \end{aligned}$$

Note that the relation (3.2) is essentially a generalization of Buffon's needle problem. Buffon (1777) solved the problem for a segment with fixed length and uniform orientation density, $f_{\mathcal{A}}(\alpha) = 1/\pi$. In this case, $P_L(\beta) = 2L_A/\pi$.

In a simple extension of the Boolean model $\Gamma$, the length of a segment $\Gamma_n$, denoted by $l_n$, is distributed by the length density $f_{\mathcal{L}}(l)$ with mean $\bar{l}$, and the segment orientation $\alpha_n$ by $f_{\mathcal{A}}(\alpha)$ on $[0, \pi)$. The length $l_n$ is assumed to be independent of $\alpha_n$ and the centre location $x_n$. Then, the length intensity and the point intensity are, respectively, $L_A = \lambda \bar{l}$ and

$$\begin{aligned} P_L(\beta) &= \lambda \int_0^\infty \int_0^\pi l|\sin(\alpha-\beta)| f_{\mathcal{A},\mathcal{L}}(\alpha, l) d\alpha dl \\ &\overset{(*)}{=} \lambda \left( \int_0^\infty l f_{\mathcal{L}}(l) dl \right) \int_0^\pi |\sin(\alpha-\beta)| f_{\mathcal{A}}(\alpha) d\alpha \\ &= L_A \int_0^\pi |\sin(\alpha-\beta)| f_{\mathcal{A}}(\alpha) d\alpha, \qquad (3.3) \end{aligned}$$

cf. Mecke and Stoyan (1980) and also Hilliard (1962); Corte and Kallmes (1962); Forgacs and Strelis (1963). The note $(*)$ emphasizes that we have used the independence of $\alpha$ and $l$. The orientational characteristic $P_L(\cdot) : [0, \pi) \to [0, \infty)$ is the so-called rose of intersections, cf. Mecke and Stoyan (1980). The connection between the elliptic segment orientation density and the point intensities $P_L(\beta)$ is illustrated in the top left images in Figures 4.2 and 4.3, respectively. The point intensities are calculated in eight directions $\beta_i$, $i = 1, \ldots, 8$, according to Tables 6.1 and 6.2; in the directions $\beta_i + \pi$ periodicity is used.

In practice, however, the length $l$ and orientation $\alpha$ of a segment $\Gamma_0$ may be dependent. For example, long paper fibres are mainly in the machine direction, whereas short fibres tend to be more isotropically distributed. In such cases, the stereological formula (3.3) is not valid anymore. Instead, one can write

$$\begin{aligned}P_L(\beta) &= \lambda \int_0^\infty \int_0^\pi l |\sin(\alpha-\beta)| f_{\mathcal{A}}(\alpha) f_{\mathcal{L}|\mathcal{A}}(l|\alpha) d\alpha dl \\ &= \lambda \bar{l} \int_0^\pi |\sin(\alpha-\beta)| f_{\mathcal{A}}(\alpha) \frac{\int_0^\infty l f_{\mathcal{L}|\mathcal{A}}(l|\alpha) dl}{\bar{l}} d\alpha \\ &= L_A \int_0^\pi |\sin(\alpha-\beta)| f_{\mathcal{A}}(\alpha) E_{\mathcal{L}|\mathcal{A}}(l|\alpha)/\bar{l}\ d\alpha \\ &= L_A \int_0^\pi |\sin(\alpha-\beta)| f_{\mathcal{R}_0}(\alpha) d\alpha, \qquad (3.4)\end{aligned}$$

where

$$f_{\mathcal{R}_0}(\alpha) = f_{\mathcal{A}}(\alpha)\ E_{\mathcal{L}|\mathcal{A}}(l|\alpha)/\bar{l} \qquad (3.5)$$

is the length-weighted segment orientation density of $\Gamma_0$. In the independent case, the formula (3.4) equals to (3.3) since $E_{\mathcal{L}|\mathcal{A}}(l|\alpha)/\bar{l} = 1$. Observing the intersection points, one can estimate the length-weighted orientation density $f_{\mathcal{R}_0}(\alpha)$ but not the marginal orientation density $f_{\mathcal{A}}(\alpha)$. In our case, it is an advantage since the first one is more relevant for paper fibres. The properties of paper are more related to fibre length fractions in different directions than to angles. Note that if $f_{\mathcal{R}_0}(\alpha)$ is, in this case, uniform, $\Gamma$ is isotropic, and vice versa.

In a more general case, the paper fibres may depend on each other and on their positions. There may further exist an anisotropic arrangement, attraction or repulsion of positions. One generalization of the Boolean model of segments is presented by a germ-grain model, cf. Hanisch (1981). Formally, we can consider a stationary marked point process $\Psi = \{[x_n; \Gamma_n]\}$, cf. Stoyan et al. (1995). Then, the germ-grain model with the positions $x_n$ (germs) and the segments $\Gamma_n$ (grains) is defined as

$$\Gamma = \cup_{[x_n,\Gamma_n]\in\Psi}(x_n + \Gamma_n) = (x_1 + \Gamma_1) \cup (x_2 + \Gamma_2) \cup \ldots .$$

The Boolean model is a special case. It should be mentioned that rotations of germ-grain models change the marks, the segment orientations, in contrast to the assumption made in the general definition of marked point processes (Penttinen and Stoyan, 1989; Stoyan et al., 1995, p. 216). Another generalization is to replace line segments $\Gamma_n$ by fibres, that is, smooth, simple curves of finite length either in Boolean models or germ-grain models. The general fibre process is needed for modelling the fibre orientation density.

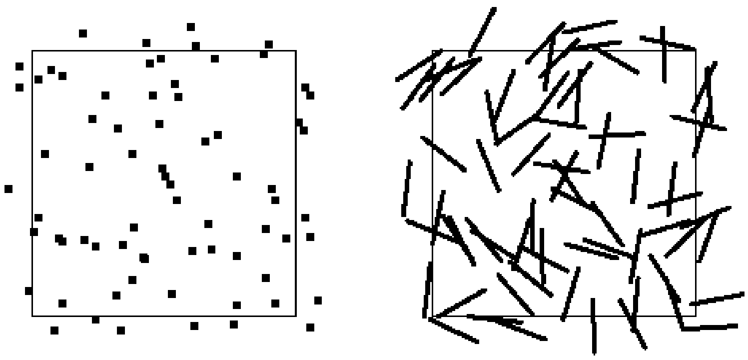

Figure 3.1: A simulation of a Boolean model of line segments. **On the left:** Centre points are from a Poisson stationary point process with intensity $\lambda = 50$. (The side length of the window is one). **On the right:** Segments with constant length $l = 0.2$ and uniform orientation are added on the points. Length intensity $L_A = \lambda l = 10$.

## 3.2 Fibre Process

The full generality of stochastic fibre systems is achieved in the construction of a planar fibre process introduced by Mecke and Stoyan (1980), see also Stoyan et al. (1995). The fibre process operates in terms of fibre length in planar sets. Properties such as length of a single fibre are unessential. This is not a restriction in our application since the fibre length distribution of paper can be measured economically from the fibre suspension at an early stage of manufacturing. The first role of the fibre process is to afford the orientation characteristic for modelling the fibre orientation distribution. The second role is to give a stereological formula for its estimation.

Let us assume that fibres are sufficiently smooth simple curves of finite length in the plane. Then, the fibre system forms a union of at most countably many fibres such that in any compact set, the number of fibres is finite and the fibres have only the end points in common. Then, the total fibre length in any compact set is finite. If $B$ is a (Borel) set of $\mathbb{R}^2$ and $\Phi(B)$ stands for the total random fibre length in $B$, then the fibre process $\Phi$ can be characterized by means of the joint distributions of $\Phi(B_1), \ldots, \Phi(B_n)$ for any collection of Borel sets $B_1, \ldots, B_n$ and for all non-negative integers $n$. Note that the line segment and more complex germ-grain models can be interpreted as a fibre process, if the length of $\Gamma \cap B$ is finite for all compact subsets $B$ of $\mathbb{R}^2$.

For a stationary fibre process $\Phi$, that is, $\Phi(B)$ has the same distribution as $\Phi(B+x)$ for all $x \in \mathbb{R}^2$, the orientation distribution is defined by a length-weighted distribution on $[0,\pi)$ called the rose of directions $\mathcal{R}$. Let $\alpha(x) \in [0,\pi)$ be the angle of the tangent of $\Phi$ at $x$. Then, if $\nu_1$ stands for the length measure, we have for an angle interval $(\alpha_1, \alpha_2)$, $0 \leq \alpha_i \leq \pi$,

$$E\nu_1\{x \in \Phi \cap [0,1]^2 : \alpha(x) \in (\alpha_1, \alpha_2)\} = \mathcal{R}((\alpha_1, \alpha_2))L_A,$$

where the length intensity $L_A = E\Phi([0,1]^2)$ is the mean total length of fibres per area unit. The left-hand side stands for the mean length of such fibre pieces that have tangent angles in $(\alpha_1, \alpha_2)$ per area unit. $\mathcal{R}$ can be considered as the distribution of the tangent angle at a typical point of the fibre process. The density for $\mathcal{R}$ is denoted by $f_{\mathcal{R}}(\alpha)$. Note that the segment orientation densities $f_{\mathcal{A}}(\alpha)$ and $f_{\mathcal{R}_0}(\alpha)$ of Boolean models are special cases of the density $f_{\mathcal{R}}(\alpha)$.

Another type of anisotropy of a fibre process can be introduced as well. The density of the rose of directions is, for example, uniform but the arrangement of locations of fibres is anisotropic. The different aspects of anisotropy have been considered by Stoyan and Beneš (1991). In this work, we consider the anisotropy in the rose of directions.

The second role of the fibre process is to give a stereological connection between the point intensities along lines and the fibre orientation distribution. If any stationary fibre process is sampled by a line $L_\beta$ at angle $\beta \in [0,\pi)$ to the $x_1$-axis, there exists an equation

$$P_L(\beta) = L_A \int_0^\pi |\sin(\alpha - \beta)| f_{\mathcal{R}}(\alpha) d\alpha, \tag{3.6}$$

when using $\int \mathcal{R}(d\alpha) = \int f_{\mathcal{R}}(\alpha) d\alpha$. The equation (3.6) is an integral part of many statistical methods for estimating the fibre orientation distribution. Among others, a second derivatives approach of (3.6) is considered in Hilliard (1962); Corte and Kallmes (1962); Mecke and Stoyan (1980); Mecke (1981); Serra (1982); Beneš (1989). A parametric model for the intersection counts, together with the second derivatives approach, is employed in Corte and Kallmes (1962); Beneš (1989). The fibre orientation density is expanded in a Fourier series in Hilliard (1962). The Steiner compact approach is used in Rataj and Saxl (1988, 1989), and Beneš and Gokhale (2000). Rataj and Saxl (1992) give a detailed summary and compare parametric estimators suggested by several authors, such as Marriott (1971), Kanatani (1984), Rataj and Saxl (1988, 1989). Parametric fibre orientation density combined with the ratios of counts is discussed in Forgacs and Strelis (1963). The ratio idea, applied in this thesis, is also considered in Corte and Kallmes (1962). Parametric methods based on intersection counts are constructed either by modelling the orientation density, or modelling the counts and deducing the model for the orientation density. Note that both design- and model-based approaches exist in these references.

The construction of a fibre process does not allow direct simulation. Hence, in applications where simulation is the means of controlling the uncertainty of the estimates, one typically utilizes such simple special cases as Boolean (or germ-grain) models of line segments or fibres.

## 3.3 Parametric Models for Fibre Orientation Distribution

Our main objective is to study the estimation of fibre orientation distribution defined by the rose of directions $\mathcal{R}$ of the fibre process, cf. Mecke and Stoyan (1980). It turns out that parametric modelling is relevant in extracting information on fibre orientation. A familiar parametric model in paper technology is an elliptic distribution. In this study, we use the elliptic density on $[0, \pi)$, cf. Article A,

$$f_{\mathcal{R}}(\alpha; \tau, \kappa) = \frac{c}{\sqrt{1 - \kappa^2 \cos^2(\alpha - \tau)}}, \tag{3.7}$$

where the orientation angle $\tau \in [0, \pi)$ is the preferred direction of the fibres, often the machine direction in paper technology, $\kappa^2 = 1 - b^2/a^2 \in (0, 1)$ with the lengths of the major and minor axes, $a$ and $b$, respectively, describes deviation from the circular model ($\kappa = 0$), and $c$ is the normalizing factor, cf. Figure 4.2. In paper technology terminology, the anisotropy parameter is $1 - e = 1 - b/a$, and, thus, $\kappa^2 = 1 - (1 - (1 - e))^2$. The elliptic density (3.7) can be deduced when an ellipse in Cartesian coordinates is presented in polar coordinates.

Alternative models for the paper orientation distribution can be applied as well. For example, in Forgacs and Strelis (1963), the elliptic density is

$$f(\alpha; a, b) = \frac{1}{\pi a b} \left( \frac{\cos^2 \alpha}{a^2} + \frac{\sin^2 \alpha}{b^2} \right)^{-1}. \tag{3.8}$$

Table 3.1 gives a summary of further alternatives, cf. Cresson (1988) and the references therein for the means of parameters. A typical statistical model is

Table 3.1: Examples of models for fibre orientation density.

| $f(\alpha)$ |
|---|
| $1 + a_1 \cos(2\alpha)$ |
| $1 + a_1 \cos(2\alpha) + a_2 \cos(4\alpha)$ |
| $1 + \sum_n a_n \cos(2n\alpha)$ |
| $1 + 2\sum_n \rho^{2n} \cos(2n\alpha)$ |
| $(1-\rho^2)/(1+\rho^2 - 2\rho\cos(2\alpha))$ |
| $\lambda/(\cos^2(\alpha) + \lambda^2 \sin^2(\alpha))$ |

the von Mises distribution with the density function

$$f(\alpha; \mu, K) = \frac{1}{2\pi I_0(K)} \exp(K\cos(\alpha - \mu)),\ 0 < \alpha \leq 2\pi,$$

where the mean direction is $\mu$, $0 \leq \mu < 2\pi$, the concentration parameter $K$, $K > 0$, and the modified Bessel function of the first kind and order zero

$$I_0(K) = \sum_{k=0}^{\infty} (1/k!)^2 (K/2)^{2k},$$

cf. e.g. Mardia (1972). Other parametric approaches are presented, for example, in Marriott (1971); Mardia (1972); Kanatani (1984); Rataj and Saxl (1988).

# Chapter 4

# Boolean Random Fields for Fibrous Structures

If real fibres have considerable physical thickness as paper fibres, the modelling with the (mathematical) fibres may not capture all the essential features. We assume from now on that the real fibres have a physical dimension and are further observed in the form of images. In this section, we introduce useful image models and outline stereological connections related to intersection points. A stereological approach concerning grey levels is, however, discussed in Section 5.

## 4.1 Binary Model

In this section, we introduce binary models for thick fibres. For simplicity, let us consider planar line segments $\Gamma_n$ having the centre at the origin, the joined density $f_{\mathcal{A},\mathcal{L}}(\alpha, l)$ of orientation $\alpha_n \in [0, \pi)$ and length $l_n \in (0, \infty)$. In this set-up, the line segments $\Gamma_n$ are called central fibres. The thick fibres $\Xi_n$ are assumed to be thickened versions of central fibres $\Gamma_n$ such that $\Gamma_n \subseteq \Xi_n$. A simple model for $\Xi_n$ is a circle-dilated segment

$$\Xi_n = \Gamma_n \oplus B(o, r),$$

where $B(o, r)$ is a circular disk with centre at the origin $o$ and radius $r$ determining the thickness. This model is considered by Molchanov et al. (1993); Molchanov and Stoyan (1994) and Article C. The second example of $\Xi_n$ is an elliptical segment, a long flattened ellipse with half $\Gamma_n$ as the major axis, cf. Figure 4.1. This model is employed in Article A, B, C and in Johansson (2002). In a third rectangle model for $\Xi_n$, $\Gamma_n$ can be thought of as a longer side of the rectangle with width $w_n$. The last model is applied for paper fibres in, for example, Dodson (1971); Molchanov et al. (1993); Molchanov and Stoyan (1994); Deng and Dodson (1994); Kellomäki et al. (2001); Johansson (2002) and Article D.

If the thick fibres $\Xi_n$ are independent and identically distributed and independent of stationary Poisson points $\{x_n\}$ in $\mathbb{R}^2$ with the intensity $\lambda$, then we obtain the Boolean model

$$\Xi = \cup_{n=1}^{\infty}(x_n + \Xi_n) = (x_1 + \Xi_1) \cup (x_2 + \Xi_2) \cup \ldots, \tag{4.1}$$

see the illustration with elliptical segments in Figure 4.2. The sets $\Xi_n$ have the same distribution as a certain random set $\Xi_0$ called a typical grain.

In the Boolean model $\Xi$, the boundaries of apparent grains $\Xi_n$ are not typically observable, but $\partial\Xi$, the boundary of $\Xi$, is. The intersection by a sampling $L_\beta$ forms a Poisson point process $\partial\Xi \cap L_\beta$ with the point intensity $P_L^1(\beta)$, cf. Matheron (1975). The central fibres $\Gamma$, whose fibre orientation density we are interested in, are not observable and neither is the point intensity $P_L(\beta)$ of central fibres. The point intensity $P_L^1(\beta)$ has, however, been shown to be related to the point intensity $P_L(\beta)$, cf. Molchanov et al. (1993); Molchanov and Stoyan (1994). The general form of the relation, according to Article C, is

$$P_L^1(\beta) = \xi \frac{Eh_\beta(\Xi_0)}{Eh_\beta(\Gamma_0)} P_L(\beta), \tag{4.2}$$

where

$$\xi = 2\exp(-\lambda E\nu_2(\Xi_0)), \tag{4.3}$$

and $E\nu_2(\Xi_0)$ stands for the mean area of the typical grain $\Xi_0$. In (4.2), $Eh_\beta(\Gamma_0)$ and $Eh_\beta(\Xi_0)$ are the mean length of the projection of $\Gamma_0$ and $\Xi_0$, respectively, in the direction perpendicular to $\beta$. For the Boolean model of line segments

$$Eh_\beta(\Gamma_0) = \bar{l} \int_0^\pi |\sin(\alpha - \beta) f_{\mathcal{R}_0}(\alpha) d\alpha,$$

cf. (3.5). The form of $Eh_\beta(\Xi_0)$ depends on the thickening model of a line segment. For example, $Eh_\beta(\Xi_0) = Eh_\beta(\Gamma_0) + 2r$ if $\Xi_0$ is a segment dilated by a circle with radius $r$. The circle-dilated and rectangular models have been considered in Molchanov et al. (1993); Molchanov and Stoyan (1994), and in Article C. In the last one, $Eh_\beta(\Xi_0)$ is calculated for an elliptical segment. In general, if the extent of the thick fibre $\Xi_n$ is large compared to its thickness, then the relation (4.2) can be approximated by

$$P_L^1(\beta) \approx \xi P_L(\beta) \tag{4.4}$$

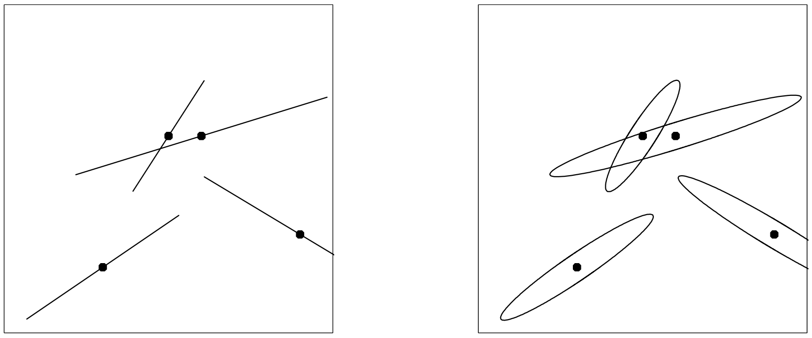

Figure 4.1: Illustration of thickening of line segments (first published in Article C). Line segments $(x_n + \Gamma_n)$'s are centred at points $x_n$'s on the left. Flat ellipses $(x_n + \Xi_n)$'s are centred at $x_n$'s on the right.

with $0 < \xi < 2$. The relation to the fibre orientation density $f_{\mathcal{R}}(\alpha)$ can be achieved through (3.6).

If the union of the thick fibres $\Xi$ has been observed in the form of a binary image, the image can be modelled by a binary random field $Z(x) = 1_{\Xi}(x),\ x \subseteq \mathbb{R}^2$. In the greyscale case instead, refinements of the random field are needed.

## 4.2 Dead Leaves Model

In this section, we consider a model for grey level variation motivated by two examples. In the greyscale images of thin paper layers, the uppermost fibres are assumed to reflect light in a different way from those at the bottom, cf. Figure 1.1. A further example is the greyscale image of stained fibres on the paper surface. We can assume that stained fibres are distributed uniformly through the vertical and horizontal direction (Fig. 1.2, left). In both cases, we can detect the uppermost fibres. This means that in the detailed modelling of the system of fibres one must take the order of construction into account. An obvious candidate is a dead leaves model.

Our description of the dead leaves model is based on random functions of grey levels on $\mathbb{R}^2$ with homogeneous time, cf. Jeulin (1989, 1993). This process class is based on a marked point process $\{[(x_n, t_n); \Theta_n]\}$, where $\{(x_n, t_n)\}$ is a stationary Poisson point process on $\mathbb{R}^2 \times [0, t]$ of intensity $\theta$, $\theta < \infty$, $t < \infty$, $\{\Theta_n\}$ are independent and identically distributed non-negative random functions on $\mathbb{R}^2$, independent of $\{(x_n, t_n)\}$, and have the centre point at the origin. An example of $\Theta_n$ is a random set $\Xi_n$ with a random grey level. It is supposed that the support of $\Theta_n$,

$$\Xi_n = \{x \in \mathbb{R}^2 : \Theta_n(x) > 0\},$$

is compact. We let $\Theta_0$ and $\Xi_0$ be a typical grey-valued grain and its support, respectively. $\Xi_0$ can be interpreted as a typical thick fibre. The dead leaves random function model

$$Z_t(x),\ x \in \mathbb{R}^2, \tag{4.5}$$

equals the grey value at $x$ of the most recently arrived grain. To be more precise, let

$$I(x) = \{n : \Theta_n(x - x_n) > 0\}$$

be the set of grains hitting $x$. If $I(x)$ is empty, let $Z_t(x) = 0$. Otherwise, $Z_t(x) = \Theta_I(x - x_I)$ where $I$ is the unique element in $I(x)$ such that $t_I > t_n$ for all $n \in I(x) \setminus \{I\}$. In a special case, the grey-valued random functions $\Theta_n$ depend on time. In all of these cases, at the time $t$, the union of supports $\Xi_n$ forms the Boolean model $\Xi$ with the intensity $\lambda = \theta t$. See the illustration in Figure 4.2.

In the dead leaves model with finite time, the boundary $\partial\Xi$ and, further, the intensity $P_L^1(\beta)$ of the intersection $L_\beta \cap \partial\Xi$ can be observed. In the dead leaves model of self-coloured fibres, we can also count the apparent intercepts formed by fibres identical with colours along the line. The question is how the number of intercepts is related to the point intensity $P_L(\beta)$ and to the fibre orientation density.

## 4.3 Shot-noise Model

Another non-binary situation appears, especially, when one is interested in the surface properties of a fibrous material. Such observations are due to the stylus instrument which measures the total local "height" of fibres at each point. The height cumulates when fibres intersect. The role of the grey level is to describe the height. The construction model should, therefore, have the additivity property in order to be a sufficient description for the surface. For situations such as these, shot-noise models have been recently suggested for paper fibres, cf. e.g. Johansson (2002). The images from paper layers have been modelled by shot-noise models in Kellomäki et al. (2001).

A simple planar shot-noise model can be defined as a random field

$$Z_s(x) = \sum_{\Xi_n \in \Xi} \delta 1_{\Xi_n}(x - x_n),\ x \in \mathbb{R}^2, \tag{4.6}$$

where $\Xi$ is the Boolean model of thick fibres $\Xi_n$ deposited on the plane. The germs $x_n$ are the points of a stationary Poisson point process in $\mathbb{R}^2$, $\Xi_n$ i.i.d. thick fibres, and independent of $\{x_n\}$. Further, $1_{\Xi_n}(x - x_n)$ indicates if $\Xi_n$ hits the point $x$, and $\delta$ stands for the height of a thick fibre. Hence, $Z_s(x)$ in (4.6) is proportional to the number of fibre crossings at $x$, see Figure 4.2. The general form of the shot-noise model is

$$Z_s(x) = \sum_{\Xi_n \in \Xi} h(x - x_n, \Xi_n), \;\; x \in \mathbb{R}^2,$$

where $h(x - x_n, \Xi_n)$ is a blur of $\Xi_n$ at $x$, cf. e.g. in Rice (1977). The moments of the process can be described directly using Campbells' theorem, see Schmidt (1985); Stoyan et al. (1995).

In the stereological observation, the number of the intersections $L_\beta \cap \partial\Xi$ can be observed and the point intensity $P_L^1(\beta)$ can be estimated as in the binary case if the image is not totally covered.

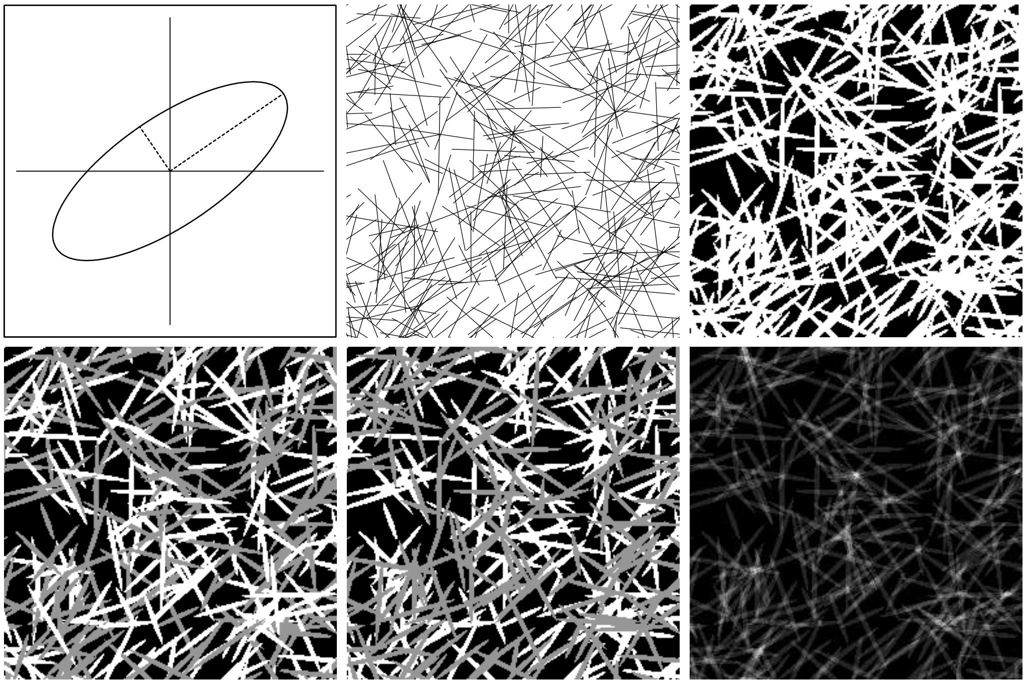

Figure 4.2: Illustration of orientation density and five fibre image models in the images of size $256 \times 256$. **From top left to right:** On the left, an elliptic orientation density $f_{\mathcal{A}}(\alpha; \tau, \kappa)$ with orientation angle $\tau = 0.60$ and the ratio $b/a = 0.28$ of the minor and the major axes of the ellipse, $b$ and $a$, respectively, cf. (3.7). Strength $\kappa = 0.92$. In the middle, a simulation of a Boolean model $\Gamma$ of line segments with mean length $\bar{l} = 55$ and the elliptic orientation density. Length intensity, the mean length per unit area, is $L_A = 0.24$. On the right, a Boolean model $\Xi$ of the same line segments thickened by ellipses with minor axis $b = 2$ in a digital two-valued image such that ellipses have grey value 255 and background 0. **From down left to right:** On the left, the same fibres coloured by a dead leaves model with homogeneous time such that fibres have either grey value 255 or 150 with half probability. In the middle, the grey value is time-dependent: the uppermost half of fibres has grey value 150 and other ones 255. On the right, the image of a shot-noise model where the grey level at each point is proportional to the number of crossed fibres.

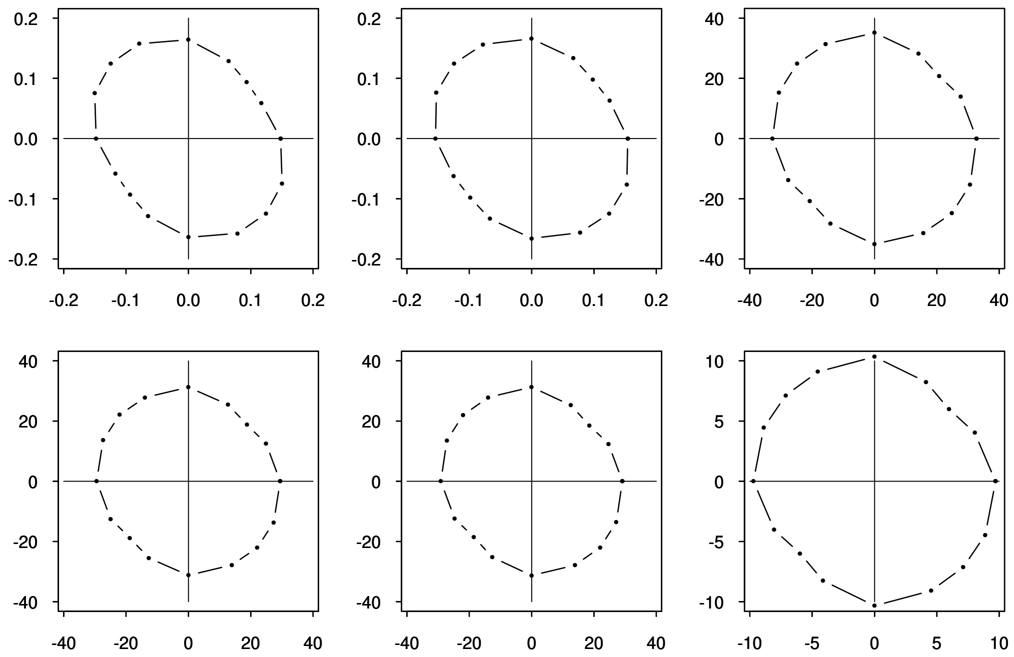


Figure 4.3: Illustration of roses of intersections and scaled variograms. Curves correspond the images in Figure 4.2. **From top left to right:** On the left, theoretical point intensities $P_L(\beta)$ calculated using (3.3) when the segment orientation density is $f_{\mathcal{A}}(\alpha;\tau,\kappa)$ with $(\tau,\kappa) = (0.60, 0.92)$, cf. (3.7), and $L_A = 0.244$. In the middle, the intensities $\hat{P}_L(\beta)$ estimated from the image with line segments. On the left, the scaled variograms $\hat{V}_L(\beta)$ in (6.8) estimated from the two-valued image. **From down left to right:** On the left and in the middle the scaled variograms estimated from the dead leaves images, and on the right from the shot-noise image.

# Chapter 5

# Stereological Connections in Random Fields

The aim has been to develop an automatic and fast system for calculating the number of points along the scan lines in both binary and greyscale images. Further, the intersection points as such are not necessarily observable due to thickness, overlapping, grey levels of fibres and digitization, cf. Figure 1.3 and 1.4. One solution is to observe the directional variation of grey levels in the form of a scaled variogram along the sampling line. The main objective is to investigate the relation between the scaled variogram and the fibre orientation distribution $f_{\mathcal{R}}(\alpha)$. If the grid of the image is small enough, it is sufficient to approximate the relation between the scaled variogram and the unobservable point intensity $P_L(\beta)$. Otherwise, the exact relation is needed.

## 5.1 Binary Images

Let us begin by thinking of a binary image (without digitization at the moment) as a random field $Z(x) = 1_{\Xi}(x),\ x \in \mathbb{R}^2$, where $\Xi$ is a Boolean model. Intersecting the image by a sampling line $L_\beta$, we observe a binary-valued function $Z(x),\ x \in L_\beta \cap \mathbb{R}^2$. We can consider two grey values $Z(x)$ and $Z(y)$ observed at points $x, y \in \mathbb{R}^2$, respectively, at the distance $d = ||x - y||$. The scaled variogram is defined as

$$V_L(d,\beta) = \frac{E|Z(x) - Z(y)|}{d}. \tag{5.1}$$

Due to the stationarity of $\Xi$, the locations of $x, y \in \mathbb{R}^2$ with $d = ||x - y||$ have no influence. Note that $V_L(d,\beta)$ is equal to $2\gamma_1(d,\beta)/d$, where $\gamma_1(d,\beta)$ is the variogram of order one, cf. Matheron (1971, 1982).

For Boolean fibres $\Xi$, the scaled variogram is

$$V_L(d,\beta) = \frac{2[\mathbf{P}(\Xi \cap \{y\} = \emptyset) - \mathbf{P}(\Xi \cap \{x,y\} = \emptyset)]}{d}, \tag{5.2}$$

where

$$\begin{aligned}
&\mathbf{P}(\Xi \cap \{y\} = \emptyset) = \exp(-\lambda E\nu_2(\Xi_0), \\
&\mathbf{P}(\Xi \cap \{x,y\} = \emptyset) = \exp(-\lambda E\nu_2(\Xi_0 \cup [\Xi_0 + \{y - x\}])).
\end{aligned} \tag{5.3}$$

In a case of small $d$, for convex fibres in (5.3)

$$\nu_2(\Xi_0 \cup [\Xi_0 + \{y - x\}]) \approx \nu_2(\Xi_0) + dh_\beta(\Xi_0), \tag{5.4}$$

cf. e.g. Matheron (1975, p. 141). Using further Taylor's approximation, we obtain

$$V_L(d,\beta) \approx P_L^1(\beta) \tag{5.5}$$

according to Article C, cf. also Jeulin (2000). Recall that $P_L^1(\beta)$ is the intensity of the point process $L_\beta \cap \partial\Xi$ with the boundary $\partial\Xi$. Combining (5.5) and (4.4), the proportional model

$$V_L(d,\beta) \approx \xi P_L(\beta) \propto P_L(\beta) \tag{5.6}$$

is obtained. The relation (5.6) depends only on $\xi$, see (4.3). The connection to the fibre orientation density of $\mathcal{R}$ can be found through the formula (3.6).

In Article C, a refined approximation for $V_L(d,\beta)$,

$$V_L(d,\beta) \approx \left(1 - \frac{\lambda dEh_\beta(\Xi_0)}{2}\right) P_L^1(\beta), \tag{5.7}$$

is suggested where $Eh_\beta(\Xi_0)$ is the mean projection length of thick fibres in the perpendicular direction to $\beta$. If we can assume that the length of fibres is large compared to their thickness, then we obtain the refined proportional model as an approximation of (5.7)

$$\begin{aligned}
V_L(d,\beta) &\approx \xi \left(1 - \frac{dP_L(\beta)}{2}\right) P_L(\beta) \\
&\propto \left(1 - \frac{dP_L(\beta)}{2}\right) P_L(\beta),
\end{aligned} \tag{5.8}$$

which is one of the main contributions of this thesis.

**Remark** If the distance $d$ is not small enough, the approximation (5.4) is not valid. In that case, under the Boolean model, we can obtain from (5.3)

$$\begin{aligned}
&\mathbf{P}(\Xi \cap \{x,y\} = \emptyset) \\
&= \exp(-\lambda(2E\nu_2(\Xi_0) - E\nu_2(\Xi_0 \cap [\Xi_0 + \{y - x\}]))).
\end{aligned} \tag{5.9}$$

Let us fix $\Xi_0$ to be a rectangle with fixed length $l$, width $w$ and random orientation from a parametric model $f_{\mathcal{R}}(\alpha;\theta)$. Then, for $x-y=d(\cos\beta,\sin\beta)$ the last term of (5.9) can be written as

$$
\begin{aligned}
&E\nu_2(\Xi_0\cap[\Xi_0+\{y-x\}]) = E\nu_2(\Xi_0\cap[\Xi_0-d(\cos\beta,\sin\beta)])\\
&=\int_0^\pi \nu_2(\Xi_0\cap[\Xi_0-d(\cos\beta,\sin\beta)]|\alpha)f_{\mathcal{R}}(\alpha;\theta)d\alpha\\
&=\int_0^\pi (w-d|\sin(\alpha-\beta)|)_+(l-d|\cos(\alpha-\beta)|)_+f_{\mathcal{R}}(\alpha;\theta)d\alpha\\
&=f_{d,\beta}(\theta,l,w),
\end{aligned}
$$

where $(w-d|\sin(\alpha-\beta)|)_+$ and $(l-d|\cos(\alpha-\beta)|)_+$ are the sides of the rectangle in the case of an intersection, cf. also Dodson (1971); Johansson (2002). Furthermore,

$$\mathbf{P}(\Xi\cap\{y\}=\emptyset)=\exp(-\lambda wl).$$

From (5.2), the exact relation

$$
\begin{aligned}
V_L(d,\beta) &= 2\exp(-\lambda wl)[1-\exp(-\lambda(wl-f_{d,\beta}(w,l,\theta)))]/d\\
&= f_{d,\beta}(\theta,l,w,\lambda) \qquad (5.10)
\end{aligned}
$$

is obtained, where the direction $\beta$ and the distance $d$ are fixed and the parameters $\theta,l,w$ and $\lambda$ unknown.

## 5.2 Images of Dead Leaves Models

In the case of grey-valued fibres, we assume that the image is partially covered by the dead leaves random function model $Z_t(x)$, cf. (4.5). The primary compact grain, the support $\Xi_0$, is grey-valued by the primary random function $\Theta_0$ with slow variation of grey levels (or constant grey value). Then, for small $d$

$$V_L(d,\beta)=\frac{E|Z_t(x)-Z_t(y)|}{d}\approx\xi' P_L^1(\beta), \qquad (5.11)$$

where $\xi'$ depends on grey levels; in a binary $\{0,1\}$-valued image, $\xi'=1$, cf. Jeulin (2000) and Article C. Recall that at the time $t$ the union of the supports $\{\Xi_n\}$ forms the Boolean model $\Xi$ with the intensity $\lambda=\theta t$, cf. (4.1), and $P_L^1(\beta)$ is the intensity of $L_\beta\cap\partial\Xi$. Assuming that the length of the convex fibres is large compared to thickness, we are back in the proportional model

$$V_L(d,\beta)\approx\xi'\xi P_L(\beta)\propto P_L(\beta), \qquad (5.12)$$

cf. (4.4).

**Remark.** If the image is completely covered ($t$ is infinite), the variation on grey level arises only from the intercepts of the apparent sets having a non-identical grey level. However, the proportionality (5.11) holds but the constant $\xi'$ is not anymore the same, cf. Jeulin (2000).

# Chapter 6

# Statistical Orientation Analysis Methods

Several authors have considered the image-based orientation analysis. Using a stereological approach, Molchanov and Stoyan (1994) consider the estimation of the rose of intersections by empirical capacity functionals for binary images of Boolean thick fibres. In a further approach, the orientation distribution is obtained from a greyscale image using sequential thresholding of grey levels and counting the intercepts along sampling lines in Serra (1982, p. 467). In Erkkilä et al. (1998), the fibre orientation analysis is based on the gradient measurements, whereas Thorpe (1999) and Xu et al. (1999) use the Hough transform. In Redon et al. (1998), the Fourier image transform is suggested for orientation analysis, concluding that the Fourier approach is less sensitive for digitization than the methods originating from the rose of directions. In Johansson (2002), the method for the parameters of a fibre system including orientation is based on a covariance function for shot-noise models.

In this work, we are looking for alternative fast stereological-based solutions, where the extraction of fibres will be bypassed. The methods for estimating the fibre orientation density are based on the ratios of the stereological formulas (3.6) in several directions. The methods are shown for binary as well as greyscale images of a dead leaves model.

## 6.1 From Point Intensities to Orientation Parameters

This work can be considered to be an extension of the idea of Forgacs and Strelis (1963) developed in paper technology. In their approach, the observed fibres constitute a realization generated by the Boolean model $\Gamma$ of line segments with independent random length and random orientation, cf. (3.1). The segments are intersected by two lines $L_{\beta_i}$ with the same length in vertical ($\beta_1 = 0$) and horizontal ($\beta_2 = \pi/2$) directions. The intensity of intersection points $P_L(\beta_i)$ along each line is related to the fibre orientation density $f_{\mathcal{A}}(\alpha)$ through (3.3). Assuming a parametric model for $f_{\mathcal{A}}(\alpha;\theta_1)$, the ratio

$$\frac{P_L(\beta_2)}{P_L(\beta_1)} = \frac{\int_0^\pi |\sin(\alpha-\beta_2)| f_{\mathcal{A}}(\alpha;\theta_1) d\alpha}{\int_0^\pi |\sin(\alpha-\beta_1)| f_{\mathcal{A}}(\alpha;\theta_1) d\alpha} = f(\theta_1) \tag{6.1}$$

will be obtained, where the orientation parameter $\theta_1$ is the only unknown. Note that the intensity $L_A$ vanishes. Considering a set $W \subset \mathbb{R}^2$ with fibres, an unbiased estimator for the point intensity is

$$\hat{P}_L(\beta_i) = \frac{\#\{L_{\beta_i} \cap \Gamma \cap W\}}{\nu_1(L_{\beta_i} \cap W)}$$

for each direction $\beta_i$, cf. e.g. Stoyan et al. (1995). In Forgacs and Strelis (1963), they assume the elliptic density (3.8), where $a$ and $b$ are, respectively, lengths of major and minor semiaxes of the ellipse, and the orientation angle is fixed to the machine direction. From $f(\theta_1)$, cf. (6.1), the estimate for $\theta_1 = a/b$ can be solved numerically when $\hat{P}_L(\beta_2)/\hat{P}_L(\beta_1)$ is observed.

This approach can be generalized. We consider the Boolean line segments whose orientation density is a parametric rose of directions $f_{\mathcal{R}}(\alpha;\theta)$ with several orientation parameters $\theta = (\theta_1, \cdots, \theta_t)$, cf. also (3.5). Therefore, we may need more than two sampling directions $\beta_i$ with $i = 1, \ldots, p$ such that $p \geq (t+1)$. Using the abbreviation

$$F_i(\theta) = \int_0^\pi |\sin(\alpha-\beta_i)| f_{\mathcal{R}}(\alpha;\theta) d\alpha, \tag{6.2}$$

we obtain from (3.6) the estimation equation

$$\frac{P_L(\beta_i)}{P_L(\beta_1)} = \frac{F_i(\theta)}{F_1(\theta)}$$

for each direction $\beta_i$. The least-squares estimation procedure leads to the (numerical) minimization of the weighted square form

$$\hat{\chi}(\theta) = \sum_{i=2}^{p} w_i \left( \hat{P}_L(\beta_i) - \frac{F_i(\theta)}{F_1(\theta)} \hat{P}_L(\beta_1) \right)^2 \tag{6.3}$$

with respect to $\theta$, cf. Article A. The optimal choice for the points from the Poisson point process gives $w_i = L_i L_1/(L_i + L_1)$ with $L_i$, which is the total length of the transect in direction $\beta_i$. Note that in cases where $L_A$ is in focus, a simple estimate is $\hat{L}_A = F_i(\hat{\theta})/\hat{P}_L(\beta_1)$, where the estimates $\hat{\theta}$ and $\hat{P}_L(\beta_1)$ are used. Another possibility is to consider the system of equations which leads to the minimization of the weighted square form

$$\hat{\chi}(\theta, L_A) = \sum_{i=1}^{p} w_i \left(\hat{P}_L(\beta_i) - L_A F_i(\theta)\right)^2 \tag{6.4}$$

with respect to $\theta$ and $L_A$ at the same time. Alternative methods for the estimation of $L_A$ are shown in Hilliard (1962); Forgacs and Strelis (1963); Marriott (1971); Mecke and Stoyan (1980); Mecke (1981); Stoyan et al. (1995).

If fibres have some thickness and are observed in a binary or greyscale image, we can detect the boundary of the union of thick fibres. Let us model these fibres by convex Boolean fibres $\Xi$, cf. (4.1). If the length of fibres is large compared with thickness, the point intensity $P_L^1(\beta)$ of the intersection $L_\beta \cap \partial\Xi$ is approximately proportional to $P_L(\beta)$ of line segments $\Gamma$, cf. (4.4). The rose of directions $\mathcal{R}$ of $\Gamma$ can be estimated, using the ratio of point intensities of $L_{\beta_i} \cap \partial\Xi$, from

$$\frac{P_L^1(\beta_i)}{P_L^1(\beta_1)} \approx \frac{F_i(\theta)}{F_1(\theta)} \tag{6.5}$$

for each direction $\beta_i$. Here, the product $\xi L_A$, where $L_A$ is the intensity of $\Gamma$ and $\xi$ constant (4.3), is eliminated. Observing

$$\hat{P}_L^1(\beta_i) = \frac{\#\{L_{\beta_i} \cap \partial\Xi \cap W\}}{\nu_1(L_{\beta_i} \cap W)}$$

in the direction $\beta_i$, one minimizes the least-squares estimation procedure, analogous to (6.3),

$$\hat{\chi}(\theta) = \sum_{i=2}^{p} w_i \left(\hat{P}_L^1(\beta_i) - \frac{F_i(\theta)}{F_1(\theta)}\hat{P}_L^1(\beta_1)\right)^2 \tag{6.6}$$

with respect to $\theta$. Note that if the length of fibres is not large enough compared with thickness, (6.6) gives the estimates for the rose of directions of the boundary $\partial\Xi$. Then, in (6.5) the intensity $L_A$ of $\partial\Xi$ is eliminated. Further, (6.4) can be used for the estimation of the boundary intensity.

Since the methods are based on optimization, it should be mentioned that optimization works better with few parameters. Hence, parsimonious modelling of orientation is preferable. The methods find the minimum better in stronger anisotropy due to flatness and identification problems near isotropic cases.

## 6.2 From Scaled Variograms to Orientation Parameters

It has been worth developing a faster stereological method for the estimation of fibre orientation from the digital images. We utilize the directional variation of grey levels along sampling lines in the form of scaled variograms, and eliminate the nuisance parameters using the ratios. The proportional method suggested in Article A is addressed to a binary or a greyscale image generated by a dead leaves model. The refined method shown in Article B and the exact method are directed to a binary image.

### 6.2.1 Proportional Method

Let us model the fibre orientation density by a parametric density $f_{\mathcal{R}}(\alpha;\theta)$ with parameters $\theta$. It can be elliptic or any of the models shown in Table 3.1. Let us consider a greyscale image generated by a dead leaves model. Using the proportional model (5.12), the relation (3.6) and (6.2), we obtain

$$\frac{V_L(d,\beta_i)}{V_L(d,\beta_1)} \approx \frac{F_i(\theta)}{F_1(\theta)} \tag{6.7}$$

for any direction $\beta_i$. The product $\xi'\xi L_A$ is eliminated since it is assumed to be the same constant in each direction. In a binary image, the proportional model (5.6) also gives the equation (6.7). In that case, $\xi L_A$ is eliminated.

In practice, we observe a digital image. It can be regarded as a discrete random field $\{Z(l,m):\ l,m \in W \cap \mathbb{Z}_+\}$ where $Z(l,m)$ denotes a random grey value at a centre point $(l,m)$ of a pixel in $W$. Due to digitization, we use a digital line $\bar{L}_{\beta_i}$ as a sampling line. Four natural directions for digital lines are horizontal, vertical and the two diagonal directions, see Table 6.1. Further, we have selected to use four intermediate directions shown in Table 6.2. In the intermediate directions $\beta_2 = 0.464,\ \beta_4 = 1.107,\ \beta_6 = 2.034,\ \beta_8 = 2.678$, every second pair of adjacent pixels is located in the main direction $\beta_{i-1}$ at the distance $d_{i-1}$, whereas every second pair is situated in the main direction $\beta_{i+1}$ at the distance $d_{i+1}$, cf. Table 6.2. For the direction $\beta_8 = 2.678$, however, the latter direction is $\beta_9 = \beta_0$ and the distance is $d_9 = d_0$. In each direction $\beta_i$, we observe a discrete grey-valued function $\{Z(l,m):\ (l,m) \in \bar{L}_{\beta_i} \cap W\}$. An estimator for the scaled variogram along $\bar{L}_{\beta_i}$ is

$$\hat{V}_L(\beta_i) = \frac{\sum_{(l,m),(l',m') \in \bar{L}_{\beta_i}} |Z(l,m) - Z(l',m')|}{L_i}, \tag{6.8}$$

Table 6.1: Main directions $\beta_{i=\{1,3,5,7\}}$ and corresponding distances $d_{i=\{1,3,5,7\}}$ between pixel centres along digital lines.

| | |
|---|---|
| $\beta_1 = 0$ | $d_1 = 1$ |
| $\beta_3 = \pi/4 = 0.785$ | $d_3 = \sqrt{2}$ |
| $\beta_5 = \pi/2 = 1.571$ | $d_5 = 1$ |
| $\beta_7 = 3\pi/4 = 2.356$ | $d_7 = \sqrt{2}$ |

Table 6.2: Intermediate directions $\beta_{i=\{2,4,6,8\}}$ of digital lines are discretized to the pairs of pixel centres in the adjoining main directions.

| | $\beta_{i-1}$ | $d_{i-1}$ | $\beta_{i+1}$ | $d_{i+1}$ |
|---|---|---|---|---|
| $\beta_2 = 0.464$ | 0 | 1 | $\frac{\pi}{4}$ | $\sqrt{2}$ |
| $\beta_4 = 1.107$ | $\frac{\pi}{4}$ | $\sqrt{2}$ | $\frac{\pi}{2}$ | 1 |
| $\beta_6 = 2.034$ | $\frac{\pi}{2}$ | 1 | $\frac{3\pi}{4}$ | $\sqrt{2}$ |
| $\beta_8 = 2.678$ | $\frac{3\pi}{4}$ | $\sqrt{2}$ | 0 | 1 |

where $(l, m)$ and $(l', m')$ are adjacent pixel centres and

$$L_i = \sum_{(l,m),(l',m') \in \bar{L}_{\beta_i}} ||(l, m) - (l', m')|| \tag{6.9}$$

is the total length of the distances between the observed centres. In each sampling direction, the whole image is scanned using a bundle of parallel digital lines. Edge correction is not employed. For the four image models in Figure 4.2, the estimate of (6.8) is calculated in eight directions $\beta_i$, $i = 1, \ldots, 8$,. Due to periodicity, they also correspond the estimates in $\beta_i + \pi$ directions, cf. Figure 4.3.

The estimator (6.8) is an unbiased estimator for the scaled variogram in the main directions since $d$ is a constant. On the contrary, the mean of the estimator in the intermediate directions is a weighted average of scaled variograms in the adjacent main directions, cf. Article C, Section 8. The corrections in the intermediate directions are needed. Then, the ratio of the means is

$$\frac{E\hat{V}_L(\beta_i)}{E\hat{V}_L(\beta_1)} \approx \frac{G_i(\theta)}{G_1(\theta)}, \tag{6.10}$$

where

$$G_i(\theta) = \begin{cases} F_i(\theta), & \text{i=1,3,5,7} \\ \frac{d_{i-1}}{d_{i-1}+d_{i+1}} F_{i-1}(\theta) + \frac{d_{i+1}}{d_{i-1}+d_{i+1}} F_{i+1}(\theta), & \text{i=2,4,6,8,} \end{cases}$$

cf. Article C. Hence, (6.10) leads to the weighted sum of squares

$$\hat{\chi}(\theta) = \sum_{i=2}^{8} w_i \left( \hat{V}_L(\beta_i) - \frac{G_i(\theta)}{G_1(\theta)} \hat{V}_L(\beta_1) \right)^2, \tag{6.11}$$

with the weights $w_i = L_i L_1/(L_i + L_1)$ for $i = 2, \ldots, 8$, cf. (6.9), cf. Article A. This is minimized with respect to the orientation parameters $\theta$.

This method is a further extension of the idea shown in Forgacs and Strelis (1963). Using the ratios, a simple system of estimation equations is obtained. This method is convenient for a greyscale or binary image generated by a dead leaves model.

### 6.2.2 Refined Method

We employ the refined connection (5.8) shown for a continuous binary image of Boolean convex, long and narrow fibres. In that case, the ratio of the scaled variograms has a presentation

$$\frac{V_L(d,\beta_i)}{V_L(d,\beta_1)} \approx \frac{[1 - dL_A F_i(\theta)/2] F_i(\theta)}{[1 - dL_A F_1(\theta)/2] F_1(\theta)},$$

cf. (3.6) and (6.2). The product $\xi L_A$ is eliminated. Due to digitization, the means of the estimators (6.8) give the ratio

$$\frac{E\hat{V}_L(\beta_i)}{E\hat{V}_L(\beta_1)} \approx \frac{G_i(\theta, L_A)}{G_1(\theta, L_A)},$$

where

$$G_i(\theta, L_A) = [1 - d_i L_A F_i(\theta)/2] F_i(\theta)$$

holds for the main directions $i = 1, 3, 5, 7$, while

$$G_i(\theta, L_A) = \frac{d_{i-1}}{d_{i-1} + d_{i+1}} G_{i-1}(\theta, L_A) + \frac{d_{i+1}}{d_{i-1} + d_{i+1}} G_{i+1}(\theta, L_A),$$

holds for the intermediate directions with $i = 2, 4, 6, 8$, cf. Article C. This leads to the numerical minimization of the sum of weighted squares

$$\hat{\chi}(\theta, L_A) = \sum_{i=2}^{8} w_i \left( \hat{V}_L(\beta_i) - \frac{G_i(\theta, L_A)}{G_1(\theta, L_A)} \hat{V}_L(\beta_1) \right)^2,$$

where $w_i = L_i L_1/(L_i + L_1)$ for $i = 2, \ldots, 8$. $L_i$ (6.9) is the total length of the digital line in the direction $\beta_i$, cf. Article B. Both $\theta$ and $L_A$ are estimated in contrast to the proportional method.

In this method, proportionality has also been used. The estimation equations are, however, more complex. Advantages of the method arise in cases where the factor $dP_L(\beta)$, cf. (5.8), depends remarkably on $\beta$ and $d$. A case like this is a binary image with strong anisotropy, high intensity $L_A$ and low resolution in a grid. The refined method has been proved for a binary image of Boolean fibres.

### 6.2.3 Exact Method

The following approach is founded on the exact relation (5.10) between the scaled variogram $V_L(d, \beta)$ and the parametric orientation density $f_{\mathcal{R}}(\alpha; \theta)$ with $\theta = (\theta_1, \cdots, \theta_t)$. Since we use a digital image $\{Z(l,m) : \; l, m \in W \cap \mathbb{Z}_+\}$, we have to fix pairs $(d, \beta)$ according to the grid such that the adjacent pixel centres are along the continuous line $L_\beta$. Therefore, the earlier estimator of the scaled variogram is not valid here.

We construct the estimator for the scaled variogram based on the constant distance as follows. For fixed $\beta_i$, the smallest possible distance between pixel centres $(l, m)$ and $(l_1, m_1)$ along the line $L_{\beta_i}$ is denoted by $d_{i_1}$. Multiples of $d_{i_1}$ can also be used denoted by $d_{i_\eta}$, where $\eta \in \{2, \ldots, n_i\}$, and $n_i$ depends on the size of the grid. At the distance $d_{i_\eta}$, we obtain a discrete binary-valued function $\{Z(l,m) : (l,m), (l_\eta, m_\eta) \in L_{\beta_i} \cap W\}$. The estimator for the scaled variogram is

$$\hat{V}_L(d_{i_\eta}, \beta_i) = \frac{\sum_{(l,m),(l_\eta,m_\eta) \in L_{\beta_i}} |Z(l,m) - Z(l_\eta, m_\eta)|}{\sum_{(l,m),(l_\eta,m_\eta) \in L_{\beta_i}} (l,m) - (l_\eta, m_\eta)},$$

where the sum goes over the "adjacent" pixel centres at the distance $d_{i_\eta} = (l,m) - (l_\eta, m_\eta)$. In the horizontal and vertical directions, $d_{i_1} = 1$, and in the diagonal directions, $d_{i_1} = \sqrt{2}$. In the intermediate directions, 0.464, 1.107, 2.034, 2.678, for example, the smallest distance is $d_{i_1} = \sqrt{5}$. Any other intermediate directions can be utilized as well.

The estimation of the orientation parameters $\theta$ can be performed by a least-squares type procedure, as in Article A and in Johansson (2002). We obtain the system of estimation equations from (5.10), which leads to the minimization of the weighted square form

$$\hat{\chi}(\theta, \lambda, l, w) = \sum_{i=1}^{p} \sum_{\eta=1}^{n_i} (\hat{V}_L(d_{i_\eta}, \beta_i) - f_{d_{i_\eta}, \beta_i}(\theta, \lambda, l, w))^2$$

with respect to the parameters $\theta, \lambda, l, w$. Note that the number of estimated $V_L(d_{i_\eta}, \beta_i)$ should be at least $t+3$.

An advantage of the exact model is the fact that it gives more possibilities in choosing the sampling directions and the distances between pixels in each direction. A disadvantage, however, is the larger number of parameters to be estimated at once. We cannot eliminate the parameters $\lambda, l$ and $w$ in a case of no interest. This method is not implemented in this thesis.

## 6.3 Estimation of Standard Errors

We have shown the parametric methods for estimating the parameters $\theta$ of the fibre orientation density $f_{\mathcal{R}}(\alpha; \theta)$. The estimates for the parameters are obtained as results of minimization procedures. The variances of the estimators can be achieved by bootstrap or parametric bootstrap methods, cf. Efron and Tibshirani (1993).

In Article A, the orientation angle $\tau$ and the strength of anisotropy $\kappa$ of the elliptic model (3.7) are estimated by the proportional method. Furthermore, the standard errors of the estimators $\hat{\tau}$ and $\hat{\kappa}$ in a paper image are studied using the bootstrap. Instead of scanning the entire image, the digital lines in the direction $\beta_i$ are sampled sparsely using systematic sampling in order to obtain almost independent grey-valued functions for the bootstrap. Let us denote the lines along which the functions are observed by $\{L_{\beta_{i_1}}, L_{\beta_{i_2}}, \ldots, L_{\beta_{i_n}}\}$. The estimate $\hat{V}_L(\beta_i)$ is calculated according to (6.8) along the lines. This is performed in each direction $\beta_i,\ i = 1, \ldots, p$. We obtain the estimates $\{\hat{V}_L(\beta_1), \hat{V}_L(\beta_2), \ldots, \hat{V}_L(\beta_p)\}$, which are used in the minimization procedure (6.11) in order to achieve the estimates $\hat{\tau}$ and $\hat{\kappa}$.

In the bootstrap, we have a resampled set of lines $\{L^*_{\beta_{i_1}}, L^*_{\beta_{i_2}}, \ldots, L^*_{\beta_{i_n}}\}$ in the direction $\beta_i$ drawn from the original set of lines with replacement. Sampling lines in each direction, the estimates $\{\hat{V}^*_L(\beta_1), \hat{V}^*_L(\beta_2), \ldots, \hat{V}^*_L(\beta_p)\}$ form a random, bootstrap sample from which the new estimates $\hat{\tau}^*$ and $\hat{\kappa}^*$ are obtained. These are called bootstrap replications. Drawing a large number of independent bootstrap samples (e.g. 1000), calculating the corresponding bootstrap replications, the standard error of $\hat{\tau}$ and $\hat{\kappa}$ can be estimated by the standard deviations of empirical distributions of $\hat{\tau}^*$ and $\hat{\kappa}^*$.

In the parametric bootstrap, one first estimates $\hat{\tau}$ and $\hat{\kappa}$ from the image data. Using these estimates in $f_{\mathcal{R}}(\alpha; \theta)$, $N$ new independent images are simulated, $\hat{\tau}^*$ and $\hat{\kappa}^*$ are estimated from each of them, and the standard deviations of empirical distributions of $\hat{\tau}^*$ and $\hat{\kappa}^*$ are calculated. They are estimates of standard errors of $\hat{\tau}$ and $\hat{\kappa}$. In this case, the image model has an important role.

In practice, when considering, for example, the images of paper fibres, it is worth evaluating the magnitude of the standard errors for each type of images but not for all analyses.

## 6.4 Evaluation of Methods by Simulation

We have implemented the proportional and refined methods for orientation analysis through images. Both methods are approximative. Their evaluation is based on the computer-based simulation of digital binary and greyscale images. They are also compared with a gradient-based method used in paper industry.

Let us consider the performances of methods in binary images. In Article D, we simulate fibres modelled by a Boolean model of rectangles (4.1). Fibres have fixed length, fixed width, and orientation distributed by an elliptic model (3.7) with orientation angle $\tau$ and anisotropy $1-e$. In the estimation according to Article C, the critical parameters may be the length intensity $L_A$ and the orientation density $f_{\mathcal{R}}(\alpha;\theta)$ if the resolution of the grid is low. In our simulation, we vary the intensity $L_A$, direction of $\tau$ and anisotropy $1-e$, cf. Article D, Section 3.2. Each realization is presented in the form of a digital binary image such that if any fibre hits a pixel, it has a grey value one.

In the comparison of the new methods, we study the effects of factors $L_A$, $\tau$ and $1-e$ on the estimation of orientation parameters, angle $\tau$ and anisotropy $1-e$. The methods are compared using a matched design. The increase in intensity and the changes of the orientation angle away from 90° worsen the estimation of both parameters obtained by the proportional method. The decrease in anisotropy has effect on the stability of the variogram-based methods due to the optimization and identification problem near the isotropic case. (If $\kappa = 0$, the orientation angle is not defined.) The refined method performs in a larger range than the proportional method, as was expected according to Article C.

In the comparison of the new methods in greyscale images generated by a shot-noise model, we have used the same realizations as in the binary case, see Article D. In the deposition of the fibres, each pixel value is proportional to the number (grammage) of fibres hitting that pixel. In shot-noise cases, the decrease in anisotropy worsen the result of the estimation of both parameters in general. Concerning the estimation of anisotropy, the increase in intensity and the changes of the orientation angle away from 90° worsen the result for both methods. In the estimation of the orientation angle, they mostly seem to affect the result of the proportional method. If the orientation angle is near the direction of the grid, both methods perform quite well. Note that the variogram-based methods are theoretically addressed to Boolean and dead leaves models.

To simulate dead leaves models, we have performed a few experiments, not as systematically as for the binary and shot-noise models. We apply a Boolean model $\Xi$ of elliptical segments. The length and the orientation of a line segment are distributed independently by a uniform and an elliptic model (3.7), respectively. In Article A, B and C, a dead leaves random functions model where uppermost fibres are lightest is used. According to our experiments, in strong anisotropic cases the refined method seems to work better than the proportional method although we do not have any rigorous proof for the present.

The new methods are compared with a gradient method used in paper industry, cf. Article D. In the comparison, the refined method seems to be best for binary models when the intensity of fibres is large, as was expected. In shot-noise cases all methods perform quite well in the ranges of paper parameters. The gradient method is more stable than the variogram-based methods.

## 6.5 Programs

The simulation of images (Articles A, B, and C), the calculations of scaled variograms, theoretical and estimated point intensities are programmed by C and the estimation of the orientation parameters is based on the function *nlm* in Splus. In Article D, the simulation of images is programmed by Matlab in VTT Processes. The gradient program used in Article D is a commercial product.

# Chapter 7

# Discussion

In this work, methods for estimating the fibre orientation distribution in a binary or greyscale image have been developed. The methods are based on observations made on the sampling lines in several directions in the image. Observation along each line is a scaled variogram of pixel values. The stereological connection between the observed scaled variograms and the fibre orientation has been considered in a binary image of a Boolean model of thickened segments and in a greyscale image of a dead leaves model. The methods are based on these connections. Assuming a parametric orientation distribution, the orientation parameters can be obtained numerically from least-squares type procedures. The proportional method is theoretically addressed to binary or greyscale images generated by a dead leaves model, whereas the refined method is directed to binary images. The methods are evaluated with each other using both simulation and real data and compared with a method used in paper industry. From the point of view of variogram-based methods, the results concerning the shot-noise models are empirical.

We have studied the effects of thickness, overlap, length intensity of fibres, grey levels and digitization on the estimation of fibre orientation of Boolean segments. The variogram-based methods rely on the determination of the boundary length in terms of directions when considering thick fibres. In the case of digital binary or greyscale images, the detection of the boundary is degraded due to grey levels and digitization. High length intensity in binary images worsens the estimation since the image is saturated. In the case of the dead leaves model, more information on grey levels is available than in the binary case. The idea of the boundary detection works. In shot-noise images, the information on the boundary is degraded in the sense of scaled variograms due to the additive process. Thus the results of simulation are acceptable. Digitization according to simulation has been noticed to be an important variable affecting the results of the estimation. There are two possibilities to improve the methods. In Serra (1982), a honeycomb lattice was used instead of a square one. In our case, the honeycomb lattice would enable the use of three similar directions. Another way, also used in Serra (1982), is to rotate the image of a squared lattice instead of rotating the sampling line. Both of these improvements should be applied in the laboratory, and they can, therefore, be time-consuming.

The methods are not specific for the model we have used as the orientation distribution. The distribution can be changed. Our methods are, however, addressed for anisotropic systems: An identification problem exists near isotropic cases. In Chaix and Grillon (1996), a method for examining the isotropy from binary images is presented. It can be one choice for checking the isotropy.

On a large scale of imaged paper fibres, a germ-grain model allowing flocculation is a relevant model. In this thesis, however, we concentrate on the development of orientation analysis for paper fibres observed in an image of a small scale. We assume, in that case, that the paper fibres can be approximated by Boolean fibres. Therefore, the orientation analysis on a large scale could be one of the tasks in the future.

# Chapter 8

# Summaries of Original Publications

In Article A, a new statistical method for estimating the orientation distribution of fibres in a fibre process is suggested, where the process is observed in the form of a degraded digital greyscale image. The method is based on the line transect sampling of the image in a few fixed directions. A well-known method based on stereology is available if the intersections between the transects and fibres can be counted. We extend this to the case where, instead of the intersection points, only scaled variograms of grey levels along the transects are observed. Non-linear estimation equations for a parametric orientation distribution are given as well as a numerical algorithm. The method is illustrated by a real-world example and simulated examples where the elliptic orientation distribution is applied. In its simplicity, the new approach is intended for an industrial on-line estimation of fibre orientation in disordered fibrous materials.

In Article B, the estimation of fibre orientation is studied for fibre systems observable as a blurred greyscale image. The estimation method is based on scaled variograms observed along a set of sampling lines in different directions. The parameters of the orientation distribution are obtained numerically. Simulated data are used to study the statistical properties of the method.

In Article C, the orientational characteristics of fibres in digital images are studied. The fibres are modelled by a planar Boolean model whose typical grain is a thick (coloured) fibre. The aim is to make stereological inference on the rose of directions of the unobservable central fibres from observations made on a digital image of the thick fibres. For central fibres, the relation between the rose of directions and the point intensity, observed on a sampling line, is known. We derive, under regularity conditions, the relation between the unobservable point intensity and the scaled variogram observed on the line in a binary and a greyscale image. Using such a relation, it is possible to draw inference about the rose of directions from the scaled variogram, which is easy and fast to determine in a digital image.

In Article D, we recall two categories of algorithms for estimating fibre orientation distribution from an image of a spatial fibre system. In the first algorithm, the estimate is a magnitude-weighted distribution from angles perpendicular to the directions of the gradients in the image. The second algorithm is based on the scaled variogram of grey values scanned along a sampling line and on its relation to the fibre orientation distribution. Using lines in several directions and assuming a parametric model for the orientation distribution, the orientation parameters are estimated numerically from a least-squares type procedure. Two versions of variogram-based methods are used in this work. We compare the potential of these three methods by simulated images of fibrous layers and their thresholded versions. All the methods were found to reproduce the original distribution with a good accuracy in the case of greyscale images where grammage, anisotropy and orientation angle are within the typical ranges of paper parameters. On the contrary, the variogram-based methods seem to handle the estimation of anisotropy in binary images more efficiently.

# Bibliography


Beneš, V. (1989). A practical approach to the stereology of anisotropic structures. *J. Microsc.*, 154:165–175.

Beneš, V. and Gokhale, A. M. (2000). Planar anisotropy revisited. *Kybernetika*, 36:149–164.

Buffon, G. L. L. (1777). Essai d' arithmétique morale. *Suppl. à l'Histoire Naturelle, (Paris)*, 4.

Chaix, J.-M. and Grillon, F. (1996). On the rose of directions measurements on the discrete grid of an automatic image analyser. *J. Microsc.*, 184:208–213.

Chapman, J., Kole, D., and Hellstrom, A. (2001). Papermaking with online fibre orientation measurement. *Pap. Technol.*, 42:43–46.

Corte, H. and Kallmes, O. J. (1962). Statistical geometry of a fibrous network. In *Formation and structure of paper*, pages 13–46, Transactions 2nd Fundamental Research Symposium, Oxford 1961. Ed. F. Bolam. BPBMA London.

Cresson, T. (1988). *The Sensing, Analysis and Simulation of Paper Formation*. PhD thesis, State University of New York, Syracuse.

Danielsen, R. and Steenberg, B. (1947). Quantative determination of fibre orientation in paper. *Svensk Papperstidn.*, 50:301–305.

Deng, M. and Dodson, C. T. J. (1994). *Paper an Engineered Stochastic Structure*. Tappi Press, Atlanta.

Dodson, C. T. J. (1971). Spatial variability and the theory of sampling in random fibrous networks. *J. Roy. Stat. Soc. B*, 33:88–94.

Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. Chapman & Hall, New York.

Erkkilä, A.-L., Pakarinen, P., and Odell, M. (1998). Sheet forming studies using layered orientation analysis. *Pulp Paper Canad.*, 99:39–43.

Erkkilä, A.-L. (1995). Paperin kerroksellisen orientaation mittaaminen ja syntymekanismit. PhL thesis, University of Jyväskylä, Jyväskylä.

Forgacs, O. L. and Strelis, J. (1963). The measurement of the quantity and orientation of chemical pulp fibres in the surfaces of newsprint. *Pulp Paper Mag. Canad.*, 64:T3–T13.

Glasbey, C. A. and Horgan, G. W. (1995). *Image Analysis for the Biological Sciences*. John Wiley, Chichester.

Hanisch, K.-H. (1981). On classes of random sets and point processes. *Serdica*, 7:160–167.

Hilliard, J. E. (1962). Specification and measurement of microstructural anisotropy. *Trans. Metall. Soc. Am. Inst. Metall. Eng.*, 224:1201–1211.

Hutten, I. M. (1994). Paper machine evaluation by fiber orientation profile analysis. *Tappi J.*, 77:187–192.

Jeulin, D. (1989). Morphological modeling of images by sequential random functions. *Signal Proc.*, 16:403–431.

Jeulin, D. (1993). Random models for morphological analysis of powders. *J. Microsc.*, 172:13–21.

Jeulin, D. (2000). *Variograms of the dead leaves model*. Research Report N-31/00/MM, Paris School of Mines, Paris.

Johansson, J.-O. (2002). *Models of Surface Roughness with Application in Paper Industry*. PhD thesis, Lund University, Lund.

Kallmes, O. J. and Corte, H. (1960). The structure of paper: the statistical geometry of an ideal two dimensional fiber network. *Tappi*, 43:737–752.

Kanatani, K.-I. (1984). Stereological determination of structural anisotropy. *Int. J. Engng. Sci.*, 22:531–546.

Kellomäki, M., Pawlak, J. J., Sung, Y.-J., and Keller, D. S. (2001). Characterization of non-stationary structural non-uniformities in paper. In *12th Fundamental Research Symposium, Keble College, Oxford, UK*, pages 1313–1342. FRC, UK.

Kärkkäinen, S. (1999). Estimation of orientation characteristics of disordered fibrous material through digital images. PhL thesis, University of Jyväskylä, Jyväskylä.

Loewen, S. (1997). Fibre orientation optimization. *Pulp Paper Canad.*, 98:67–71.

Mardia, K. V. (1972). *Statistics of Directional Data.* Academic Press, London.

Marriott, F. H. C. (1971). Buffon's problem for non-random directions. *Biometrics*, 27:233–235.

Matheron, G. (1971). The theory of regionalized variables and its applications. Cahiers du Centre de Morphologie Mathématique Vol. 5, 211 pp., *Paris School of Mines Publication.*

Matheron, G. (1972). Ensembles fermés aleátoires, ensembles semi-markoviens et polyèdres poissoniens. *Adv. App. Prob.*, 4:508–541.

Matheron, G. (1975). *Random Sets and Integral Geometry.* John Wiley, New York.

Matheron, G. (1982). *La Destructuration des Hautes Teneurs et le Krigeage des Indicatrices.* Research Report N-761/82/G, Paris School of Mines, Paris.

Mecke, J. (1981). Formulas for stationary planar fibre processes III – intersections with fibre systems. *Math. Operationsforsch. Statist., Ser. Statistics*, 12:201–210.

Mecke, J. and Stoyan, D. (1980). Formulas for stationary planar fibre processes I – general theory. *Math. Operationsforsch. Statist., Ser. Statistics*, 11:267–279.

Molchanov, I., Stoyan, D., and Fyodorov, K. (1993). Directional analysis of planar fibre networks: application to cardboard microstructure. *J. Microsc.*, 172:257–261.

Molchanov, I. S. and Stoyan, D. (1994). Directional analysis of fibre processes related to Boolean models. *Metrika*, 41:183–199.

Penttinen, A. and Stoyan, D. (1989). Statistical analysis for a class of line segment processes. *Scand. J. Statist.*, 16:153–168.

Rataj, J. and Saxl, I. (1988). Analysis of planar anisotropy by means of the Steiner compact: a simple graphical method. *Acta Stereol.*, 7:107–112.

Rataj, J. and Saxl, I. (1989). Analysis of planar anisotropy by means of the Steiner compact. *J. Appl. Prob.*, 26:490–502.

Rataj, J. and Saxl, I. (1992). Estimation of direction distribution of a planar fibre system. *Acta Stereol.*, 11:631–636.

Redon, C., Chermant, L., Chermant, J.-L., and Coster, M. (1998). Assessment of fibre orientation in reinforced concrete using fourier image transform. *J. Microsc.*, 191:258–265.

Rice, J. (1977). On generalized shot noise. *Adv. Appl. Prob*, 9:553–565.

Schmidt, V. (1985). Poisson bounds for moments of shot noise processes. *Statistics*, 16:253–262.

Serra, J. (1982). *Image Analysis and Mathematical Morphology*. Academic Press, London.

Stoyan, D. and Beneš, V. (1991). Anisotropy analysis for particle systems. *J. Microsc.*, 164:159–168.

Stoyan, D., Kendall, W. S., and Mecke, J. (1995). *Stochastic Geometry and its Applications*. John Wiley, Chichester, 2nd edition.

Thorpe, J. (1999). Exploring fiber orientation within copy paper. In *1999 International Paper Physics Conference, San Diego, CA, USA*, pages 447–458. Tappi Press, Atlanta, GA, USA.

Weibel, E. R. (1979). *Stereological Methods Vol. 2 Theoretical Foundations*. Academic press, London.

Xu, L., Parker, I., and Filonenko, Y. (1999). A new technique for determining fibre orientation distribution through paper. In *1999 International Paper Physics Conference, San Diego, CA, USA*, pages 421–427. Tappi Press, Atlanta, GA, USA.

# Yhteenveto – Summary in Finnish

Digitaalisessa kuvassa havaittujen kuitujen suuntajakauman estimoimiseksi on kehitetty kaksi kuva-analyysiin perustuvaa menetelmää. Lähestymistapa perustuu sekä kuvassa olevien piirteiden stereologiseen havainnointiin että näiden havaintojen ja alla olevan kuiduista muodostuvan systeemin yhteyteen. Tässä työssä harmaasävykuvasta havaitaan yksiulotteisia harmaasävyarvoisia funktioita useisiin suuntiin olevilta suorilta. Harmaasävyjen vaihtelua suoralla mitataan skaalatulla variogrammilla. Suorilta havaittujen skaalattujen variogrammien ja suuntajakauman välistä yhteyttä on arvioitu useille parametrisille kuvamalleille. Kehitetyt estimointimenetelmät perustuvat näihin yhteyksiin. Ensimmäinen näistä menetelmistä on tarkoitettu erityisesti harmaasävykuville (tai binäärisille), jotka ovat peräisin ns. kuolleitten lehtien mallista. Toinen menetelmä, joka on ensimmäisen menetelmän tarkennus, on osoitettu binäärisille kuville. Olettamalla parametrinen suuntajakauma mallin parametrit voidaan estimoida numeerisesti optimoimalla painotettua neliösummaa molempien menetelmien yhteydessä. Uusia suuntajakauman analysointimenetelmiä on verrattu keskenään sekä kaupallisen paperiteollisuudessa laajasti käytetyn gradienttipohjaisen menetelmän kanssa.