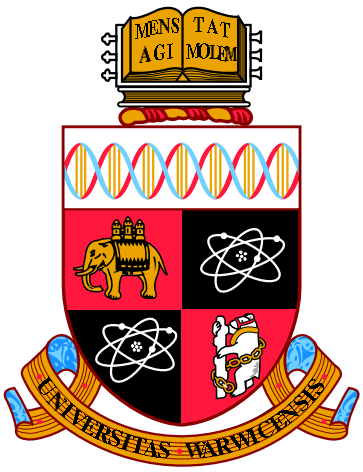


An Orientation Field Approach to Modelling Fibre-Generated Spatial Point Processes

by

Bryony J. Hill

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy

Department of Statistics

September 2011

THE UNIVERSITY OF WARWICK

Contents

List of Tables

List of Figures

Acknowledgments

First and foremost, my deepest gratitude goes to my two supervisors, Elke Thönnes and Wilfrid Kendall who have guided and supported me over the past four years. I have learnt a great deal through working with both Elke and Wilfrid, from developing presentation and writing skills to gaining an inventory of techniques and approaches to tackling problems.

I am grateful to the Århus Mathematics Department who funded my visit to Denmark and especially to Eva Vedel Jensen, Ute Hahn and Markus Kiderlen, for the collaborative discussions and their kind hospitality.

I would also like to thank all the friends I have made throughout my PhD for keeping me sane and making the years enjoyable. Special thanks go to Kat Abrahams, Mouna Akacha, Leo Bastos, Maria Costa, Thais da Fonseca, Flaávio Gonçalves, Stasia Grinberg, Ben Jacoby, Jason Laurie, Silvia Liverani, Chris Nam, Siren Veflingstad, Peter Windridge and Piotr Zwiernik.

I am indebted to to my family who kept me going in the final few months by providing a bit of perspective. Finally, I'd like to thank Richard Tyson for his unwavering support and patience (and for proof-reading my work).

Declarations

I hereby declare that this thesis is the original work of myself, Bryony Hill. This work builds on initial exploratory work by a previous PhD student (Su, 2009) that investigates pore patterns in fingerprints and develops a basic approach to estimating ridge lines. Full attribution has been given to this author, and others, where applicable.

Much of the new work of this thesis is summarised in Hill et al. [2011], including the Bayesian model described in Chapter 3, and the tensor field calculation through a kernel-smoothing interpolation, weighted by the probability that each point is signal (see Chapter 4). The new anisotropy measure, the modified square Fractional Anisotropy ($msFA$), proposed in Chapter 7 was motivated by joint work with Ute Hahn and Eva Vedel Jensen on anisotropy in pressed point patterns.

Abstract

This thesis introduces a new approach to analysing spatial point data clustered along or around a system of curves or *fibres* with additional background noise. Such data arise in catalogues of galaxy locations, recorded locations of earthquakes, aerial images of minefields, and pore patterns on fingerprints. Finding the underlying curvilinear structure of these point-pattern data sets may not only facilitate a better understanding of how they arise but also aid reconstruction of missing data.

We base the space of fibres on the set of integral lines of an orientation field. Using an empirical Bayes approach, we estimate the field of orientations from anisotropic features of the data. The orientation field estimation draws on ideas from tensor field theory (an area recently motivated by the study of magnetic resonance imaging scans), using symmetric positive-definite matrices to estimate local anisotropies in the point pattern through the *tensor method*. We also propose a new measure of anisotropy, the modified square Fractional Anisotropy, whose statistical properties are estimated for tensors calculated via the tensor method.

A continuous-time Markov chain Monte Carlo algorithm is used to draw samples from the posterior distribution of fibres, exploring models with different numbers of clusters, and fitting fibres to the clusters as it proceeds. The Bayesian approach permits inference on various properties of the clusters and associated fibres, and the resulting algorithm performs well on a number of very different curvilinear structures.

Chapter 1

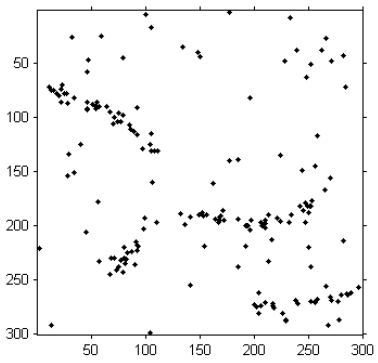
Introduction

Spatial point patterns arise throughout nature as the locations of apparently random objects or events. In statistical analysis, these data are commonly modelled as an instance of a random point process, i.e. a random and locally finite collection of points. There is substantial literature on the statistical analysis of such point data, however most research focuses on rotationally invariant or *isotropic* point processes. The work presented here is concerned with anisotropic point processes, in particular spatial point data clustered around a collection of curves or *fibres*.

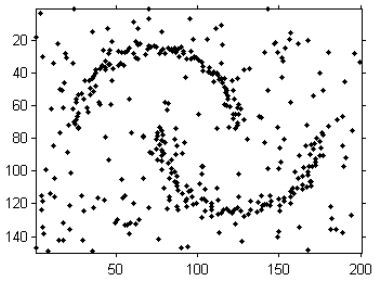
The motivation for this thesis is the identification of systems of fibres generating noisy point processes. Identification of curvilinear elements (i.e. point clusters resembling curves) and elucidation of their relationship with the point data is both an interesting theoretical problem and a useful tool for gaining insight into the origins of the data.

Point patterns exhibiting a filamentary structure often arise in nature when events occur near some latent curvilinear generating feature. For example, earthquakes occur around seismic faults which lie on the boundaries of tectonic plates and hence are naturally curvilinear. Similarly, sweat pores in fingerprints lie on the fingertip ridges lines which have a curvilinear structure. Estimation of ridge lines from the pore pattern could be used to develop a process for reconstructing smudged or patchy fingerprints. Figure 1 presents examples of these data together with two simulated examples of point patterns clustered around underlying families of curves with additional background noise. Our approach is flexible to the features of fibres, producing consistently strong results when applied to each of the four examples in Figure 1; these results are presented in Chapter 6.

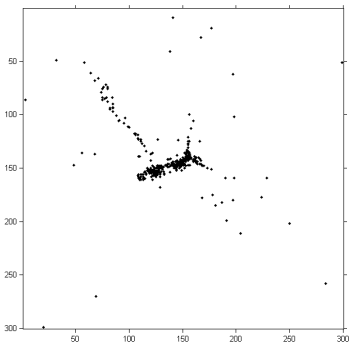
Other data exhibiting a curvilinear structure include land mines located on thin strips of minefield amongst background clutter; cenotes (or sinkholes) clustered



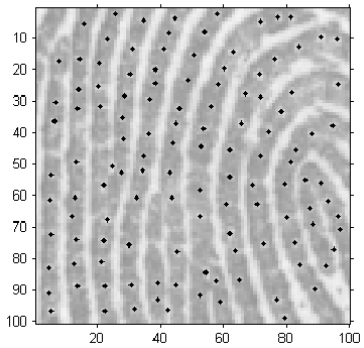
(a) Simulated point pattern.



(b) Simulated point pattern from Stanford and Raftery [2000].



(c) Earthquake epicentres in the New Madrid region. Data are taken from the earthquake catalogue at CERI (Center for Earthquake Research and Information).



(d) Pores along ridges of a section of fingerprint a002-05 from the *NIST (National Institute of Standards and Technology) Special Database 30* [Watson, 2001].

Figure 1.1: Four examples of point patterns clustered around latent curvilinear features with background noise.

around the edge of the Chicxulub crater on the Yucatán Peninsula; and galaxies that cluster in filaments around huge voids creating a 3-dimensional web-like structure. The detection of minefields is a high priority for defence forces, which has prompted investment from the United States Navy into project COBRA, the development of purpose-built, unmanned reconnaissance aircraft (Witherspoon et al., 1995). Estimation of the radius of the Chicxulub crater may clarify the extent to which the asteroid impact affected prehistoric life (Hildebrand et al., 2002). Analysis of the large-scale distribution of matter and identification of the 3-dimensional cosmic web is a subject of great scientific interest, see Martínez and Saar [2002] for further details. These data are described in further detail in Section 2.2.1.

The majority of the current approaches to estimating curvilinear features from noisy point patterns (see Section 2.2.2) provide only a point estimate of the fibres or the

associated point clustering. We show how properties of the underlying distribution of fibres can be estimated using Monte Carlo techniques applied to the spatial point data. This approach has the significant advantage that it can be used to quantify uncertainty on a range of parameters and does so effectively for different types of curvilinear structure. The use of a field of orientations to identify fibres leads to a strong performance on data such as the fingerprint pore pattern shown in Figure 1.1(d), despite the difficulty of there being noticeable alignment of points perpendicular to the fibres.

The data in consideration typically arise from multiple fibres. Simultaneous estimation of the number and location of the generating fibres is a difficult problem; existing approaches to estimating the number of isotropic clusters in a point pattern are generally not-well suited to the long curvilinear clusters. Our approach uses trans-dimensional Monte Carlo methods to explore the full posterior distribution of the set of fibres, thus models with different numbers of fibres can be compared directly.

An important consideration in modelling fibre-generated point processes is the choice of state space for the random fibre process. There is no single natural choice for this state space, although it is generally assumed that the random fibres are smooth and continuous. A common approach is to approximate the smooth fibre by a piece-wise linear curve, however this results in the loss of the curvilinear details of the fibres. The model introduced here describes families of non-intersecting curves *via* a field of orientations (a map from the window of observation W to $[0, \pi)$ assigning an undirected orientation to each point in the window). The curves are identified by segments of streamlines integrating the field of orientations. We say that a curve *integrates* the field of orientations if the curve is continuous and its tangent agrees with the field of orientations at each point. The term *streamline* is used to describe a curve which integrates the field of orientations and has no end points in the interior of the window $W \backslash \partial W$. This novel approach to modelling the generating fibre process permits, in principle, any smooth collection of non-intersecting fibres.

Note that we work with an orientation field rather than a vector field (a map from the window of observation W to $[0, 2\pi)$). The distinction is drawn between the two in fingerprint analysis where a field of orientations is used to model the ridge lines, see for example Ratha et al. [1995], and in engineering when studying the orientation of fibres forming in compressed fluids (Lee et al., 1997).

We choose to use a variant on an empirical Bayes approach to estimate the field of orientations, since a fully Bayesian approach would involve infinite dimensional distributions and be computationally very intensive. The empirical Bayes component consists of estimating the field of orientations from the data *via* a tensor field

as detailed in Chapter 4. In this work, a tensor field is represented by the assignation of a symmetric positive definite matrix to each point of the planar window. Tensor fields of this kind play an important role in diffusion tensor imaging (DTI), as reviewed in Chanraud et al. [2010]. The field of orientations is constructed by simply calculating the orientations of the representative matrices' principal eigenvectors; singularities in the field of orientations correspond to points where there is equality of the two eigenvalues. This empirical Bayes approach enables the reliable estimation of orientation fields (estimates are integrated by fibres producing high likelihoods), through the extension of previous work on tensors by Su et al. [2008], Su [2009] and in diffusion tensor imaging (Dryden et al., 2009). Essentially, tensors that estimate the local orientation of alignments in the point pattern data are smoothly interpolated to create a field of tensors; the orientation field is determined from this tensor field.

The following chapter provides a background in the relevant areas of statistical analysis, together with an overview of existing approaches to solving the problem of identifying filamentary structure in point pattern data. The original contributions of this thesis begin in Chapter 3 with a full description of the proposed Bayesian model; details and justification of the empirical Bayes approach to estimating the field of orientations are given in Chapter 4. Chapter 5 specifies the associated rates and acceptance probabilities of a birth-death Markov chain Monte Carlo process (BDMCMC) algorithm, which is used to draw samples from the posterior distribution of fibres given a particular instance of the point process. Results of the implementation of our approach on the four data sets depicted in Figure 1 are presented in Chapter 6.

In Chapter 7 we return our focus to the positive-definite symmetric tensor, a mathematical object used in the orientation field estimation of Chapter 4 to summarise directional information in point patterns. This penultimate chapter describes how tensors can be used to measure *anisotropy*, the extent to which a point process deviates from isotropy. Here we propose a new measure of anisotropy, motivated by our choice of tensor estimator, the *tensor method* (Su et al., 2008 and Su, 2009). An analysis of the robustness of the tensor method is presented, and we propose some potential applications of anisotropy measures in fibre-generated point processes. Possible areas for further research are suggested in Chapter 8, together with a discussion summarising the work of this thesis. Much of this work is reported on in Hill et al. [2011].

Chapter 2

Background and Related Work

This thesis is primarily concerned with anisotropy (a lack of directional invariance) in spatial point patterns. More specifically, it focuses on point processes that exhibit curvilinear structure, with the aim of making inferences on the generating fibre process given an instance of the point process. This chapter presents an overview of the relevant statistical theory for this thesis, together with a summary of existing approaches to the analysis of fibre-generated point processes.

Section 2.1 provides some examples of how anisotropy can appear in point pattern data together with some of the known approaches to analysing this data. Anisotropic point processes in which points are clustered around a number of curvilinear features or *fibres* are described in Section 2.2. The aim of this thesis is to infer properties of the fibres given an instance of such a point process. We appraise existing approaches to solving this problem, and briefly discuss their strengths and limitations.

The treatment advocated in this thesis is based on the formulation of a general Bayesian model for families of curves and the point patterns clustered around them. Samples are drawn from the posterior distribution of fibres using a birth-death Markov chain Monte Carlo (BDMCMC) process. Section 2.3 gives an introduction to fully Bayesian and empirical Bayes techniques, together with an overview of relevant work in Markov chain Monte Carlo methods.

An important consideration in the formulation of the model is how to define the random fibre process. We choose to model fibres as segments of streamlines that integrate a smooth *field of orientations* $v_{\mathrm{FO}} : W \to [0, \pi)$ where $[0, \pi)$ represents the space of planar directions (with 0 and π identified); orientation fields are discussed in Section 2.4.

This thesis builds on the initial exploratory work of Su et al. [2008] (see also the earlier PhD thesis Su, 2009), which is motivated by the fingerprint pore data. Positive-definite symmetric matrices or *tensors* are used to produce local estimates of point pattern orientations. These tensors form the basis of an empirical Bayes estimation of the field of orientations. Tensors of this form, discussed in 2.5, are used in a range of disciplines to summarise directional information. We describe the uses of tensors in diffusion tensor imaging, a technique in magnetic resonance imaging that has recently stimulated research into tensor analysis, with similar aims to our own. Particular reference is made to current work on the different metrics prescribed for tensors.

2.1 Anisotropy in Spatial Point Processes

We briefly describe the notion of a point process, provide a few examples of anisotropic point processes, and list some available approaches for analysing anisotropic point patterns. The focus of this thesis is on anisotropic point patterns that exhibit curvilinear structure, which are discussed in Section 2.2.

A spatial point process is defined in Stoyan et al. [1995] as a random collection of points in $\mathbb{R}^n$, which is locally finite (each bounded subset contains a finite number of points) and contains no repeated points. We focus primarily on planar point processes over $\mathbb{R}^2$, but many ideas extend naturally to higher dimensions.

Inter-point distances and the local density of spatial point processes have been studied in some detail; Ripley [1981] and Diggle [1983] describe a number of the statistics typically used. The hypothesis that a point pattern is an instance of a homogeneous Poisson point process can be tested using such statistics. They are also helpful for identifying other structures in the point pattern, such as clustering or regularity. However, discussions are usually restricted to stationary (invariant under translation) and isotropic (invariant under rotation) point processes.

A point processes is said to exhibit anisotropy if it is not invariant under rotation. Anisotropy may appear in different forms including the local alignment of points, and global structures, such as anisotropic clusters of points. The focus of this thesis is on the second of these two types of anisotropic point processes, specifically those that exhibit clusters in the form of curvilinear features. This type of point process is described in Section 2.2.

First, we provide a few examples of where such point patterns can be found in nature, followed by an overview of existing approaches to studying anisotropy in point processes.

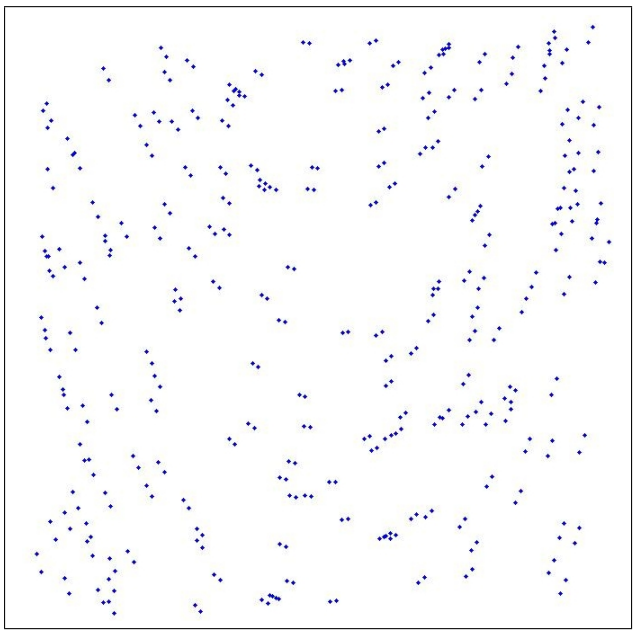


Figure 2.1: A simulated example of a Glass pattern using an exponential transformation.

2.1.1 Examples from Nature

It is suggested (Guan et al., 2006) that the locations of certain shrubs in the North American Mojave desert are directionally associated. Surviving seedlings tend to lie on the north side of existing shrubs where shadow keeps the soil from drying.

Illian et al. [2008] provide two further examples of anisotropy. The first is of 573 carbide particles in rolled steel that tend to cluster in bands parallel to the direction of the rolling. The second example is of two proteins on the surface of a cell that appear to be aligned in pairs of proteins, one of each type, with pairs similarly oriented across the cell.

2.1.2 A Simulated Example - Glass Patterns

Glass patterns, named after the physiologist Leon Glass (see Glass, 1969), consist of a random set of points, superimposed with a geometrically transformed copy. An example is displayed in Figure 2.1. By looking at the point pattern the brain can easily visualise the underlying pattern. These patterns are predominantly used for investigations into psychophysical study of how the brain perceives form. However, they also provide an interesting example of anisotropy in point patterns, and are of particular note as they have a locally parallel structure, similar to the pores on fingerprint ridge lines (see Section 2.2).

Section 2.1.3 describes how Stevens [1978] proposed to identify local orientations in Glass patterns.

2.1.3 Testing for Anisotropy and Estimating Orientation

This section briefly describes some of the existing approaches to analysing anisotropic point patterns.

Nearest neighbour methods, the analysis of second-order orientations and the rose of directions provide the basis for tests of anisotropy and the determination of the global dominant orientation of a point pattern. Steven's method supports the estimation of local orientation within a point pattern.

This thesis uses methodology based on the tensor method, an extension of kernel principal components analysis (kPCA), to estimate local orientations of a point pattern.

Nearest Neighbour and Second-Order Orientation Analysis

Illian et al. [2008] suggest exploring the distribution of the orientations of line segments that connect each point to its nearest neighbour. This is appropriate for identifying local anisotropy (and its direction) where the direction of anisotropy is constant throughout. However, if the direction of anisotropy varies (e.g. in the Glass pattern of Figure 2.1), or arises from anisotropic clusters, this approach is less effective.

A second-order orientation analysis, also described in Illian et al. [2008], is used to explore the distribution of orientations of line segments connecting pairs of points with an inter-point distance lying in some interval $[r_1, r_2]$. The values r_1 and r_2 are usually found by experimentation. This approach is rather more suitable for investigating anisotropic clusters which exhibit anisotropy on a larger scale, such as the 573 carbide particles in rolled steel described previously. However, a global estimate of the orientation of line segments is less informative when the data exhibits anisotropy that varies in orientation.

The distribution of these orientations is equivalent to the *rose of directions* (Stoyan et al., 1995) of the line process given by the collection of lines between pairs of points whose lengths lie in the interval $[r_1, r_2]$, as described in Su et al. [2008]. The rose of directions was primarily developed for use in hypothesis testing and is therefore not well suited to estimating any dominant orientation.

Steven's Method

Stevens [1978] is concerned with the visual processing of point patterns in artificial intelligence, and makes the following hypothesis about Glass patterns.

> One perceives in these patterns a structure that is locally parallel. Our ability to perceive this structure is shown [...] to be limited by the local geometry of the pattern, independent of the overall structure [...]

This idea relates to our approach where the local relative orientation of points is considered first, then interpolated to estimate the global structure.

Steven's method, an extension of the second-order orientation analysis that produces local estimates of dominant orientation in a Glass pattern, proceeds as follows.

A histogram approach is used to produce local estimates of the rose of directions for the lines connecting pairs of points. Each local estimate is based on the lines connecting all pairs of points within a disc (of predetermined fixed radius) centred at a point. The dominant orientation is estimated by smoothing the histogram and choosing the peak of the resulting distribution.

The algorithm presented in this paper produces an effective estimator for the local orientations in Glass patterns, however this approach was not designed for making more general inferences on the properties of an underlying random point process.

Kernel Principal Components Analysis and the Tensor Method

A principal components analysis or PCA (see, for example, Marriott, 1974) is a technique for reducing the dimensionality of a data set by transforming the data to a new coordinate system. Under the new coordinate system, the first coordinate (or first *principal component*) indicates the direction that maximises the variance of the data projected onto the equivalent axis. Each subsequent principal component is orthogonal to all previous components, but similarly maximises the variance of the projected data. A subset, usually the first k principal components for some $k < n$ (n being the dimensionality of the data) are proposed as a new basis for the data.

The principal components are found by identifying the eigen-decomposition of the empirical covariance matrix calculated from the mean-centred data. Specifically, the first coordinate is indicated by the principal eigenvector (with the largest corresponding eigenvalue), and further principal components are given by the eigenvectors corresponding to successively decreasing eigenvalues.

Kernel principal components analysis or kPCA (Schölkopf et al., 1997) is the extension of principal components analysis where the data is first projected on to a different coordinate system, usually of a higher dimension. This permits the detection of nonlinear trends in the data.

The data $y_1, ..., y_m \in \mathbb{R}^N$ are mapped into feature space F by the function $\Phi : \mathbb{R}^N \to F$. In principle, the analysis proceeds following the linear PCA approach on the transformed data $\Phi(y_1), ..., \Phi(y_m)$, i.e. eigenvectors $\mathbf{v}$ and eigenvalues λ are found satisfying

$$\mathbf{C}\mathbf{v} = \lambda\mathbf{v} \tag{2.1}$$

$$\text{where } \mathbf{C} = \frac{1}{m}\sum_{i=1}^{m} \Phi(y_i)^{\mathrm{T}}\Phi(y_i). \tag{2.2}$$

The dimension of covariance matrix $\mathbf{C}$ could be arbitrarily large, depending only on the dimension of feature space F. For this reason the problem is restated as the eigen-decomposition of an N-dimensional matrix, specifically the 'kernel' matrix, defined in terms of a kernel function $k(\cdot, \cdot)$,

$$K_{i,j} = k(y_i, y_j) := \Phi(y_i) \cdot \Phi(y_j) \tag{2.3}$$

(recall N is the dimensionality of the data). We briefly describe the motivation for using the kernel matrix and how the corresponding eigenvectors and eigenvalues relate to the data.

First note that any eigenvector $\mathbf{v}$ solving Equations (2.1) and (2.2) must be spanned by the vectors $\Phi(y_1), ..., \Phi(y_m)$, i.e.

$$\mathbf{v} = \sum_{i=1}^{m} \alpha_i \Phi(y_i). \tag{2.4}$$

Therefore, consider instead the following system of equations:

$$\Phi(y_k) \cdot \mathbf{C}\mathbf{v} = \Phi(y_k) \cdot \lambda\mathbf{v} \quad \text{for all } k = 1, ..., m. \tag{2.5}$$

Substituting Equations (2.4) and (2.2) into Equation (2.5) we see that, for $k =$

$1, ..., m$

$$\Phi(y_k) \cdot \frac{1}{m} \sum_{j=1}^{m} \Phi(y_j)^{\mathrm{T}} \Phi(y_j) \sum_{i=1}^{m} \alpha_i \Phi(y_i) = \Phi(y_k) \cdot \lambda \sum_{i=1}^{m} \alpha_i \Phi(y_i)$$

$$\frac{1}{m} \sum_{j=1}^{m} \sum_{i=1}^{m} \left(\Phi(y_k) \cdot \Phi(y_j)\right) \left(\Phi(y_j) \cdot \Phi(y_i)\right) \alpha_i = \lambda \sum_{i=1}^{m} \Phi(y_k) \cdot \Phi(y_i) \alpha_i. \tag{2.6}$$

Alternatively, this can be written in matrix form using the kernel matrix $\mathbf{K}$ defined in Equation (2.3) and writing the vector $(\alpha_1, ..., \alpha_m)^T$ as $\boldsymbol{\alpha}$:

$$\frac{1}{m} \mathbf{K}^2 \boldsymbol{\alpha} = \lambda \mathbf{K} \boldsymbol{\alpha}. \tag{2.7}$$

Solutions of Equation (2.7) can be found by solving

$$\frac{1}{m} \mathbf{K} \boldsymbol{\alpha} = \lambda \boldsymbol{\alpha}, \tag{2.8}$$

for $\boldsymbol{\alpha}$. The projection $(\Phi(x))$ of the image of a point x onto the $k-$th eigenvector $\mathbf{v}^k$ is given by

$$\sum_{i=1}^{m} \alpha_i^k \Phi(y_i) \cdot \Phi(x) \tag{2.9}$$

where α_i^k is the i-th element in the k-th eigenvector. Hence the projection $\Phi(y_i)$ need not be directly calculated, just the kernel function, $k(x, y) = \Phi(x) \cdot \Phi(y)$.

Examples of typical kernel functions include

$$k(x, y) = (x \cdot y)^d \qquad \text{(polynomial kernels)}, \tag{2.10}$$

$$k(x, y) = \exp(||x - y||^2 / 2\sigma^2) \qquad \text{(radial basis functions)}, \tag{2.11}$$

$$\text{and } k(x, y) = \tanh(a(x \cdot y) + b) \qquad \text{(sigmoid kernels)}. \tag{2.12}$$

The advantage of posing the problem as the eigen-decomposition of kernel matrix $\mathbf{K}$ rather than the covariance matrix $\mathbf{C}$ is that we can choose a high-dimensional feature space with little impact on the computing time required. If the size of the dataset is very large, the data may be de-noised as in Minier and Csató [2007], or by partitioned into smaller subsets (see for example Shi et al., 2009) to reduce the dimensionality of kernel matrix $\mathbf{K}$.

This approach to principal components analysis allows the extraction of nonlinear features from data. The drawbacks are that it is generally not possible to calculate the principal components in the original space making interpretation of the results

non-trivial, and that the dimension of the matrix to be eigen-decomposed grows with the number of data. It is also worth noting that kPCA requires some prior knowledge of the nonlinear features to be extracted, although the scope of this class of features is controlled only by the dimensions of the feature space which may be arbitrarily high at a low computational cost.

In this thesis, we build on the work of Su [2009] and Su et al. [2008], where the *tensor method*, a variant on kPCA, is used to estimate the local orientations of a point pattern. The term *tensor* is used to describe the sum of the outer product of each vector representing a data point with itself. An overview of tensors is given at the end of this chapter.

The tensor method proceeds as follows. Let $P_1, ..., P_n$ denote the points in an instance of a point process. A tensor is created at point P_j by applying a non-linear transformation to the vectors $v^i = (v_1^i, v_2^i) = \overrightarrow{P_j P_i}$ for $i \neq j$. Specifically,

$$\tilde{v}^i = (\tilde{v}_1^i, \tilde{v}_2^i) = \frac{\exp\left(-\left((v_1^i)^2 + (v_2^i)^2\right)/2\sigma^2\right)}{\sqrt{(v_1^i)^2 + (v_2^i)^2}}(v_1^i, v_2^i) \tag{2.13}$$

where σ is a scaling parameter. The Gaussian transformation was chosen because it is continuous, decreases with distance, and the properties of the Gaussian function are well understood.

The tensor at P_j is then calculated by

$$T(P_j) = 2\sum_{i \neq j}(\tilde{v}_1^i, \tilde{v}_2^i)^{\mathrm{T}}(\tilde{v}_1^i, \tilde{v}_2^i). \tag{2.14}$$

The multiple of 2 arises because all vectors $\tilde{v}^i$ are copied and rotated 180 degrees about P_j to centre the mean of the transformed vectors.

Two main differences between the tensor method and kPCA are: (1) - in kPCA the equivalent to the sum in Equation (2.14) is over the vectors between all pairs of points rather than just those including the point P_j; (2) - the tensor method omits the normalising constant $1/(n-1)$, therefore as the number of points increases, so does the 'size' of the tensor.

As with kPCA, the tensor's principal eigenvector gives the principal axis along which the variance of the transformed points are maximised. Hence if the untransformed vectors v^i were projected onto the principal axis, their endpoints (the locations of P_i) would lie relatively close to the initial point P_j suggesting that the principal axis is a good estimate of the fibre orientation.

2.2 Curvilinear Features in Point Processes

We now focus our attention on a particular class of anisotropic point process, those containing long, thin, curved clusters.

A fibre process is a random collection of curvilinear geometric objects; it is a natural generalisation of a line process (see Stoyan et al., 1995). Interest in the literature focuses on the stationarity of fibre processes and the number of intersections with lines or other objects. Stationary fibre processes are often described in terms of their intensity (mean length per unit area) and the rose of orientations given by the orientations of the tangents to the fibres.

A *fibre-process generated Cox process* is a Poisson point process whose driving intensity measure relates to a random fibre process. The name originates from Illian et al. [2008] who present the example of a Poisson point process along an instance of a random fibre process, with intensity λ_f points per unit length of fibre.

In this thesis we consider the more general fibre-process generated Cox process where points are distributed around fibres rather than along them. This type of point process is further generalised to *fibre-process generated point processes*, that depend on a fibre process without the restriction of being Poisson-distributed.

2.2.1 Example Data Sets

Point patterns with a filamentary structure exist in many different areas of study and at greatly varying scales. Some examples are provided below.

Earthquake Epicentres

Earthquake epicentres are typically clustered around an underlying curve structure defined by seismic fault lines. An illustration of the clustering of earthquake epicentres around the world is can be found at `http://pubs.usgs.gov/gip/earthq4/severitygip.html`. There is some interest in using statistical methods to describe the underlying structure, particularly in the principal curve analysis described in Stanford and Raftery [2000].

Minefields

The need to locate minefields before and during assaults makes minefield detection a high priority for armed forces. Reconnaissance aircraft provide images that identify

mines as well as a number of miscellaneous objects or artefacts of a region of interest. This is studied in papers such as Cressie and Collins [2001] and Fraley and Raftery [1998], although currently most published work is only applied to simulated data. The simulated data typically consists of homogeneous Poisson processes on multiple wide strips superimposed on background noise. An example is presented in Figure 2.2. As such, approaches to identifying the minefields generally take no account of the anisotropy or filamentary nature of the point process, partly because it is not evident on a local scale.

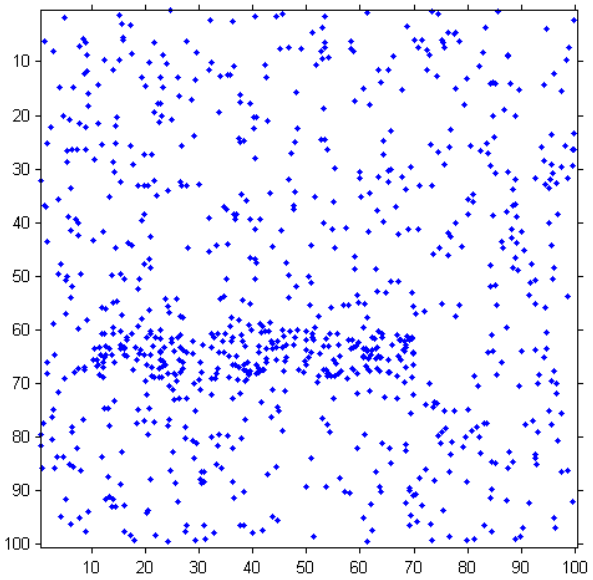


Figure 2.2: A simulated example of minefield data. Dots indicate objects detected through reconnaissance imagery; the dense region of points is suggestive of a minefield.

Pores in Fingerprints

Fingerprints are widely used in forensics, biometric identification and security systems. They have benefits over other forms of biometric identifier (e.g. iris scanning, DNA testing, voice recognition) of being unobtrusive, highly distinctive, relatively permanent and easily collectable (Maltoni et al., 2003). As a result, there has been a large amount of research into the investigation of claims of individuality of fingerprints, creating new recognition and classification systems, and building high resolution fingerprint scanners that capture all the fingerprint details.

Sweat pores are tiny holes along the ridge on a fingertip where the ducts of the sweat glands open. The underlying fibre structure is the dense set of approximately locally-parallel ridge lines (which form the fingerprint) along which pores are located,

usually close to the centre of the ridge (see Figure 2.3).

Ridges on fingertips usually form concentric patterns with loops and/or arches, which help to resist slipping in all directions (particularly concentric patterns). The ridges are constructed from ridge units, each having one sweat gland and one pore opening at some point on its surface. Consequently the distance between adjacent pores on a ridge appears to be proportional to the width of the ridge (see Ashbaugh, 1999).

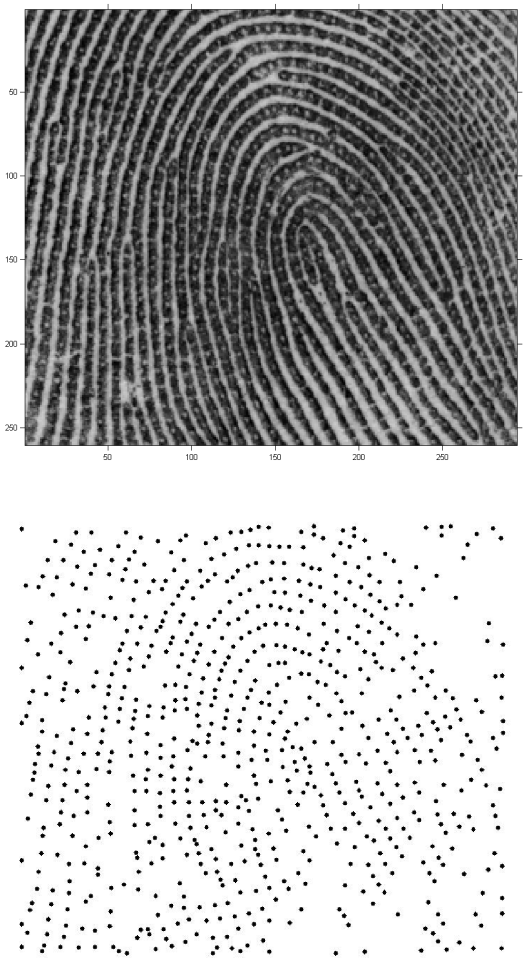


Figure 2.3: Top: Fingerprint a002-05 from the NIST Special Database 30 (Watson, 2001). The sweat pores appear as small light-coloured circles along the ridges. Bottom: The pore pattern of fingerprint a002-05, identified using empirical image analysis techniques (see Su et al., 2008).

An example of a fingerprint and the pattern of pores extracted from it are presented in Figure 2.3. The curved structure of the fingerprint ridge lines is clearly discernible from the pattern of sweat pores shown in the second figure.

An inference of the ridge-line structure from the pore pattern, which is robust to

noise, has potential for aiding reconstruction of patchy fingerprints and may also allow for more efficient storage of fingerprints in huge databases.

Cenotes

A third geographical example is of cenotes (surface connections to underground water bodies), typically found in the Yucatán Peninsula where they are clustered along the circumference of the Chicxulub crater. The cenotes are clustered around just one curve, with non-uniform background noise.

Galaxies

A final application, extending the problem to 3 dimensions, is that of the locations of galaxies in the universe. Galaxies tend to cluster along filaments forming a 3-dimensional web-like structure with large voids between the filaments. In this application the points are clustered around a large number of intersecting curves. Martínez and Saar [2002] describe a number of the statistical methods used to analyse the large-scale structures; however the focus of this thesis is on 2-dimensional data.

A simulation of the web-like cosmic structure can be found at `http://cosmicweb.uchicago.edu/filaments.html`.

Varying Features of Data

The above examples exemplify the variety of features that can be found in this type of data. They have different numbers of curvilinear features, which, in turn, are of varying curvature and thickness. Where multiple curvilinear features exist they may be densely packed or sparsely located, and they can be locally parallel (as in the fingerprint pore data) or connect in the web-like structure of the galaxy data. There are also different types of background noise (e.g. homogeneous, clustered).

The approach described in this thesis is flexible enough to draw inferences on most of these types of data, although it is restricted to fibres that do not cross.

2.2.2 Existing Approaches

Of the existing approaches to identifying curvilinear clusters in background noise, some take no account of the curvilinear nature of the point clusters.

One such approach involves finding the Voronoi tessellation of the point pattern. This is the set of regions that partitions the window of observation (or $\mathbb{R}^2$) such that any point in the space is in the same region as the nearest data point, and no two data points are in the same region. These regions are called Voronoi polygons as they are necessarily polygons for a finite point pattern.

Allard and Fraley [1997] propose a method for detecting the support domain of a uniformly distributed point pattern within a second uniformly distributed point pattern with a larger support domain. They simply take the union of Voronoi polygons that maximises the likelihood of their model. This approach has the advantage that it can be adapted for any shape of point cluster. However, if geometrical constraints such as the number of clusters are made then the maximum likelihood estimator can only be approximated. Also, as with most of these approaches, only a point estimate of the underlying structure is obtained.

Byers and Raftery [1998] propose a method for detecting features in noisy point pattern data, using the distance to the K-th nearest neighbour to separate dense point clusters from the background noise. This produces similar results to the Voronoi tessellation approach of Allard and Fraley [1997], but leads to a classification of points to noise or signal rather than identifying a union of regions in which the feature is expected to lie. While this approach is easily extended to higher dimensions, it is based on the assumption that the signal point pattern is an instance of a Poisson process. Byers and Raftery [1998] also note that the parameter K needs to be chosen with some care.

The following approaches to analysing anisotropic point patterns put a greater focus on identifying the curvilinear structure.

Density-Comparison Approaches

A piecewise linear *Candy Model* is used by Stoica et al. [2005] to model filaments in galaxy data, and extended to the 3-dimensional *Bisous Model* in Stoica et al. [2007]. The Candy Model comprises of random linear objects or *network segments* that 'link' under certain conditions (such as proximity and relative orientation) to form a collection of networks, with connected filamentary structure.

To fit the model to the data, the empirical densities of galaxies in two disjoint regions are compared. The first region is the interior volume of the linear object, and the second is a region surrounding the linear object. Linear objects are more likely to be accepted as part of the filamentary structure if the density is higher in the first region than the second. In three dimensions this extends to comparing

the densities of points in concentric cylinders. An energy function is defined for a network of segments, and a simulated annealing approach is used to determine the network that minimises this energy. This approach is restricted to piecewise linear fibre models, the lengths of the linear segments being limited by the density of points.

A similar approach to the Candy Model is proposed in Arias-Castro et al. [2005] who consider the detection of filamentary structures in point patterns over the $[0,1]\times[0,1]$ square. In particular they test whether the point pattern, consisting mainly of points distributed uniformly at random over the unit square, also contains a set of points sampled from a continuous curve. Note that in this instance the signal points are assumed to lie directly on the curve or fibre.

They proceed by counting the number of data points that lie in thin regions or strips and accepting each strip if it contains more than a certain number of points. Accepted strips are connected if they satisfy certain continuity properties. The null hypothesis that the unit square contains no fibre-dependent points is rejected if the total length of the accepted strips exceeds a predetermined threshold. As is mentioned, this approach does not consider the estimation of the fibre, only the detection to see if one actually exists.

The strips are defined as functions $y = f(x)$ under a Cartesian coordinate system (x, y). Hence, as it stands, the currently proposed method is not rotationally invariant and will not detect some curvilinear structures, although Arias-Castro et al. [2005] have suggested an extension to solve this problem.

Path-Density Approach

A density estimator of the point pattern can be obtained using techniques such as kernel smoothing. Kernel smoothing is a statistical methodology for describing point data by a curve or surface, and is commonly used to estimate the density function given a data sample. The kernel density estimator at a point x is given by

$$\hat{f}(x;h) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{(x-y_i)}{h}\right), \tag{2.15}$$

where $y_1, ..., y_n$ are a data sample, and $K(\cdot)$ is a kernel, typically a positive function of the distance between two points that decreases with increasing distance, with bandwidth parameter h. See Wand and Jones [1995] for further details.

Fibres can be directly estimated from this density; an example of this can be seen in Genovese et al. [2009] where steepest ascent paths along the density estimate are

constructed and the density of these paths is considered an estimator for the density of an underlying fibre process.

The data are modelled as an inhomogeneous Poisson process with density described by a mixture of three components: curves, clusters and background noise. The component arising from fibres is modelled as the convolution of a Gaussian kernel with the fibre; clusters are the equivalent of zero-length fibres. They proceed to show that the paths of steepest ascent (of the empirical estimate of the point process density) concentrate near the fibres of the model. A second kernel smoothing is applied to the paths providing an estimator of the underlying fibre process.

While this technique is an improvement over other approaches that only provide a point estimate of the curvilinear features, it has the shortcoming that it does not implicitly classify points into noise and signal components or support inference of the properties of individual fibres (such as lengths, curvature, etc.). Examples show that the main curvilinear clusters are identified, however, the approach leads to density estimates that require trimming or choosing high level sets. It is also rather sensitive, in that it often identifies artefacts from the background noise as potential clusters.

Minimal Spanning Trees and the Skeleton Model

A further approach discussed in Barrow et al. [1985] is based on the construction of the *minimal spanning tree* of the point pattern. A spanning tree is a set of points and edges (lines between pairs of points), such that all points are included and connected (through paths of edges), and no loops (closed paths) occur. The minimum spanning tree is the spanning tree with minimal total edge length. Reducing the minimum spanning tree, by 'pruning' or removing edges if they fail to meet certain requirements, leaves a simple tree describing the filamentary structure of the point data. It is particularly useful in three dimensions, where it provides a useful insight into the overall characteristics of the filamentary structure. However, it does not provide the means for much further inference and relies on an appropriate choice of the level of pruning. It is also unsuited to walls of galaxies (points aligned in 2-dimensional surfaces in 3-dimensional space).

An alternative to the minimal spanning tree is proposed in Novikov et al. [2008], where a skeleton model is used to describe the structure of a density estimate of a point process. A smooth density estimate is found by applying a kernel smoothing to an instance of the point process. The skeleton is formed by considering curves perpendicular to the iso-contours of the smooth density field, originating from local maxima. The length of the skeleton can be used as a test for Gaussianity of the

random field.

Both of the above approaches are well adapted for identifying branching or bifurcation of fibres, an important aspect of galaxy data, and an aspect that is ignored in our approach.

Principal Curves

An existing method for estimating the curves in the underlying structure of a point process is Stanford and Raftery [2000]'s use of *principal curves* (Hastie and Stuetzle, 1989), a nonlinear generalisation of the first principal component line. Specifically, a principal curve of a density h is defined as a curve f, parameterized by arc-length λ, such that

$$\mathbb{E}\left[X | f_\lambda(X) = \lambda\right] = \lambda \tag{2.16}$$

for almost all λ, where X is a random vector with density h, and $f_\lambda(X)$ is the value of λ such that $f(\lambda)$ is the orthogonal projection of X onto f. The principal curve is fit to the data by iteratively applying this definition.

The approach is based on the assumptions that the background noise arises as a homogeneous Poisson process and features in the point pattern are modelled by normally distributed orthogonal perturbations from points uniformly distributed along unknown curves (identified as cubic B-splines). The features are then combined in a mixture model.

For each combination of number of components and degree of smoothness an optimal clustering of points is estimated. A classification version of the Expectation-Maximisation algorithm is used to cluster the data into features that maximises the likelihood of the model, and simultaneously fit principal curves. An optimal choice of smoothness and number of components is then selected using Bayes factors.

This technique generally performs very well; however it is sensitive to the initial clustering of the data in the Expectation-Maximisation algorithm, and also has difficulties reconstructing fibres where signal points are sparse (for example the fingerprint pore data - Figure 2.3). The authors also mention that a lower bound on the variance of the perturbation of points from curves must be chosen, otherwise the principal curves may be over-fitted.

The remainder of this chapter describes some of the statistical theory and methodology drawn upon within the thesis. Brief explanations are provided here; for a more comprehensive review the reader is referred to the relevant references.

2.3 Bayes, Empirical Bayes and Markov Chain Monte Carlo

The approach proposed in this thesis involves modelling the point process using a Bayesian hierarchical model as formulated in Chapter 3. An empirical Bayes approach is used to estimate the prior of the field of orientations (used to describe fibres) from the data. Properties of the posterior distribution of fibres, conditional on the data, are then estimated using Markov chain Monte Carlo methods.

This section briefly describes how such Bayesian inference proceeds.

2.3.1 Bayesian Inference

Bayesian inference involves estimating features of the posterior distribution determined by Bayes' Theorem (Bayes, 1763) as

$$f(\theta|y) = \frac{f(\theta)L(\theta|y)}{\int f(\theta)L(\theta|y)\,\mathrm{d}\theta}, \tag{2.17}$$

where θ are the parameters of interest, y is the observed data, $f(\theta)$ is the prior on θ and $L(\theta|y)$ is the likelihood function. By sequentially using Bayes theorem, a hierarchical model of priors and hyperparameters is created. This permits great flexibility and allows the propagation of uncertainty throughout the model. Advances in computing over the last 20 years have made it easier to study complex Bayesian models.

Through Bayes' Theorem, point estimates and confidence (or credible) intervals of the posterior distribution of parameters given the data can be found.

2.3.2 Empirical Bayes

Empirical Bayes, a term coined by Robbins [1964], means that the prior distribution (or a Bayes decision rule) is estimated directly from the data. It is argued that all Bayesian methods are empirical as, when postulating the prior, the data is almost always taken into consideration. However, the term *empirical Bayes methods* is used to describe a more rigorous framework in which these empirical estimates are made.

As described in Maritz and Lwin [1989], an empirical Bayes approach is typically implemented when the same experiment is executed repeatedly generating a series of data components. Bayes theorem is then adapted so that the historical data of

previous components can be used in the calculation of the posterior distribution for the current component. The data are used to estimate the prior distribution, or alternatively the Bayes decision rule is directly estimated from the data. Empirical Bayes is an approximation to the fully Bayesian approach described in the previous section.

Criticisms of the empirical Bayes methodology include that it uses the data twice, contradicting the Bayesian philosophy (Gelman, 2008). It also assumes exchangeability of the data components, which is not always reasonable.

It is usually assumed that a hyperparameter η is unknown but can be estimated from the data. The empirical Bayes approach involves estimating this parameter using the marginal distribution of the data,

$$\int L(\boldsymbol{\theta}|\mathbf{y}) f(\boldsymbol{\theta}|\eta)\, \mathrm{d}\boldsymbol{\theta}. \tag{2.18}$$

Here $\mathbf{y}$ is the data, and $\boldsymbol{\theta}$ denotes all other unknown parameters with joint prior density function $f(\boldsymbol{\theta}|\eta)$, conditional on hyperparameter η, and likelihood $L(\boldsymbol{\theta}|\mathbf{y})$.

In parametric empirical Bayes methods (see Carlin and Louis, 2008) it is assumed that there is a family of prior distributions $F(\boldsymbol{\theta}|\eta)$ indexed by η. The parameter η is then estimated (for example, as a maximum likelihood estimator) and plugged back into Equation (2.18) to estimate the posterior distribution.

Empirical Bayes methods reduce the bias in the posterior density associated with choosing hyperparameters. However, it should be noted that empirical Bayes confidence intervals such as highest posterior density intervals often have insufficient coverage, or are too short. This is because they do not account for the uncertainty in the posterior distribution induced by estimating the hyperparameter.

2.3.3 Markov Chain Monte Carlo

Markov chain Monte Carlo (MCMC) methods, which have been used extensively over the past 50 years in statistical physics, are now commonly used in statistics for estimating properties of posterior distributions in Bayesian models.

MCMC methods allow us to draw samples from the posterior distribution without the need to fully evaluate the normalising constant,

$$\int f(\theta) L(y|\theta)\, \mathrm{d}\theta. \tag{2.19}$$

Properties of the posterior density are estimated by consideration of the properties

of these samples.

The idea behind MCMC sampling is that a Markov chain can be constructed, that will explore the state space, and has stationary distribution equal to the target distribution - in this case, the posterior distribution. The Markov chain is typically constructed by proposing moves and accepting or rejecting them according to some calculated probability.

A popular choice of move is the Metropolis-Hastings update. An update θ', of one or more variables θ, is proposed from the proposal density $Q(\theta'|\theta)$ and accepted with probability

$$\alpha = \min\left\{1, \frac{\pi(\theta')Q(\theta|\theta')}{\pi(\theta)Q(\theta'|\theta)}\right\}, \tag{2.20}$$

where $\pi(\cdot)$ is the target distribution. If the proposal density is symmetric, i.e. $Q(\theta'|\theta) = Q(\theta|\theta')$, then the terms cancel leaving a Metropolis update, with acceptance probability

$$\alpha = \min\left\{1, \frac{\pi(\theta')}{\pi(\theta)}\right\}. \tag{2.21}$$

Brooks et al. [2011] provides a recent overview of MCMC methods.

In Chapter 5, MCMC methods are used to sample from the posterior distribution of parameters (including fibres) given the data points. As already identified, our model has the flexibility of not fixing the number of fibres (or point clusters), and so we require a type of MCMC method that enables the exploration of states with different numbers of fibres. This is referred to as a variable dimension problem, and the two main solutions are Reversible-Jump MCMC and Birth-Death MCMC (collectively termed *trans-dimensional MCMC*, see Roeder and Wasserman, 1997).

Reversible-Jump MCMC or RJMCMC is proposed in Green [1995] and extends the Metropolis-Hastings update to a move that varies the number of parameters in the model. Birth-Death MCMC or BDMCMC is a continuous-time approach to the variable dimension problem, and is an extension of the more general birth-death process (see Preston, 1977). RJMCMC and BDMCMC are very similar, indeed a sequence of RJMCMC samplers can be shown to converge to a BDMCMC under an appropriate rescaling of time; see Cappé et al. [2003] for further details.

Reversible-Jump MCMC

In RJMCMC, a move from a k-component state with parameters θ to a proposed state with k' components with associated parameters θ' is accepted with probabil-

ity

$$\alpha = \min\left\{1, \frac{\pi(k', \theta')Q(k|k')Q(u')}{\pi(k, \theta)Q(k'|k)Q(u)} \left|\frac{\partial g(\theta_k, u)}{\partial(\theta_k, u)}\right|\right\}, \tag{2.22}$$

where $Q(\cdot)$ denotes a proposal density and u, u' are the vectors of random variables such that a bijective function $(\theta', u') = g(\theta, u)$ maps the current state to the proposed state see Green [1995] for further details.

The bijective function and random variables can be chosen to create pairs of moves such as birth and death, where a component is created or destroyed without directly affecting the other components, or split and join, where one component is replaced by two similar components, or two components replaced by one, described in Richardson and Green [1997]. These dimension-jump moves, together with moves within a fixed dimension (e.g. Metropolis Hastings updates), form a RJMCMC.

Birth-Death MCMC

The BDMCMC as described in Stephens [2000a] is an extension of the spatial Birth-Death process described in Ripley [1979]. As with the RJMCMC, it is used to draw samples from a posterior distribution with an unknown number of components, however, here the time scale is continuous and events occur at a predetermined or calculated rate. As is evident from the name, the two main types of event are birth and death moves.

During a birth move a component is proposed from some birth density, and during a death move a component is deleted. This is similar to RJMCMC, the main difference being that rather than proposing and then accepting or rejecting moves, the events are proposed at varying rates and always accepted. Hence, the rates of birth moves and the death rates of components are chosen so that detailed balance holds and the limiting distribution of the chain will therefore be the target distribution. Rejection sampling is sometimes incorporated into the birth proposal where it is not feasible to draw samples directly from the birth density. Other moves, can also be proposed at some rate using Metropolis-Hastings probabilities to accept or reject them, for example, split and combine (join) moves are implemented in Cappé et al. [2003].

There is an issue with processing the output of RJMCMC and BDMCMC, known as the *label switching* problem (see Jasra et al., 2005), in that it is not trivial to identify components across samples. Hence, the marginal distribution of parameters of individual components are often unidentifiable. One approach to solving this problem is to put artificial identifying constraints on the components as described in McLachlan and Peel [2000]. This can have the undesirable effect of causing a bias on

output statistics. Suppose each component is identified by some parameter μ_i with the constraint $\mu_1 < \mu_2 < ... < \mu_i < ...$, then estimates of μ_i calculated from a series of MCMC output samples would be 'pushed apart'. This is because, in estimating μ_1 for example, we are estimating the random variable taking the minimum value of all component means, rather than the mean value of the component believed to have the lowest mean. Stephens [2000b] provides a brief overview of existing solutions to this problem.

There is particular difficulty in choosing convergence diagnostics for RJMCMC and BDMCMC as most parameters, on which convergence diagnostics are typically based, are non-identifiable across samples. The approach described in Richardson and Green [1997] is to test the number of components k for convergence, and then test the convergence across samples with fixed k. This has the disadvantage that models with a particular number of components may be so infrequently sampled that it is difficult to determine whether they have converged. Both Brooks and Giudici [1998] and Castelloe and Zimmerman [2002] suggest alternative convergence diagnostics suitable for RJMCMC, based on the work of Gelman and Rubin [1992], that compare the variation of a random variable between chains, within chains, and between models.

Full details of the BDMCMC algorithm are provided in Chapter 5.

The two final sections of this chapter describe two mathematical objects - the *field of orientations* which is instrumental in our construction of a random fibre process; and the *tensor*, used in the empirical Bayes estimation of the field of orientations.

2.4 Field of Orientations

We define a *field of orientations* (or simply an *orientation field*) as a map ν_{FO} from the window of observation W to the interval of orientations $[0, \pi)$, where 0 is associated with π. The interval of orientations corresponds to the collection of points on a circle of unit radius where antipodal points are equivalent. Thus, the field of orientations can be thought of as the scalar field obtained by projecting a vector field onto the half unit circle.

The integral curve of a field of orientations is defined as a map $\gamma : I \to W$ where I is a real interval, and $\left|\frac{\partial \gamma(t)}{\partial t}\right| = |\upsilon_{FO}(\gamma(t))|$. Theory from dynamical systems (see, for example, Irwin, 1980) tells us that if the field of orientations is C^r (has an r-th derivative that is continuous), then the integral curves are also C^r.

Orientation fields are commonly used in fingerprint analysis (see, for example, Mar-

dia et al., 1997 and Ratha et al., 1998), where integral curves of the field of orientations provide a reasonable model for fingerprint ridge lines. Singular points (or singularities) in the field of orientations, where the orientation is undefined, define the overall pattern of the fingerprint.

Like vector fields, fields of orientations are often visualised by evaluating the field over a grid of points and representing each orientation by a fixed length correspondingly oriented. An alternative approach is to integrate the field of orientations and plot the resulting streamlines. Zhang and Deng [2009] describe a method for placing streamlines in a vector field while keeping them as evenly spaced as possible.

2.5 Tensors

A tensor, frequently used in physics, is the term used for a geometric object that describes a linear relationship between scalars, vectors, or even other tensors. Although a tensor is basis-independent, it is often represented by a multidimensional array. The number of indices of such an array is given by the dimension of the tensor. Depending on the order (or rank) of the tensor it will be represented by a scalar, a vector, a matrix, or some higher dimensional array.

The work presented in this thesis uses only order 2 positive-definite symmetric tensors, which identify with positive-definite symmetric matrices.

2.5.1 The Tensor Method

As described in Section 2.1.3 this thesis uses the tensor method as described by Su [2009] and Su et al. [2008], where a variant on kPCA, is used to estimate the local orientations of a point pattern. The equivalent of the empirical covariance matrix calculated from the kPCA determines a tensor at each point in the pattern that estimates the local orientation.

2.5.2 Diffusion Tensor Imaging

Tensors are similarly used in diffusion tensor imaging, or DTI (Basser et al., 1994), to understand brain pathologies such as multiple sclerosis, schizophrenia and strokes. DTI is used to analyse images of the brain collected from magnetic resonance imaging (MRI) machines. The MRI scan detects diffusion of water molecules in the brain and uses the data to infer the tissue structure that limits water flow. It also

helps to identify axons (or nerve fibres) as diffusion is considerably faster parallel to axons.

A 3-dimensional diffusion tensor is used to describe the orientation dependence of the diffusion. The eigenvalues, roughly-speaking, indicate a measure of the proportion of water molecules flowing in the associated eigenvector direction.

Diffusion tensors are constructed at points on a grid (called *voxels*). The tissues in the brain are then identified using fibre tracking techniques (Basser et al., 2000) by assuming that the direction of fibres is co-linear with the principal eigenvector. Fibre tracking can often fail where the voxels are spaced far apart, and so interpolation of the grid of tensors to a smooth *tensor field* has been proposed. Interpolation requires the notion of a mean tensor and hence raises the question of which metric should be used when dealing with tensors and has motivated recent research developments in tensor metrics. A selection of possible tensor metrics are briefly reviewed and compared in Dryden et al. [2009].

Fletcher and Joshi [2007] consider a general statistical analysis on diffusion tensors with the aim of quantifying the variability of the structure of brain matter across patients.

2.6 Conclusions

This chapter lists and briefly describes the main areas of research relevant to this thesis, along with some examples of the type of data in consideration. Over the next five chapters, we present our solution to the problem of identifying curvilinear features in point patterns. The approach is flexible, providing a structure for Bayesian inference on a variety of 'types' of curvilinear structure.

In the following chapter we define our Bayesian hierarchical model for fibre-generated point processes; the empirical Bayesian estimation of the field of orientations is described in Chapter 4. The details of the birth-death MCMC are provided in Chapter 5, and results of this Bayesian inference on simulated and real data are presented in Chapter 6. Chapter 7 further investigates the statistical analysis of tensors, with a focus on anisotropy measures, which describe the extent of the deviation from isotropy, and can be directly calculated from a tensor.

Chapter 3

Bayesian Model for Fibre-Generated Point Processes

xThis chapter describes and details our Bayesian hierarchical model for fibre-process generated point processes. We begin with a statement of the main problem addressed by this thesis.

3.1 Problem and Outline of Solution

The objective is to model, and make inference on, a random point process Π viewed in a planar window $W \subset \mathbb{R}^2$; we write the observed part of the point process as $W \cap \Pi = \{y_1, ..., y_m\}$ for some arbitrary ordering of points. The point process arises from a mixture of homogeneous background noise and an unknown number of point clusters, each clustered along a curve, henceforth called a fibre. A fibre is defined as a one-dimensional object embedded in a higher dimensional space (the space containing the point process). Random sets of fibres or *fibre processes* are discussed in Stoyan et al. [1995] and Illian et al. [2008].

Figure 3.1 presents an example of the type of point pattern under consideration.

We proceed by specifying details of both the structural and probabilistic model, and mention some model variations that have also been considered. A method for analysing the posterior distribution of fibres given an instance of the spatial point process is described in Chapter 5. Chapter 6 provides illustrative examples of our

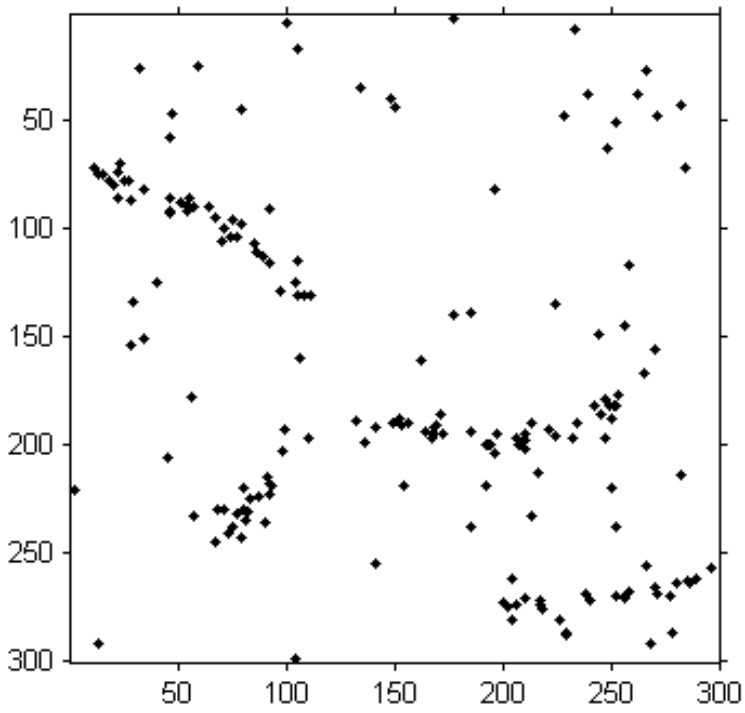


Figure 3.1: Simulated example of a point pattern arising from a fibre-process generated point process.

approach.

See Appendix A for a full table of the notation used in this model summary.

3.2 Hierarchical Bayes Model for Fibre-Generated Cox Process

The system of points and fibres is described using a Bayesian hierarchical model: Bayes rule is repeatedly applied to describe the relation between points, fibres and other parameters. An advantage of using this approach is that inferences can then be made on all parameters in the model. Of the frequentist and Bayesian approaches to inference, the Bayesian is preferred due to its flexibility for working with complex models. A discussion of the advantages of Bayesian inference over frequentist is found in Carlin and Louis [2008], along with a description of Bayesian hierarchical models. An argument against Bayesian statistics is that it is not objective, as two statisticians with different prior beliefs would get different results. However it is often argued by Bayesian statisticians that this is can be an advantage, see for example Goldstein [2006] or Howson and Urbach [1991]. Among the advantages of

Bayesian analysis are that it allows propagation of uncertainty through the model, and that analysis follows directly from the posterior distribution, removing the need for the formulation of further analysis techniques.

The hierarchical model is built sequentially by describing the relations between the model parameters.

The next section outlines the structural model for the points and fibres, and is followed by the full probability model given in Section 3.2.2.

3.2.1 Structural Model

Points

A natural, yet somewhat restrictive, choice is to model the spatial point process as a mixed Poisson process or *Cox process* driven by a random fibre process. However, this model is based on the assumption that points are independently located given the random collection of fibres, a property that does not correspond, for example, with the fingerprint pores which are evenly spaced along ridge lines. Such a point process is called a *fibre-process generated Cox process*, see Illian et al. [2008].

In our model we do associate points with particular fibres but we remove the Poissonian character of the distribution of points along fibres, replacing this by a renewal process based on Gamma distributions for inter-point distances. This allows us to model a tendency to regularity in the way in which points are distributed along a fibre. Extending the model to include the possibility of dependence between points allows us to analyse varying types of data exhibiting curvilinear structure.

Fibres

In this work we use the novel approach of defining fibres as integral curves (or *partial streamlines*) of a field of orientations. This means that at any point on a fibre, the tangent to the fibre agrees with the field of orientations at that point. Note that a field of orientations is equivalent to a vector field except that each point in the field is assigned a *directionless* orientation. An instance of a random field of orientations Υ_{FO} is written as $\upsilon_{\mathrm{FO}} : W \to [0, \pi)$ where $[0, \pi)$ represents the space of planar orientations (with 0 and π identified). A geometric interpretation of the space of planar orientations is that of a unit circle with antipodal points identified.

The simplest way to determine a fibre F is to choose a reference point ω on the fibre and specify the arc lengths of $F \backslash \{\omega\}$. Here we assume that F integrates the

field of orientations. For a fixed field of orientations this will characterise a fibre, although the parametrisation by reference point and length is evidently not unique. We model the fibres in terms of these parameters (the reference points, arc lengths and field of orientations). Consequently, a random fibre is identified by sampling a random reference point and integrating the orientation field in both directions from the reference point to distances respectively determined by the two random lengths.

We note that taking the reference points to be uniformly distributed over the window W will lead to a non-stationary distribution of fibres in that the intensity measure (mean fibre length per unit area) is not constant across W. This issue has been considered and a solution involving modelling the random fibre process via a birth death process with 'time' running tangential to the field of orientations is described in Section 3.3.2.

The field of orientations is a useful intermediary in modelling fibres as it decomposes the construction problem. In practice we seek to identify a suitable field of orientations through analysis of properties of the data; this is the focus of Chapter 4.

Noise

Finally we include background noise in the form of an independent homogeneous Poisson process superimposed onto the fibre-generated *signal* point process.

3.2.2 Probability Model

A Directed Acyclic Graph (or DAG) showing the conditional dependencies for the model is shown in Figure 3.2. The DAG, alternatively called a Bayesian network, provides an effective way to represent the structure of a model; specifically, nodes indicate random variables and directed edges (or arrows) indicate the direction of dependence. See Pearl [1988] for further details.

Here we provide the details of the priors that, together with a likelihood and hyper-priors, form the hierarchical model.

Fibres

Henceforth let $\mathbf{F} = \{F_1, ..., F_k\}$ denote a set of k random fibres. As outlined earlier, the fibre F_j is determined by a reference point ω_j and arc lengths $l_{j,1}, l_{j,2}$. It is also written $F_j = F_j(\omega_j, l_j, \upsilon_{\mathrm{FO}})$ (where $l_j = (l_{j,1}, l_{j,2})$ and υ_{FO} is a field of orientations) to indicate that it is a deterministic function of ω_j and l_j once υ_{FO} is

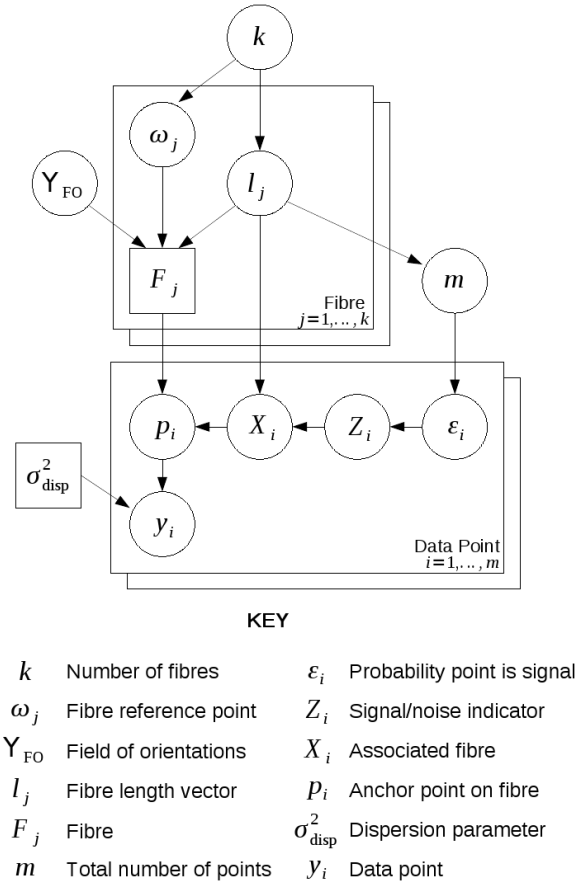


Figure 3.2: Directed Acyclic Graph (DAG) of model: arrows indicate conditional dependencies, elements in squares are deterministically calculated or constant, whilst those in circles are random variables. For simplicity we have not included hyperparameters λ, κ, η, α_{signal} and β_{signal}.

given. This construction is depicted in Figure 3.3. For the list of reference points we write $\boldsymbol{\omega} = \{\omega_1, ..., \omega_k\}$, and the arc length vectors are $\mathbf{l} = \{l_1, .., l_k\}$. We use $l_{j,T} = l_{j,1} + l_{j,2}$ as a shorthand for the total arc length of the jth fibre.

Note that in general the orientation field v_{FO} may possess singularities (where the orientation is undefined), which would constrain the choice of the lengths $l_j = (l_{j,1}, l_{j,2})$. However, it is assumed that the field of orientations contains only finitely many singularities and that the probability of reaching one through integration from a random reference point is 0.

There is an issue with the fibre process that, when a portion of the fibre process over $\mathbb{R}^2$ is viewed through the window W, the reference points of visible fibres crossing the boundary may not lie within W. This is a similar issue to that of

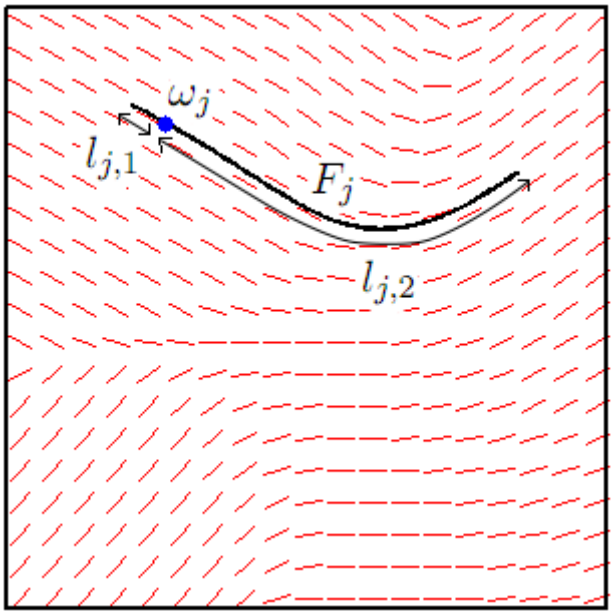


Figure 3.3: The construction of fibre F_j with reference point ω_j and pair of lengths $l_{j,1}, l_{j,2}$. The field of orientations is indicated by the orientations of short lines evaluated over a grid of points.

drawing samples from a cluster point process where daughter points are clustered around unobserved parent points, which may lie outside the window of observation. Approaches to correcting these edge effects have been suggested, see for example Brix and Kendall [2002] and Stoyan et al. [1995]. However we choose to use a minus-sampling approach (see, for example Miles, 1974) and stipulate that each fibre should lie entirely within the window $F_j \subset W$. Assigning a prior probability of 0 to fibres that do not lie completely within W reduces edge effects arising from the unavailability of information on reference points outside the window.

Field of Orientations

We must of course determine the field of orientations to be integrated by these fibres, whether it be a distribution of fields or a single instance. It is computationally advantageous to estimate a single field of orientations which is likely to generate (be integrated by) fibres that fit the data well (produce a high likelihood). The most natural way to do this is to base the calculation of the field of orientations on the data, using an empirical Bayes technique. Full details and justification of the empirical Bayes approach to constructing a field of orientations are given in Chapter 4.

Signal Points

The points are typically clustered around fibres. We use a model that assigns an anchor point p_i on some fibre to each data point y_i. The data point is then displaced from p_i by an isotropic bivariate normal distribution (i.e. $y_i \sim \text{MVN}(p_i, \sigma^2_{\text{disp}}\mathbf{I}_2)$ where $\mathbf{I}_2$ is a 2×2 identity matrix).

The fibre on which anchor point p_i is located is determined by an auxiliary variable X_i, so $X_i = j$ if and only if $p_i \in F_j$. The anchor points on the jth fibre are spaced such that the vector of arc-length distances between adjacent points is proportional to a Dirichlet distributed random variable. This follows directly from modelling the arc-length distances between adjacent anchor points as independently Gamma-distributed random variables, as described in Section 3.2.1. Setting an appropriate parameter for the Dirichlet distribution will encourage points to be either evenly spread, clustered, or placed independently at random along the fibre.

A priori, the probability that point y_i is allocated to the jth fibre ($X_i = j$) is proportional to the total length of fibre F_j. This ensures that the mean number of points per unit streamline remains approximately constant.

Noise Points

Noise is then added as a homogeneous Poisson process. This is included in the model by allocating each point y_i to noise or signal (stored in auxiliary variable $Z_i = 1$ or 0 for signal or noise respectively). Point y_i is allocated to signal independently of the allocations of all other points. The probability that y_i is allocated to signal is given by ε_i. If the point is signal then its location is distributed as outlined in the previous subsection. Otherwise, if the point is noise, it is distributed uniformly across the window W.

Total Number of Points

A Poisson distributed random variable is used to model the total number of points m. The mean total number of points μ_{total} is defined to be equal to some function of μ_{signal} the mean number of signal points, and ρ, a parameter governing the number of noise points. For simplicity we set ρ to be an estimate of the proportion of the total points that are noise points and define $\mu_{\text{total}} = \mu_{\text{signal}}/(1-\rho)$. The assumption that the mean number of noise points is proportional to the mean number of signal points is suited to the fingerprint pore data (see Figure 1) where the fibres are evenly distributed across the window, and noise points arise as artefacts of the

pore detection process. It is less appropriate for the earthquake data (also in Figure 1), where noise points (earthquake epicentres far from the main features) are independent of the signal points. The mean number of signal points μ_{signal} is assumed to be proportional to the total sum of the fibre arc lengths. Hence m is assumed to be Poisson distributed with mean

$$\mu_{\text{total}} \quad = \quad \left(\sum_{j=1}^{k} l_{j,T} \right) \frac{\eta}{1-\rho} \tag{3.1}$$

where $\rho = \beta_{\text{signal}}/(\alpha_{\text{signal}} + \beta_{\text{signal}})$ is the prior estimate of the proportion of points that are signal and η is a density parameter.

Priors

We use the following independent priors:

$$P(k|\kappa) \sim \text{Poisson}(\kappa) \tag{3.2}$$

$$P(\mathbf{l}|k,\lambda) = \prod_{j=1}^{k} P(l_{j,1}|\lambda)P(l_{j,2}|\lambda) \qquad \text{where } l_{j,\cdot} \sim \text{Exp}(1/\lambda) \tag{3.3}$$

$$P(\boldsymbol{\omega}|k) = \prod_{j=1}^{k} P(\omega_j) \qquad \text{where } \omega_j \sim \text{Uniform}(W) \tag{3.4}$$

$$P(\mathbf{F}|\boldsymbol{\omega},\mathbf{l},\upsilon_{\text{FO}}) = \prod_{j=1}^{k} P(F_j|\omega_j,l_j,\upsilon_{\text{FO}}) \tag{3.5}$$

$$\text{where } P(F_j|\omega_j,l_j,\upsilon_{\text{FO}}) = \mathbf{1}_{[F_j(\omega_j,l_j,\upsilon_{\text{FO}})\subset W]}$$

The function determining the fibres $F_j(\omega_j, l_j, \upsilon_{\text{FO}})$ is described in Section 3.2.2, and depicted in Figure 3.3.

$$P(\boldsymbol{\epsilon}|\alpha_{\text{signal}},\beta_{\text{signal}}) = \prod_{i=1}^{m} P(\epsilon_i|\alpha_{\text{signal}},\beta_{\text{signal}}) \tag{3.6}$$

$$\text{where } \epsilon_i \sim \text{Beta}(\alpha_{\text{signal}},\beta_{\text{signal}})$$

$$P(\mathbf{Z}|\boldsymbol{\epsilon}) = \prod_{i=1}^{m} P(Z_i|\epsilon_i) \tag{3.7}$$

$$\text{where } Z_i \sim \text{Bernoulli}(\epsilon_i).$$

Here m is the total number of points in $\{y_1, ..., y_m\}$, and $\mathbf{1}_{[...]}$ is the indicator function.

The prior for the field of orientations $P(\upsilon_{\text{FO}})$ is defined as part of an empirical Bayes step, see Chapter 4.

The prior distribution of the anchor points requires careful construction.

For each fibre in $F_j \in \mathbf{F}$ identify the set of anchor points that lie on F_j as $\mathbf{p}^j = \{p_i : X_i = j\}$, and write $n^j = |\mathbf{p}^j|$ for the number of points in this set. The distances between adjacent points on the ridge, $\mathbf{q}^j = q_0^j, ..., q_{n^j}^j$ are proportional to a Dirichlet distributed random variable. Here, q_0^j and $q_{n^j}^j$ are the distances from each end of the fibre to the nearest point in $\mathbf{p}^j$. We choose a constant vector of length $n^j + 1$ for the parameter of the Dirichlet distribution $(\alpha_{\text{Dir}}, ..., \alpha_{\text{Dir}})$.

Hence, the prior for the anchor points is

$$\begin{aligned} P(\mathbf{p}|\mathbf{F}, \mathbf{X}, \alpha_{Dir}) &= \prod_{j=1}^{k} P(\mathbf{p}^j|F_j, \alpha_{Dir}) && (3.8)\\ &= \prod_{j=1}^{k} P(\mathbf{q}^j|\alpha_{Dir}) \frac{1}{n^j!(l_{j,T})^{n_j}} && (3.9)\\ &= \prod_{j=1}^{k} \frac{1}{B((\alpha_{\text{Dir}}, ..., \alpha_{\text{Dir}}))} \prod_{i=0}^{n_j} (q_i^j)^{\alpha_{\text{Dir}}-1} \frac{1}{n^j!(l_{j,T})^{n_j}} && (3.10) \end{aligned}$$

where $B((\alpha_{\text{Dir}}, ..., \alpha_{\text{Dir}}))$ is the Beta function. The factor $1/(n_j!(l_{j,T})^{n_j})$ arises from the number of orderings of $\mathbf{p}^j$ and the change of variables from $\mathbf{p}^j$ to $\mathbf{q}^j$. Taking $\alpha > 1$ promotes regularity on the spacing of points along the fibre.

For ease of presentation we write

$$D(\mathbf{q}(\mathbf{p}), \alpha_{\text{Dir}}) := \prod_{j=1}^{k} P(\mathbf{p}^j|F_j, \alpha_{Dir}) l_{j,T}^{n_j}\,. \tag{3.11}$$

Posterior

We are interested in the posterior distribution of fibres (and other parameters) given a particular instance of the point process. This posterior is given by

$$\begin{aligned}
&\pi(\mathbf{F}, \mathbf{l}, \boldsymbol{\omega}, k, \upsilon_{\mathrm{FO}}, \boldsymbol{\epsilon}, \mathbf{Z}, \mathbf{X}, \mathbf{p}) \qquad (3.12)\\
&= P(\mathbf{F}, \mathbf{l}, \boldsymbol{\omega}, k, \upsilon_{\mathrm{FO}}, \boldsymbol{\epsilon}, \mathbf{Z}, \mathbf{X}, \mathbf{p}|\mathbf{y})\\
&\propto P(\mathbf{F}, \mathbf{l}, \boldsymbol{\omega}, k, \upsilon_{\mathrm{FO}}, \boldsymbol{\epsilon}, \mathbf{Z}, \mathbf{X}, \mathbf{p})\\
&\quad \times L(\mathbf{F}, \mathbf{l}, \boldsymbol{\omega}, k, \upsilon_{\mathrm{FO}}, \boldsymbol{\epsilon}, \mathbf{Z}, \mathbf{X}, \mathbf{p}|\mathbf{y})\\
&= P(\mathbf{l}|k)P(\boldsymbol{\omega}|k)P(k)P(\upsilon_{\mathrm{FO}})P(\mathbf{F}|\mathbf{l}, \boldsymbol{\omega}, \upsilon_{\mathrm{FO}})P(\boldsymbol{\epsilon})P(\mathbf{Z}|\boldsymbol{\epsilon})\\
&\quad \times P(\mathbf{X}|\mathbf{Z}, \mathbf{l})P(\mathbf{p}|\mathbf{F}, \mathbf{X})L(\mathbf{F}, \mathbf{l}, \boldsymbol{\omega}, k, \upsilon_{\mathrm{FO}}, \boldsymbol{\epsilon}, \mathbf{Z}, \mathbf{X}, \mathbf{p}|\mathbf{y}).
\end{aligned}$$

Here $P(\cdot)$ indicates a prior distribution.

Chapter 5 describes how to sample from this posterior distribution using Markov chain Monte Carlo techniques.

Likelihood

The likelihood in Equation (3.12) is given by

$$\begin{aligned}
&L(\mathbf{F}, \mathbf{l}, \boldsymbol{\omega}, k, \upsilon_{\mathrm{FO}}, \boldsymbol{\epsilon}, \mathbf{Z}, \mathbf{p}, \mathbf{X}|\mathbf{y})\\
&= L(\mathbf{p}, \mathbf{Z}, \sigma^2_{\mathrm{disp}}|\mathbf{y}) \qquad (3.13)\\
&= \prod_{i=1}^{m} \left(\frac{1}{2\pi\sigma^2_{\mathrm{disp}}} \exp\left(-\frac{\mathrm{dist}(y_i, p_i)^2}{2\sigma^2_{\mathrm{disp}}} \right) \mathbf{1}_{[Z_i=1]} + \frac{\mathbf{1}_{[Z_i=0]}}{|W|} \right), \qquad (3.14)
\end{aligned}$$

where $\mathrm{dist}(y_i, p_i)$ denotes the Euclidean distance between points y_i and p_i, and $|W|$ is the Lebesgue measure of window W. From the prior distributions and likelihood we can calculate the necessary death rates, acceptance probabilities and full conditionals required in a birth-death process such that detailed balance holds.

Computational Simplifications

Computer implementation makes it necessary to approximate the field of orientations by a discrete structure. We adopt the simple approach of estimating the field of orientations at a dense regular grid of points over W. Integral curves are calcu-

lated stepwise by estimating the orientation at a point by its value at the nearest evaluated grid point and extending the curve a small distance in that direction. Note that the choice of direction (from the two available for each orientation) is made so that the angle between adjacent linear segments is greater than $\pi/2$. If the angle between adjacent linear segments is precisely equal to $\pi/2$, then a further rule for the choice of direction must be specified, however in practical applications with noisy data, this will happen with probability 0. Similarly, the probability of there being a singularity in the field of orientations at one of the finite grid points over which it is estimated is also 0 (assuming the field of orientations contains only finitely many singularities).

Consequently fibres are stored as piecewise-linear curves and further calculations are performed on these approximations. Of course this discretisation can be arbitrarily reduced (at a correspondingly large computational cost) to improve the accuracy of the approximation.

3.3 Alternative Models

The model, as outlined in this chapter, is deemed an appropriate and sufficiently flexible model for the type of data we are considering. Two of the possible variations on the model which were considered are outlined in this section.

3.3.1 A Fibre-Process Generated Cox Process

In Section 3.2.1 it was suggested that points could be modelled as a Cox process generated by a fibre process. In a Cox process, points are Poisson distributed with intensity function given by a random field over W which, in turn, depends on the fibre process. Such a model gives rise to an independent point process unlike our model where points are perturbed from anchor points, which are Dirichlet-distributed along a fibre.

If, in the construction of our model, we choose a Dirichlet parameter $\alpha_{\mathrm{Dir}} = 1$, so that anchor points are uniformly distributed along the fibres, then the resulting point process is a Cox process. Variations on this Cox process are found by choosing an alternative likelihood function $L(p_i|y_i, Z_i = 1)$, or indeed, dropping the auxiliary variables and anchor points, and determining a likelihood $L(F|y_i)$.

For example, the likelihood could be a function of the distance from y_i to the *nearest* fibre. The intensity function of the resulting Cox process is not quite the same as that of the Cox process derived from our construction, where the intensity is determined

by integrating a kernel along each fibre. However for a field of orientations with low curvature, these intensity measures can be nearly equal. The advantage of using anchor points in the Cox process construction is that the model can be easily extended to control the regularity of points along fibres.

3.3.2 Towards an Unbiased Fibre Process

As described in Section 3.2.1, fibres are identified by partial integral curves of a field of orientations. We have chosen a prior distribution for the fibre process based on the sampling mechanism of drawing a reference point uniformly at random from the window W and integrating the field of orientations to a random length in each direction. However, this prior fibre distribution is biased, in that some regions of the window are more likely to contain a random fibre than others due to the inherent curvature of the field of orientations. The bias is depicted in Figure 3.4, where a large number of fibres have been drawn from the prior distribution. There is clearly a long region from $(140, 143)$ to $(170, 129)$ that is integrated by a greater density of fibres than other areas of the window.

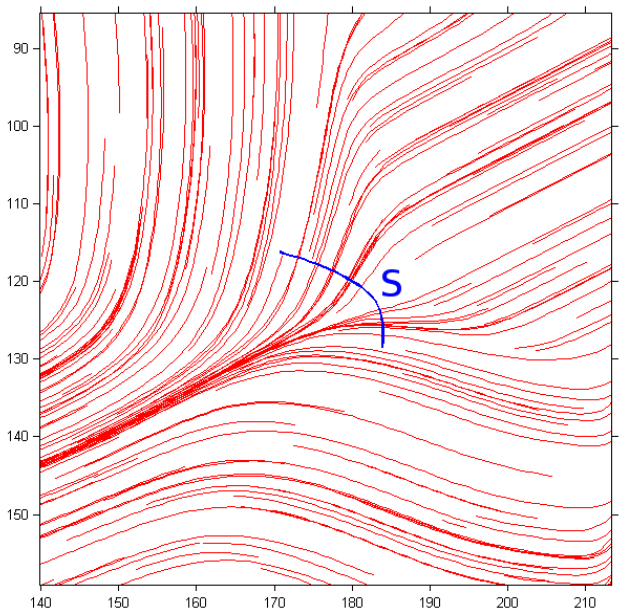


Figure 3.4: A cropped window showing a large sample of fibres drawn from the prior fibre distribution with a diverging field of orientations. There is clearly a bias on the number of fibres we would expect the curve S, orthogonal to the field of orientations, to intersect.

We can quantify the biasing effect by considering the mean number of fibres, $m(S)$,

that intersect the curve segment S which runs perpendicular to the field of orientations. A natural condition to impose, in order to reduce the bias on the density of fibres, is to require that $m(S)$ is proportional to the length of S. However, we need to construct a fibre process model satisfying this condition.

One approach is to model the number of fibres intersecting S as a time-homogeneous birth-death process by setting S_t to be the curve perpendicular to the field of orientations, conditioned on two streamlines on which its end points are located, and meeting one of these streamlines at point t, measured in arc length along the streamline (see Figure 3.5). By taking the two streamlines to be very close, we can arbitrarily choose either streamline to measure the arc length along.

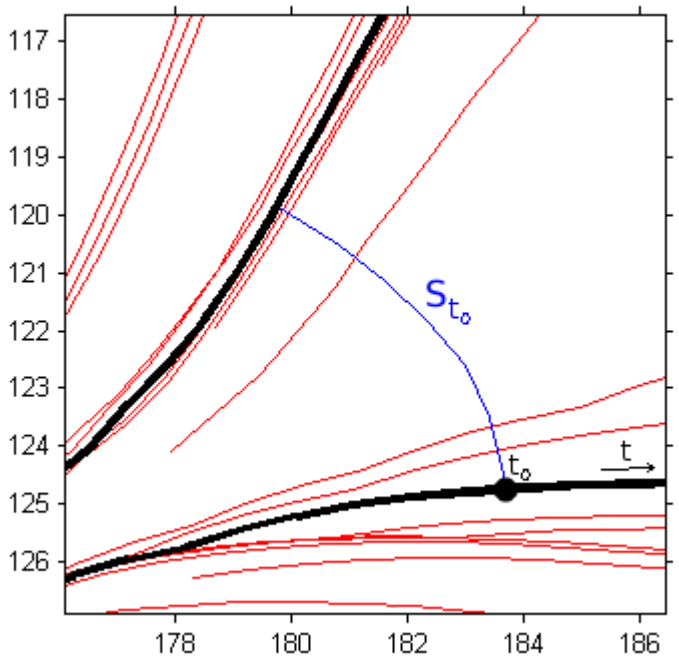


Figure 3.5: A section of Figure 3.4 motivating the construction of a birth-death process. The number of fibres integrating curve S_t will vary as t increases or decreases from t_0. The two thick streamlines that S_t connects are assumed fixed.

The birth-death process describes the number of fibres X_t intersecting S_t as t varies (from $-\infty$ to ∞ in principle). This means that $X_t \to X_t + 1$ at rate b_t, the birth rate, and $X_t \to X_t - 1$ at rate $d_t X_t$, where d_t is the death rate, and both b_t and d_t are functions of time. It is possible to estimate birth and death rates which ensure that the mean number of fibres intersecting S_t is proportional to the length of S_t, for a chosen pair of streamlines. By considering the birth and death rates for an arbitrarily close pair of streamlines we can draw samples from an unbiased distribution of fibres - where 'unbiased' is defined as 'the mean number of fibres

intersecting S_t is proportional to the length of S_t.

In practice, the prior fibre distribution described in Section 3.2.1 leads to simple calculations, and generally there is sufficient data largely to eliminate this bias from the posterior. However, finding an unbiased fibre-process prior is both of theoretical interest and informative as to how unbiased and biased priors differ.

3.4 Conclusions

In this chapter we presented a Bayesian hierarchical model for a general point process exhibiting clustering around an unknown number of curvilinear features. The Bayesian approach is motivated by the inherent complexity of the data clustering and the prior belief that there exists a random fibre process with random points clustered about each fibre. In Chapter 5 we will show how Markov chain Monte Carlo (MCMC) methods can be used to draw samples from the posterior distribution of fibres and other parameters (e.g. signal/noise allocations and fibre lengths), given an instance of the random point process. Examples and statistics of samples drawn from the MCMC methods are given in Chapter 6.

Fibres are identified as partial integral curves of a field of orientations. This is a relatively novel characterisation of fibres in the study of fibre-generated point processes, yet integral curves have been used in other areas such as image analysis (Kass and Witkin, 1987), diffusion tensor imaging (Mori et al., 2001) and fingerprint topology (Sherlock and Monro, 1993). An important consideration is how the field of orientations should be estimated so that integral curves produce high likelihoods, this is the focus of the following chapter.

Chapter 4

Estimation of the Orientation Field

The model, as described in the previous section relies on an appropriate choice of prior for the field of orientations. It is highly desirable to generate a field of orientations which is likely to contain (be integrated by) fibres that fit the data well (produce a high likelihood), as sets of integral fibres sampled conditional on the field of orientations will potentially produce high posterior densities. A natural way to achieve this is to base the calculation of the field of orientations on the data, using an empirical Bayes technique. The use of empirical Bayes to find the prior for the field of orientations distribution means that aspects of the prior, or parameters of the prior, are estimated from the data. This is fully motivated in Section 4.1, and an overview of the construction of the field of orientations estimator is presented in Section 4.2.

Tensors, in particular 2×2 positive-definite symmetric matrices, are instrumental in the construction of the field of orientations estimator. Section 4.3 provides a full description on how we define tensors, and explains how the data are used to make local orientation estimates which are smoothly interpolated to produce a field of orientations estimator. The estimator, based on orientation estimates arising from signal data, is extended to include noise data by weighting the contribution of each point to the field of orientations estimator by how likely it is to be noise or signal. This is described in Section 4.4. The second of these estimators defines the empirical Bayes prior on the field of orientations used in the examples of Chapter 6.

The smooth interpolation induces a bias on the field of orientations in regions of high curvature, defined as the areas where integral curves exhibit high curvature. The extent of the smoothing-induced bias is estimated in Section 4.5, and in Section

4.6 three possible corrections to the bias are described.

First, we justify the use of an empirical Bayes approach, explaining why our prior distribution of the field of orientations depends on the data.

4.1 Empirical Bayes

An empirical Bayes approach (see Robbins, 1956) is used to estimate the field of orientations; the data are directly used in the computation of its prior.

The empirical Bayes approach involves estimating the field of orientations using the marginal distribution of the data,

$$\int L(\boldsymbol{\theta}|\mathbf{y}) f(\boldsymbol{\theta}|\upsilon_{\mathrm{FO}})\,\mathrm{d}\boldsymbol{\theta}. \tag{4.1}$$

Here $\mathbf{y}$ is the data, and $\boldsymbol{\theta}$ denotes all other unknown parameters with prior density function $f(\boldsymbol{\theta}|\upsilon_{\mathrm{FO}})$, conditional on the field of orientations, $\Upsilon_{\mathrm{FO}} = \upsilon_{\mathrm{FO}}$, and likelihood $L(\boldsymbol{\theta}|\mathbf{y})$. The set of fibres $\mathbf{F}$ is the only parameter that directly depends on the field of orientations.

An empirical Bayes approach would usually be implemented by substituting a point estimate $\hat{\upsilon}_{\mathrm{FO}}$ (for example the maximum likelihood estimator) into the posterior distribution:

$$p(\boldsymbol{\theta}|\mathbf{y}, \hat{\upsilon}_{\mathrm{FO}}) \propto L(\boldsymbol{\theta}|\mathbf{y}) f(\boldsymbol{\theta}|\hat{\upsilon}_{\mathrm{FO}}), \tag{4.2}$$

see, for example Maritz and Lwin [1989]. If the distribution $f(\boldsymbol{\theta}|\upsilon_{\mathrm{FO}})$ is fully known then a parametric empirical Bayes approach can be implemented, estimating $\hat{\upsilon}_{\mathrm{FO}}$ directly from the marginal distribution in Equation (4.1). Point estimates like maximum likelihood estimators are often estimated using Expectation Maximisation algorithms (see for example Dempster et al., 1977). However, due to the high dimensionality of the field of orientations and the high complexity of the marginal distribution of the data given the field of orientations, Expectation Maximisation algorithms are unsuitable. Hence we have not attempted to directly estimate the field or orientations from the marginal distribution, but rather we use prior assumptions, such as the smoothness of the field of orientations, to create an estimator based on estimates of the local orientation of the point clusters.

An alternative approach would be to use a fully Bayesian model, where the field of orientations is modelled as an independent random variable Υ_{FO}. A suitable state-space and a corresponding σ-algebra, transition kernel and prior on this state space would all need to be identified. These could be derived from random field theory (see, for example Adler and Taylor, 2007), using an appropriate covariance function

to maintain smoothness in the field of orientations, however there are a number of issues with this approach. In practice sampling a random field of orientations can be computationally expensive, particularly if the covariance function does not have a simple form (as is likely in this model). Calculations relating to the conditional distribution of the field of orientations given the fibres are likely to lead to unfeasible computational complexity. A further issue is that this approach leads to a huge space of possible fibres. It becomes very difficult to ensure that the Markov chain Monte Carlo (MCMC) methodology, designed to draw samples from the posterior distribution of fibres and outlined in the following chapter, adequately explores this space. Using information given in the data helps to limit this space to a more easily explorable restricted class of suitable fields of orientations.

It should be noted that empirical Bayes confidence intervals such as the highest posterior density intervals obtained in Chapter 6 often have insufficient coverage (intervals are too short) as they do not account for the uncertainty induced by estimating the field of orientations (see Morris, 1983). This bias, induced by the field of orientations, is particularly evident in the two simulated examples of Chapter 6.

We refer the reader to Carlin and Louis [2008] for further details of empirical Bayes techniques.

4.2 Overview

We now give an overview of the construction of the field of orientations estimator.

The tensor method described in Su et al. [2008] (and further discussed in Su, 2009) is applied to the point pattern to construct a tensor (identified by a 2×2 symmetric positive-definite matrix) at each point, which summarises the directions and distances of nearby points. To this set of tensors we apply a 2-dimensional Gaussian kernel smoothing in the log-Euclidean metric to produce a field of tensors. The tensor field is represented by an assignation to each point of a 2×2 symmetric positive definite matrix whose principal eigenvector indicates the dominant orientation at that point; the relative magnitude of the eigenvalues indicates the strength of the dominant orientation. The field of orientations assigns the orientation of this principal eigenvector to each respective point. If, at a certain point the principal eigenvector is not unique (the eigenvalues are equal), the field of orientations will not be defined creating a singularity.

Images illustrating the main steps in the estimator calculation are presented in

Sections 4.3.3 and 4.4.

First, we introduce the notion of a tensor.

4.3 Tensors

The tensor is an abstract concept used in a range of disciplines, particularly physics and engineering, and is simply defined as the generalisation of a vector (or matrix). We follow the approach of Su et al. [2008] and identify tensors with positive-definite symmetric matrices. Hence, when we describe properties of a tensor, we are referring to the properties of a symmetric positive-definite matrix. A tensor field is simply the assignation of a tensor to each point in the window W. This is also the definition of a tensor used in diffusion tensor imaging (DTI), we refer the reader to Section 2.5.2 for further details.

4.3.1 Decomposition of Tensors

A tensor of order 2 is represented by an $n \times n$ matrix. The size n of the matrix corresponds to the dimensionality of the data; we work with $n = 2$ as $W \subset \mathbb{R}^2$, but in the extension to 3 dimensions (see for example the galaxy data in Section 6.3.2) 3×3 tensors are used.

A 2×2 tensor T is written

$$T = \begin{pmatrix} A & B \\ B & C \end{pmatrix}, \tag{4.3}$$

where $AC > B^2$ and $A, C > 0$. The tensor represented by this matrix is used to summarise directional information. The principal eigenvector of T indicates the dominant direction, while the ratio of the eigenvalues describes 'how dominant' this direction is. This ratio is a measure of anisotropy, see Chapter 7 for some other measures of anisotropy. Equal eigenvalues indicate isotropy - a lack of a dominant direction. The eigenvalues are both positive as T is identified with a positive-definite matrix.

4.3.2 The Tensor Method

Tensors summarising directional information are commonly used in diffusion tensor imaging, where a tensor is calculated at each voxel (3-dimensional pixel) by estimating the diffusion of water molecules in three directions. From this information the dominant orientation of the flow of water molecules can be estimated.

To construct a tensor for a point pattern we follow the tensor method described in Su et al. [2008]. At each point a tensor is created which summarises the local orientation of the cluster in which it lies. This orientation is estimated by considering the distance and direction of other points.

Let $y_1, ..., y_m$ denote the spatial data points. A tensor is constructed at a point y_j using a non-linear transformation applied to the vectors $v^i = (v_1^i, v_2^i) = \overrightarrow{y_j y_i}$ for $i \neq j$ (Su et al., 2008; Su, 2009). Specifically,

$$\tilde{v}^i = (\tilde{v}_1^i, \tilde{v}_2^i) \quad = \quad \frac{\exp\left(-\left((v_1^i)^2 + (v_2^i)^2\right)/4\sigma_{\mathrm{FO}}^2\right)}{\sqrt{(v_1^i)^2 + (v_2^i)^2}}(v_1^i, v_2^i) \tag{4.4}$$

where σ_{FO} is a scaling parameter.

The *initial tensor* at y_j is then represented by

$$T_0(y_j) = \sum_{i \neq j} (\tilde{v}_1^i, \tilde{v}_2^i)^{\mathrm{T}} (\tilde{v}_1^i, \tilde{v}_2^i). \tag{4.5}$$

Note that the parameter σ_{FO} in Equation (4.4) has been rescaled, as Su et al. [2008] and Su [2009] use transformed vectors

$$\tilde{v}^i = (\tilde{v}_1^i, \tilde{v}_2^i) \quad = \quad \frac{\exp\left(-\left((v_1^i)^2 + (v_2^i)^2\right)/2\sigma_{\mathrm{FO}}^2\right)}{\sqrt{(v_1^i)^2 + (v_2^i)^2}}(v_1^i, v_2^i). \tag{4.6}$$

The factor is changed so that the contribution of each unit vector

$$\frac{(v_1^i, v_2^i)}{\sqrt{(v_1^i)^2 + (v_2^i)^2}} \tag{4.7}$$

to the tensor is weighted by a Gaussian function with variance σ_{FO}^2.

Also, Su et al. [2008] and Su [2009] append the collection of transformed vectors $\tilde{v}^1, ..., \tilde{v}^m$ with the same collection rotated by π, i.e. $-\tilde{v}^1, ..., -\tilde{v}^m$. While this is theoretically appealing as it creates a mean-centred collection of vectors, corresponding to the mean-centred data in a principal components analysis, it has no effect on the orientation or anisotropy of the tensor, merely doubling every component, for that reason we choose to omit this step.

The tensor calculated in Equation (4.5) will have a zero-eigenvalue if the points are collinear (lie on a single straight line), and therefore not be positive definite. If all points are truly collinear then our approach breaks down as it is not intended for such noise-free data sets. The more common situation is that one vector $\tilde{v}^i$ dominates the

Figure 4.1: Fingerprint a002-05 from the NIST database, (Watson, 2001).

tensor as calculated in Equation (4.5) due to the relative distances between points. Typically this occurs if two points are close while other points are far from the pair. Due to rounding errors, the contribution of other points to the matrix becomes zero, and the two remaining points are collinear by definition. In order to avoid errors in further calculations we set tensors with at least one zero-eigenvalue to the identity matrix, suggesting a lack of directional information.

We leave an analysis of the robustness of this orientation estimate to Chapter 7, where we also propose a new measure of anisotropy suited to the tensor method construction.

4.3.3 Example of Tensor Calculation

Before continuing, we provide an example of the tensor calculation by applying it to the fingerprint pore data. We use fingerprint a002-05 from the *NIST (National Institute of Standards and Technology) Special Database 30* (Watson, 2001). Pores were extracted from the fingerprint image following the procedure described in Su et al. [2008]. The full fingerprint is shown in Figure 4.1.

Figure 4.2 shows the extracted pore locations, a pictographic summary of the cal-

culation is provided in Figure 4.3, and Figure 4.4 shows the principal eigenvector orientations of the initial tensors. In Figure 4.3 we use the popular approach of visualising tensors as ellipses. The lengths of the axes of the ellipse are proportional to eigenvalues of the tensor and the major axis is parallel to the major eigenvector.

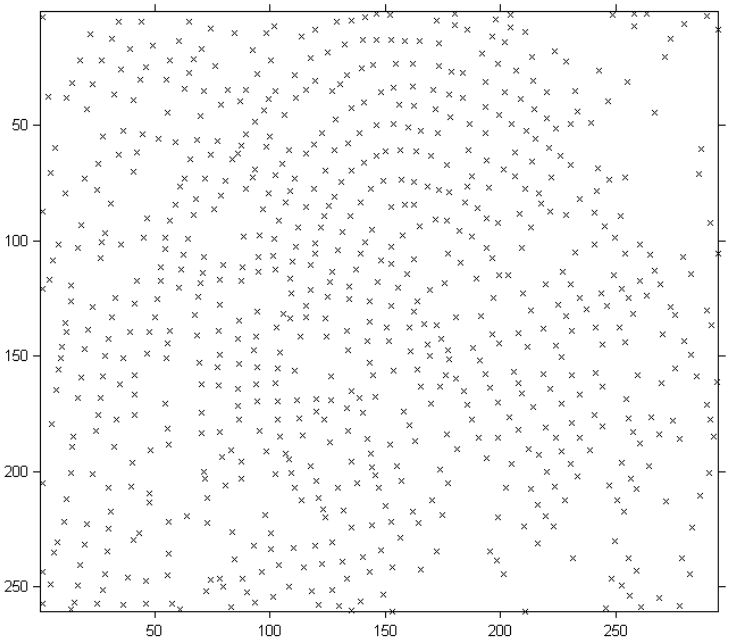


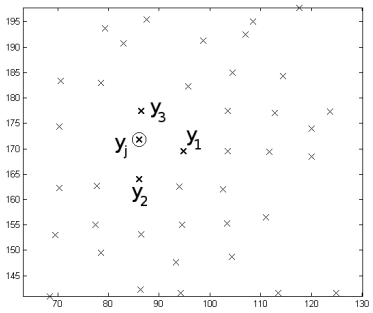
Figure 4.2: Pore data (represented by $\times$) extracted from fingerprint a002-05 from the NIST database (Watson, 2001).

Figure 4.4 shows the orientations of the principal eigenvectors of the tensors created in Section 4.3.2.

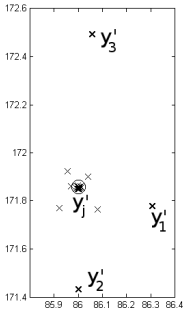
Like the sample covariance matrix, the principal eigenvector of the tensor indicates the axis along which the variance of the data is maximised (compare with Principal Components Analysis). This implies that the end points of vectors v^i (see Equation (4.4)) are more dispersed along this axis than any other. Hence if the un-transformed pores y_i were projected onto this axis, they would lie relatively close to the initial pore y_j. This illustrates the merit of taking the eigenvector as an estimator for the local orientation at y_i.

4.3.4 Interpolation

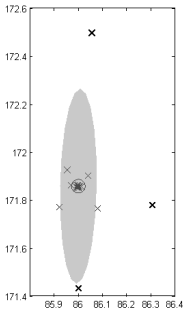
The tensor method allows us to construct an estimate, represented by a tensor, at each point y_i of the orientation of the cluster in which it lies. By the interpolation of



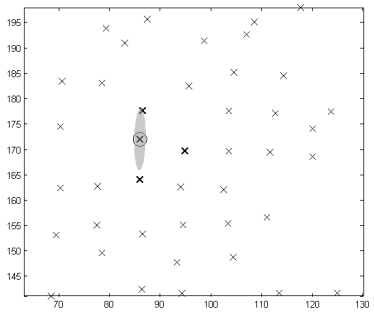
(a) Point pattern with point y_j and 3 other points y_1, y_2, y_3 labelled.



(b) End points of transformed vectors $\tilde{v}^i$, with start points at y'_j.



(c) The tensor at P is shown in ellipse form.



(d) The tensor in relation to the initial pore pattern.

Figure 4.3: Four stages of the tensor method: From the initial point data in (a), vectors $y_i - y_j$ are transformed, the transformed end points are shown in (b). Notice that all but the three nearest points have been transformed very close to y_j and therefore have little effect on the tensor calculated at y_j, represented by an ellipse with the transformed end points in (c) and the original data in (d). The major axis of the ellipse is oriented in the principal eigenvector direction. The ratio of the lengths of the ellipse's axes corresponds to the ratio of the eigenvalues. Crosses indicate end points of transformed vectors $\tilde{v}^i$, with start points at y'_j.

these tensors, a field of orientations is created providing a field of local orientation estimates. An alternative way to estimate the field of orientations is to interpolate the principal eigenvectors of the initial tensors, for example using barycentric, bilinear or bicubic interpolation. However, this approach takes no account of the anisotropy measure - of how dominant the orientation estimate is in each tensor. We desire a smooth prior estimate of the field of orientations to ensure that integral curves are also smooth, and direct interpolation of the principal eigenvectors does not generally have this property. Therefore it is beneficial to impose a smoothness constraint on the field of tensors. We propose a kernel smoothing approach, where

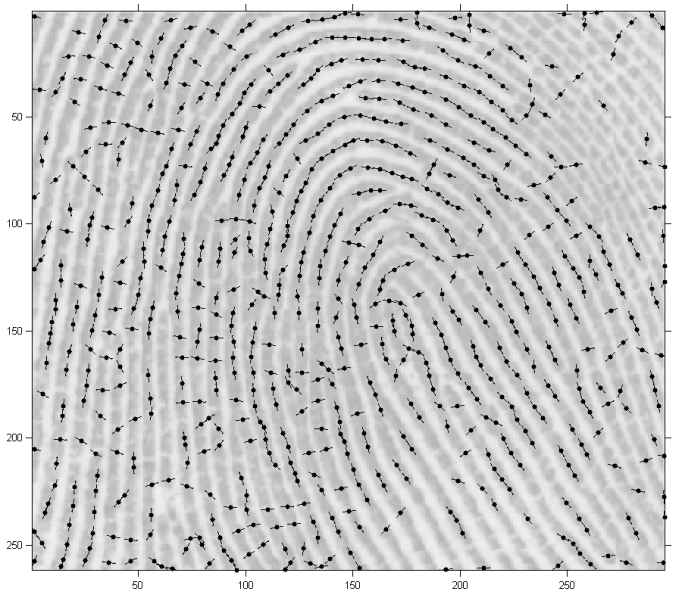


Figure 4.4: Principal eigenvectors of the tensors created by the tensor method (with $\sigma_{\text{FO}} = 3$). Lines through each data point indicate the orientation of the principal eigenvector.

weighted average tensors are calculated which vary smoothly across W.

4.3.5 Tensor Metrics

The tensor metric defines the distance between two tensors and is necessary to calculate any form of *mean tensor*, used for interpolation. Choosing an appropriate tensor metric is a problem that has been encountered in diffusion tensor imaging (Fletcher and Joshi, 2007). For this reason there has been increased interest in developing tensor metrics in recent years. We elect to work in the log-Euclidean metric (see Arsigny et al., 2006) as it satisfies a number of invariance properties and is convenient to work with, providing quick calculations.

We briefly discuss a few other metrics which have been proposed (primarily for use in DTI applications), for a more extensive account of tensor metrics see Dryden et al. [2009].

Euclidean Metric

The standard metric used when working with general matrices is the *Euclidean*

metric where the distance between two matrices M_1, M_2 is

$$\text{dist}_{\text{E}}(M_1, M_2) = \sqrt{\text{tr}\left((M_1 - M_2)^2\right)}, \tag{4.8}$$

effectively each element of the matrix is operated on independently. This is not an appropriate metric for tensor calculations as Euclidean calculations can have strange effects on the properties of eigenvalues. For example, Euclidean extrapolation of two positive definite matrices can create matrices with negative eigenvalues. Also, Euclidean interpolation can lead to a 'ballooning' effect of the determinant: instead of increasing (or decreasing) as one interpolates between two tensors it may reach a maximum for some intermediate tensor. See Pennec et al. [2006, pg 16, Figure 3] for a visual depiction of this effect.

Affine-Invariant Riemannian Metric

An alternative to the Euclidean metric is the affine-invariant Riemannian metric proposed in Pennec et al. [2006]. The strong theoretical properties of this metric make it ideal for working with tensors, however multilinear interpolation, such as weighted averages, can be computationally expensive. In general, the Fréchet mean (which minimises the least square distances) does not have an explicit solution, hence these weighted averages are approximated through a Newton gradient descent method.

Log-Euclidean Metric

Arsigny et al. [2006] propose an alternative metric without the complexity of interpolation associated with the affine-invariant Riemannian metric, named the *log-Euclidean metric*. Log-Euclidean calculations are simply Euclidean calculations on the tensor logarithms which are transformed back to tensor space by taking the exponential. The tensors arising in this study can all be represented by positive definite matrices. Tensor logarithms are therefore well defined as logarithms of these matrices.

The distance between two tensors T_1, T_2 in the log-Euclidean metric is defined by

$$\text{dist}_{\text{LE}}(T_1, T_2) = \sqrt{\text{tr}\left((\log(M_1) - \log(M_2))^2\right)}. \tag{4.9}$$

The log-Euclidean Fréchet mean of m tensors $T_0(y_1), T_0(y_2), \ldots, T_0(y_m)$ with associated weights $w_1, w_2 \ldots, w_m$, i.e. the tensor T that minimises

$$\sum_{i=1}^{m} w_i \text{dist}_{\text{LE}}(T, T_0(y_i))^2 \tag{4.10}$$

where dist_{LE} is the log-Euclidean distance, is given by

$$T = \exp\left\{ \frac{\sum_{i=1}^{m} w_i \log(T_0(y_i))}{\sum_{i=1}^{m} w_i} \right\}. \tag{4.11}$$

The log-Euclidean metric is not fully affine-invariant but it is similarity invariant, and is therefore invariant to orthogonal transformations and scaling, making it acceptable as a tensor metric. The actual results of log-Euclidean interpolation are a significant improvement on the undesirable effects of Euclidean interpolation, and very similar to the results of interpolation in the affine-invariant Riemannian metric - the only difference being that Log-Euclidean means are, in general, slightly more anisotropic. See Arsigny et al. [2006] for further details.

4.3.6 Kernel Smoothing

The field of tensors is calculated as the weighted log-Euclidean average of *all* initial tensors $T_0(y_i)$ at data points y_i. The weights are a function of the Euclidean distance from y_i, i.e. the tensor field evaluated at $x \in W$ is

$$T_{h_{\text{FO}}}(x) = \exp\left(\frac{\sum_i f_{h_{\text{FO}}}(\text{dist}(x, y_i)) \log(T_0(y_i))}{\sum_i f_{h_{\text{FO}}}(\text{dist}(x, y_i))} \right), \tag{4.12}$$

where h_{FO} is a smoothing parameter.

The function $f_{h_{\text{FO}}}$ is chosen to be a multivariate Gaussian function centred at 0,

$$f_{h_{\text{FO}}}(\text{dist}(x, y_i)) = \exp\left(-\frac{\text{dist}(x, y_i)^2}{2h_{\text{FO}}^2} \right). \tag{4.13}$$

We choose a Gaussian kernel as it is infinitely continuous, assigning large weights to points within a certain radius of x and thereafter weights decrease very quickly, however, as discussed in Wand and Jones [1995], the smoothing is robust to the choice of function. This type of interpolation is based on ideas from kernel smoothing, hence we call it kernel smoothing interpolation.

Kernel smoothing interpolation has the advantages over barycentric interpolation (used in Su, 2009) that, for sufficiently large h_{FO}, it produces a smooth field of orientations with fewer undesirable artefacts.

A field of orientations is estimated by assigning to each point $x \in W$ the orientation of the principal eigenvector of tensor $T_{h_{\text{FO}}}(x)$.

In most instances this procedure will give a good estimation of a suitable field of orientations for modelling the point process with integral fibres. By this we mean

that the field of orientations is integrated by fibres with high likelihoods for the observed point pattern. The smoothing method has the drawback that it can create a bias around areas of high curvature (rapidly varying orientation) in the field of orientations. The magnitude of the bias appears to be proportional to the smoothing parameter h_{FO}, see Section 4.5. Three possible corrections to this curvature bias are described in Section 4.6.

4.4 Construction of the Field of Orientations Estimator

We now summarise how a field of orientations is estimated under the assumption that the data $y_1, ..., y_m$ are all part of the signal point process. The estimation is then extended for more general data to include the probabilities that each point is signal, $\boldsymbol{\epsilon}$.

4.4.1 Estimation for all Signal Points

The result of the tensor method described in Section 4.3.2 is a set of tensors located over a sparse set of locations. This sparse set of tensors is interpolated as described in the previous section and the orientation of the principal eigenvector, where defined in the field of tensors, determines a field of orientations.

We calculate the interpolated tensor field $T_{h_{\mathrm{FO}}}(x)$ for $(x \in W)$ as a kernel smoothing procedure, using a Gaussian kernel with variance parameter h_{FO}^2 in the log-Euclidean metric. Hence for $h_{\mathrm{FO}} > 0$,

$$\begin{aligned} T_{h_{\mathrm{FO}}}(x) &= \exp\left(\frac{\int_W f(\mathrm{dist}(x,z)) \log(T_0(z))\,\mathrm{d}z}{\int_W f(\mathrm{dist}(x,z))\,\mathrm{d}z}\right) \\ &= \exp\left(\frac{\sum_{y_i \in \{y_1,\ldots,y_m\}} f(\mathrm{dist}(x,y_i)) \log(T_0(y_i))}{\sum_{y_i \in \{y_1,\ldots,y_m\}} f(\mathrm{dist}(x,y_i))}\right) \end{aligned} \tag{4.14}$$

as we set $\log(T_0(z)) = \underline{0}_2$ (the zero matrix) for $z \notin \{y_1, ..., y_m\}$.

The field of orientations $v_{FO}(x)$ for $x \in W$ is defined to be $\tan^{-1}(v_1(x)/v_2(x))$ where $(v_1(x), v_2(x))$ is the principal eigenvector of the matrix representation of $T_{h_{\mathrm{FO}}}(x)$.

4.4.2 Estimation using Signal Probabilities

We extend this orientation field estimation to take account of the vector of probabilities $(\varepsilon_1, \varepsilon_2, ..., \varepsilon_m)$ that points are signal by weighting the construction of the initial tensor and also weighting the contribution of each initial tensor to the kernel smoothing.

Specifically, the initial tensors (equivalent to those calculated in Equation (4.5)) are represented by

$$T_0(y_i) = \sum_{j \neq i} \epsilon_j (\tilde{v}_1^j, \tilde{v}_2^j)^{\mathrm{T}} (\tilde{v}_1^j, \tilde{v}_2^j) \tag{4.15}$$

for each point y_i, and the tensor field becomes

$$T_{h_{\mathrm{FO}}}(x) = \exp\left(\frac{\sum_{y_i \in \{y_1,...,y_m\}} f(\mathrm{dist}(x, y_i)) \log(\epsilon_i T_0(y_i))}{\sum_{y_i \in \{y_1,...,y_m\}} f(\mathrm{dist}(x, y_i))} \right). \tag{4.16}$$

The field of orientations $\upsilon_{\mathrm{FO}(\mathbf{y},\boldsymbol{\varepsilon},h_{\mathrm{FO}},\sigma_{\mathrm{FO}})}$ is calculated from the tensor field in Equation (4.16) by taking the orientation of the principal eigenvectors. This weighting allows points that are more likely to arise from the signal component to have a greater effect on the field of orientations estimation. As $\varepsilon_i \to 0$ the effect of the point y_i on the field of orientations tends to zero, whereas if $\varepsilon_i = 1$ for all i we would be performing the calculation described in Section 4.4.1.

In Chapters 5 and 6 the empirical Bayes prior distribution of the field of orientations is:

$$P(\Upsilon_{\mathrm{FO}} = \upsilon_{\mathrm{FO}(\mathbf{y},\boldsymbol{\varepsilon},h_{\mathrm{FO}},\sigma_{\mathrm{FO}})} | \mathbf{y}, \boldsymbol{\varepsilon}, h_{\mathrm{FO}}, \sigma_{\mathrm{FO}}) = 1. \tag{4.17}$$

4.4.3 Example of Tensor Field Estimation

The above method is applied to the fingerprint pore data shown in Figure 4.2. Generally the tensor method performs very well in identifying local orientations, however there is still a substantial amount of noise, justifying a combined smoothing and interpolation step.

Figure 4.5 shows the resultant field of orientations corresponding to the tensor field estimated by Equation (4.16). It is evident from this example that the kernel smoothing approach is effective in interpolating over areas with missing pores. The smoothing effect is particularly beneficial in this application as the ridgelines of fingerprints are locally parallel, so we expect the field of orientations to be smooth (have a low variation over a local neighbourhood).

The interpolated field of orientations fails to follow the underlying ridges around

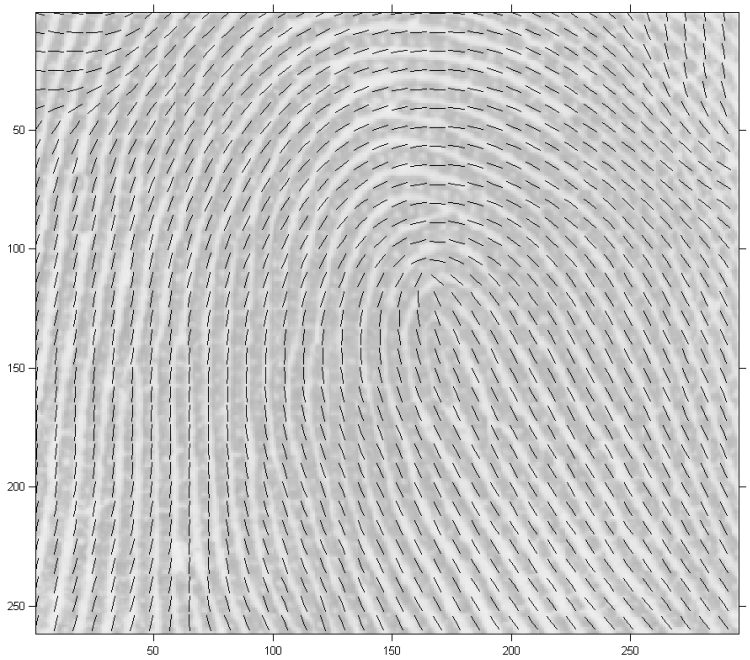


Figure 4.5: Principal eigenvector field of the tensor field empirically estimated from the pore data extracted from fingerprint a005-05 (from the NIST database, Watson, 2001) calculated with $h_{\mathrm{FO}} = 30$, $\sigma_{\mathrm{FO}} = 3$, $\epsilon_i = 1$ for all i. Lines indicate the orientations of principal eigenvectors over a regular grid.

the edges of the window where pores are very sparse and also in the central region near the loop of the fingerprint. The central loop feature has been pushed upwards in the field of orientations due to the effect of smoothing on areas of high curvature. This bias is directly related to the smoothing parameter h_{FO} as described in Section 4.5; three possible corrections are outlined in Section 4.6.

In the following sections we focus on fingerprint pore data. The bias is most evident in the pore data because the fibres (ridge-lines) are long and lie approximately parallel to one another, with points located almost centrally and evenly spaced apart. In comparison, the bias appears to have only a small effect on data where clusters are thicker (points are more dispersed), and the distance between clusters is greater.

For the remainder of this chapter we write the smoothing parameter h_{FO} as h for clarity.

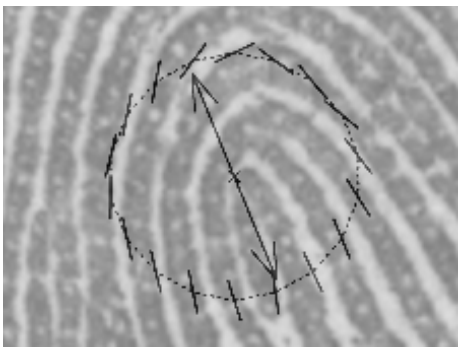


Figure 4.6: The approximate orientations of the underlying fingerprint ridge-lines, around a circle centred at the loop of fingerprint a002-05 (from the NIST database, Watson, 2001). A large proportion of the orientations appear to be near-vertical. The large central arrows indicate the average of these orientations.

4.5 Curvature Bias

At the centre of Figure 4.5 there is evidence that smoothing can affect the field of orientations around an area of high curvature. The term *high curvature* is used to describe an orientation field integrated by curves that have high curvature. We assume that the initial tensors calculated at each point give a reasonable indication of the cluster orientation at those points.

If a set of tensors have principal eigenvectors that are mostly oriented in the same direction then the principal eigenvector of the log-Euclidean mean of the set will be similarly oriented. This is conditional on the set of tensors having similar eigenvalues. Figure 4.6 shows the approximate orientations of the ridge-lines around a circle centred at the loop of the fingerprint. A large proportion of the orientations along the circle are near-vertical (the average orientation is indicated by the large arrows in the centre). If the initial tensors are good estimators of the underlying ridge orientation it is expected that the principal eigenvector of the interpolated tensor at the centre of the loop has a near-vertical orientation too. See Figure 4.5 where the orientation field at the central loop of the fingerprint follows this average orientation.

In the following section the effect of this bias is estimated by considering the location of the singularity (or loop) of the interpolated field compared with its true location in the fingerprint.

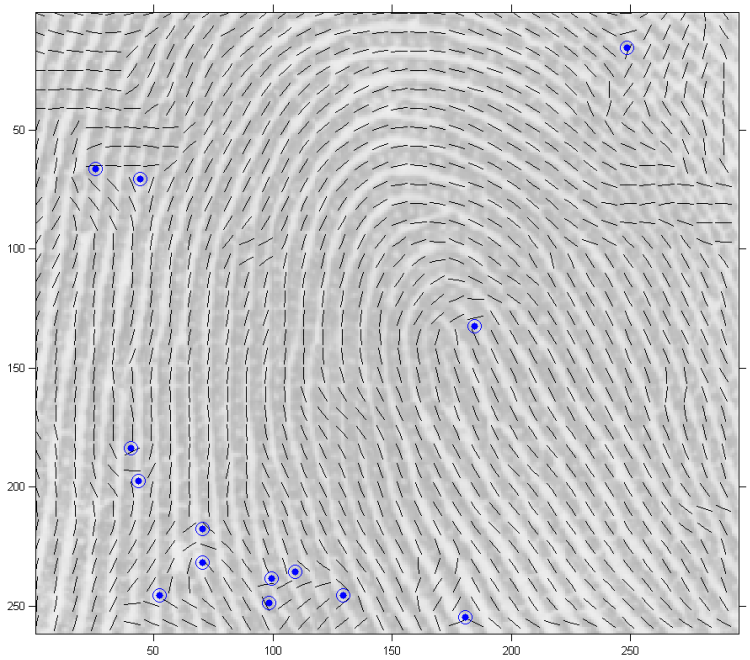


Figure 4.7: Principal eigenvector field of the tensor field empirically estimated from the extracted pore data from fingerprint a005-05 (from the NIST database Watson, 2001) calculated with $h = 10$. Lines indicate the orientations of principal eigenvectors over a regular grid. Circled points denote singularities in the tensor field.

4.5.1 Singularities in a Tensor Field

Singularities in a tensor field are locations where the tensor has equal eigenvalues. Equivalently, singularities in a field of orientations are the points where the orientation is undefined.

If the smoothing parameter h is sufficiently small, many singularities appear due to the variation in initial tensor estimates of the local orientations. Compare Figure 4.7 and Figure 4.8, the fields of orientations from tensor fields with smoothing parameters 10 and 60 respectively. The first has many singularities whereas the larger smoothing parameter of Figure 4.8 has 'smoothed out' the anomalous singularities, leaving the key singularity which defines the overall shape of the field of orientations.

Assume for now that there exists an underlying field of orientations that follows the true ridge-line orientation and which we are trying to estimate. The underlying orientation field may have singularities; in the case of the fingerprint in Figure 4.1, there appears to be one central singularity at the loop of the fingerprint. Applying a

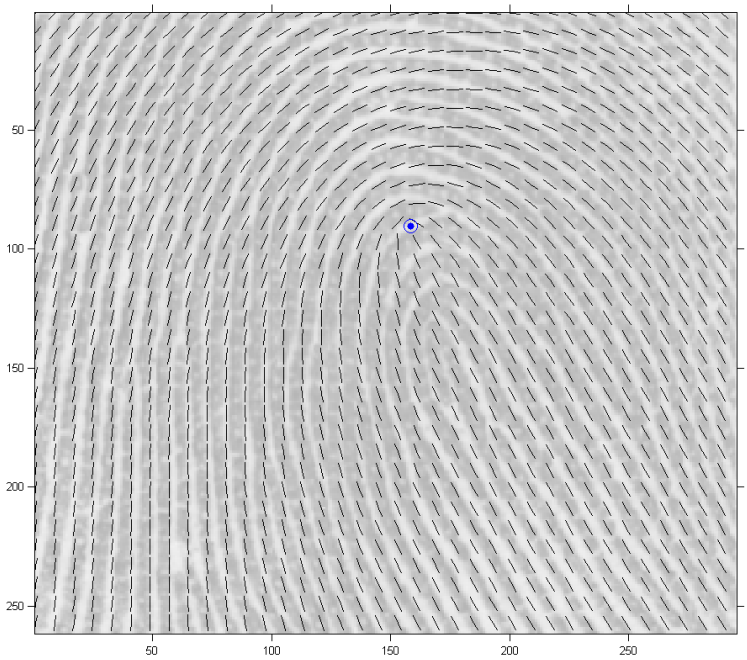


Figure 4.8: Principal eigenvector field of the tensor field empirically estimated from the extracted pore data from fingerprint a005-05 (from the NIST database Watson, 2001) calculated with $h = 60$. Lines indicate the orientations of principal eigenvectors over a regular grid. The circled point denotes a singularity in the tensor field.

kernel smoothing to the initial tensors causes a bias on the location of the singularities. This bias is evident in Figure 4.8 where the central singularity is significantly displaced.

Through comparison with the application of kernel smoothing to a tensor field over the window W, we estimate the extent of the bias arising from applying kernel smoothing to a set of tensors sparsely located across the window W.

We begin with an arch-model tensor field based on fingerprint a002-05 (Watson, 2001) in Figure 4.1, and show in the subsequent section how a similar model, the parabola, gives very different results.

Estimation of Singularity Location Bias: Basic Arch Model

Theorem 1 (Singularity Location Bias: Arch Model)**.** *Let* $T_0^{arch} : \mathbb{R}^2 \to [0, \pi)$ *be a tensor field with constant eigenvalues* $\lambda_1 > \lambda_2 > 0$*, and principal eigenvectors that agree with the arch model pictured in Figure 4.9. Explicitly, using a Cartesian*

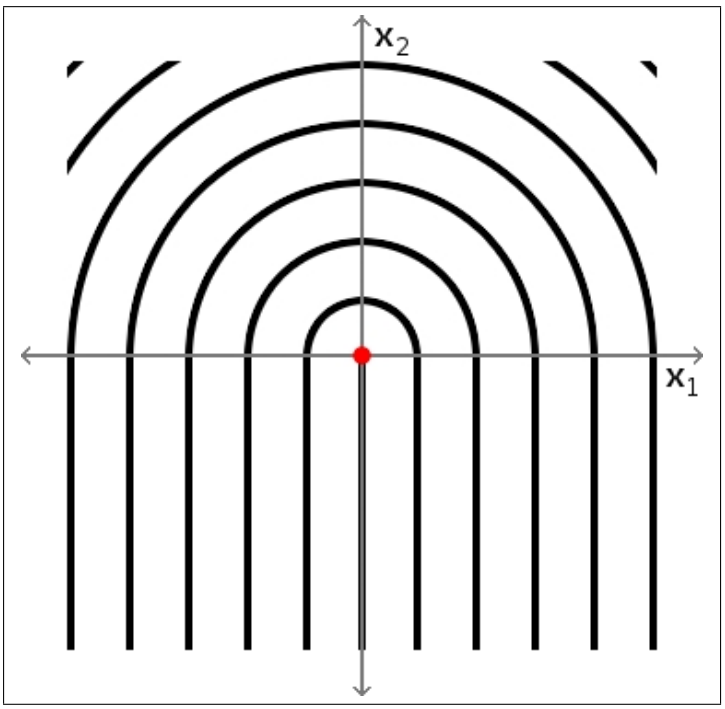


Figure 4.9: A basic fingerprint structure of concentric arches (not to be confused with the arch fingerprint pattern). The singularity is located at the origin with concentric circles above the horizontal axis and parallel vertical lines below.

coordinate system $(x_1, x_2) \in \mathbb{R}^2$ *with origin* $(0,0)$ *at the singularity, eigenvectors are tangent to a circle centred at* $(0,0)$ *if* $x_2 > 0$ *and equal to* $(0,1)$ *if* $x_2 \leq 0$. *Denote by* T_h^{arch} *the result of applying a kernel smoothing in the log-Euclidean metric to* T_0^{arch}, *using a Gaussian kernel with parameter* h. *Then* T_h^{arch} *has a singularity at* $(x_1, x_2) = (0, hc)$. *The constant* c *is the solution of*

$$\begin{aligned}
&\int_0^c \left(\int_{\cos^{-1}\left(\frac{c}{r'_x}\right)-\frac{\pi}{2}}^{\frac{3\pi}{2}-\cos^{-1}\left(\frac{c}{r'_x}\right)} \cos^2\theta_0 - \sin^2\theta_0 \; d\theta_x \right. \\
&\left. + 2\cos^{-1}\left(\frac{c}{r'_x}\right) \right) \frac{\exp(-(r'_x)^2/2)}{2\pi} dr'_x \\
&+ \int_c^\infty \int_0^{2\pi} \cos^2\theta_0 - \sin^2\theta_0 \; d\theta_x \frac{\exp(-(r'_x)^2/2)}{2\pi} dr'_x = 0
\end{aligned} \tag{4.18}$$

where

$$\theta_0 = \theta_0(r_x, \theta_x) = \tan^{-1}\left(\tan\theta_x + \frac{\beta}{r\cos\theta_x}\right). \tag{4.19}$$

The proof of this theorem is given in Appendix B.1.

The equality in Equation (4.18) must be solved computationally. An estimation

based on Riemann integrals gives $c \approx 0.772$ to 3 decimal places. This implies that the result of smoothing the tensor field is that the singularity is displaced vertically upwards to a distance of $0.772 \times h$.

Theorem 1 describes the extent of the singularity location bias arising from the application of kernel smoothing to a tensor field. In our tensor field construction, the same kernel smoothing is applied but to a discrete set of tensors sparsely located over W, rather than an infinite tensor field. Approximating the sparse set of tensors by an infinite tensor field is necessary to simplify calculations. Noise, or variation from the arch model orientation, in the point estimates of the tensor field will likely affect the extent of the bias, but with no further information on the nature of the noise it is assumed that the infinite tensor field provides a reasonable approximation to the discrete collection of tensors. The effect of noise on the initial tensors is discussed in Section 7.3, in order not to distract from the current discussion.

Approximating the tensor calculation by integrating over the whole of $\mathbb{R}^2$, rather than just W, will only affect the estimation of the extent of the bias if the distance from the singularity to the boundary of W is small in comparison to h. Heuristic evidence shows that singularities are 'pulled' to the boundary of W as the smoothing parameter is increased beyond a certain threshold. This is believed to be a result of edge effects in kernel smoothing.

The assumption that eigenvalues are constant and therefore that tensors are all equal except for orientation is unlikely to be true in practice as variations in the point pattern intensity and local anisotropy directly affect the eigenvalues. However, it is a reasonable approximation, particularly as the variation in eigenvalues across the set of tensors is not easily predictable.

Applying the result of Theorem 1 to the example of Figure 4.5, we find that increasing h from 30 to 60 displaces the singularity by a further 36 length units which is of the same order as the estimate $0.772 \times 30 = 23.2$. The difference is likely due to the inaccurate assumptions, in particular the fingerprint seems to have higher curvature than the arch model. Increasing h to 90 displaces the singularity by a further 41 length units in approximately the same direction. This appears to support the claim that the displacement is approximately linear in h.

Estimation of Singularity Location Bias: Parabolic Model

In the previous section it was shown that smoothing an arch-shaped field of tensors will displace the singularity by a distance proportional to the smoothing parameter h. We now consider a second model based on parabolas which, although it

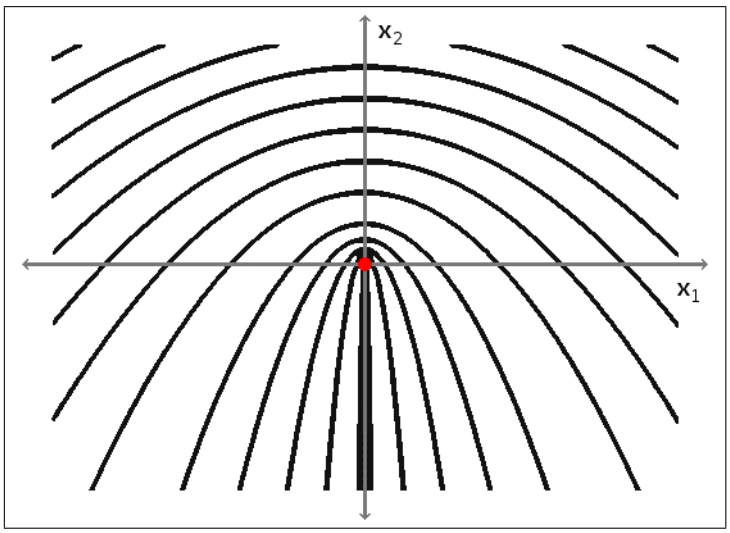


Figure 4.10: The parabolic tensor field, illustrated by integral fibres of the associated field of orientations.

appears similar in shape to the arch model gives quite different results. For the parabolic model, the smoothing causes no bias on the location of the singularity of the parabolic tensor field.

Let (x_1, x_2) denote a Cartestian coordinate system over $\mathbb{R}^2$. The parabolic tensor field $T_0^{\text{para}}(x_1, x_2)$ is defined at (x_1, x_2) to be the tensor with constant eigenvalues $\lambda_1 > \lambda_2 > 0$ and principal eigenvector tangent to the parabola $x_2 = \frac{1}{4a} - ax_1^2$ for $x_1 \neq 0$, where $a > 0$ takes different values depending on x_1, x_2. For $x_1 = 0$, the principal eigenvector is proportional to $(1, 0)$ if $x_2 > 0$ and $(0, 1)$ if $x_2 < 0$. The corresponding orientation field of the parabolic tensor field is continuous over $\mathbb{R}^2 \backslash \{(0, 0)\}$ (verified by considering the derivatives $\frac{dx_1}{dx_2}$ and $\frac{dx_2}{dx_1}$) and has a single singularity located at the origin $(0, 0)$. Any integral curve of the associated field of orientations is a parabola with a focus at $(0, 0)$ and directrix given by $x_2 = c$ for some $c > 0$, hence the term, parabolic tensor field. These parabolic integral curves are illustrated in Figure 4.10.

Theorem 2 (Singularity Location Bias: Parabolic Model)**.** *Let $T_0^{para}(x_1, x_2)$ be a parabolic tensor field with constant eigenvalues $\lambda_1 > \lambda_2$, and let $T_h^{para}(x_1, x_2)$ be the result of applying a convolution in the log-Euclidean metric to this tensor field with a Gaussian kernel and smoothing parameter h. Then $T_h^{para}(x_1, x_2)$ contains a singularity located at the origin $(0, 0)$.*

The proof of this result is a consequence of the following lemma

Lemma 1. *Let $e(x_1, x_2) = (e_1(x_1, x_2), e_2(x_1, x_2))$ denote the principal eigenvector of the tensor $T_0^{para}(x_1, x_2)$. Then $e(x_1, x_2)$ is perpendicular to $e(-x_1, -x_2)$ for all $(x_1, x_2) \neq (0, 0)$.*

The proofs of both lemma and theorem are given in Appendix B.2.

These two plausible models - the arch and parabola, make it clear that when estimating the bias on the location of the singularity, choosing an appropriate approximation is crucial.

4.5.2 The 3 Stages of Singularity Displacement due to Smoothing

The previous two sections describe the biasing effect of kernel smoothing by considering the displacement of the singularity. However, when applying kernel smoothing to a discrete set of tensors located sparsely over W the bias resulting from the curvature of the tensor field only accounts for one of three stages of singularity displacement, that occur with increasing h.

For small h, the tensor interpolation smooths out the noise of the initial tensors. During this stage the singularity will be moved in a random manner which is not easily predictable.

Once h is larger than some first threshold t_{h_0}, the Gaussian kernel will assign sufficient weights to enough initial tensors that the tensor field will become locally stable. At this stage the bias outlined in the previous section is apparent; areas of high curvature induce a bias in the kernel smoothed tensor field.

When h is higher than a second threshold t_{h_1} this bias is dominated by another effect. As h increases pairs of singularities attract one other and eventually cancel each other out. Pairs of singularities can only cancel if their Poincaré indices sum to 0. The Poincaré index of a singularity (described in Maltoni et al., 2003) is calculated by choosing a small closed curve C around the singularity and moving once clockwise around the closed curve. The Poincaré index is the sum of the gradient differences (in degrees) between consecutive points on the curve. Poincaré indices take one of 5 values: $\pm\pi$ radians, $\pm 2\pi$ radians or 0 radians if there is no singularity within the curve. Two singularities with Poincaré indices that sum to 0 may cancel each other out as h passes some threshold. As h increases the two singularities move closer until they meet and both disappear from the tensor field. It is not clear which singularities will cancel but this opens up an interesting area of research on the topology of kernel-smoothed tensor fields. If there is no complementary singularity to cancel with, a singularity will usually move towards the edges of the window. It is not entirely clear why this happens but it is believed to be due to the fact that as $h \to \infty$ the weights assigned to each tensor in the kernel smoothing approach equality. Hence the tensor field approaches uniformity, and as the location of singularities is continuous with

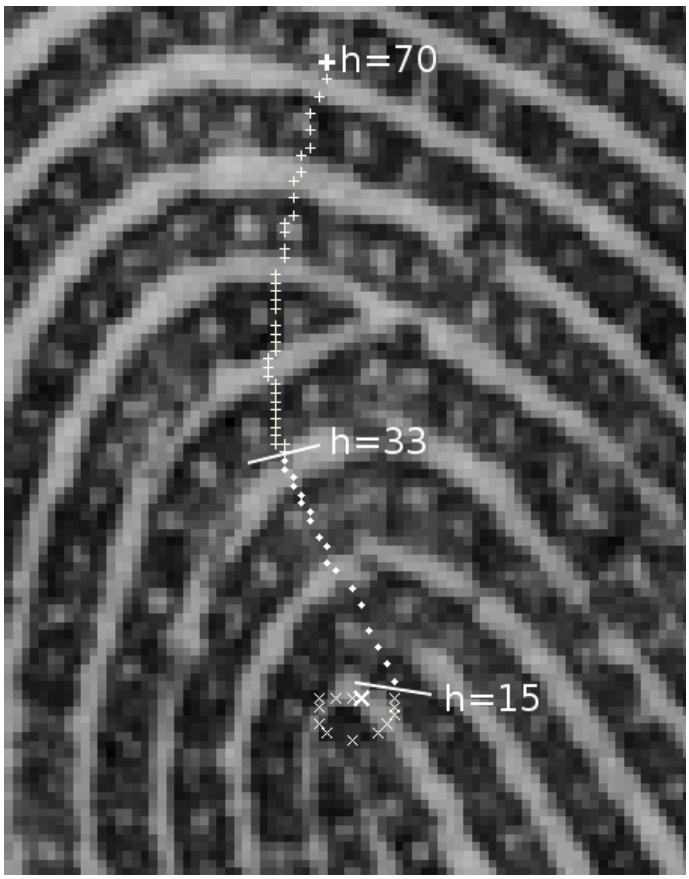


Figure 4.11: The main singularity of the interpolated tensor field for $h = 1$ ($\times$), $2, \ldots, 69, 70$ ($+$).
1st Stage ($\times$): singularities are displaced in a random manner due to the noise of the initial tensors.
2nd Stage ($\cdot$): Singularities are displaced approximately linearly in h due to the high curvature bias of subsection 4.5.1.
3rd Stage ($+$): A third effect dominates the high curvature bias causing the singularity to be displaced towards the edge of the image.

respect to h, the singularity will usually leave W.

Figure 4.11 gives an example of these 3 stages, showing the path of the central singularity as h increases. It appears that the threshold values for these data are $t_{h_0} = 15$ and $t_{h_1} = 33$, although there seems to be an overlap between the stages.

4.6 Bias Correction

In this section, 3 separate approaches to correcting the curvature bias, based on the dependency of the bias on h are proposed. The first is based on a Taylor series

expansion, the second follows from describing the bias correction as an extrapolation problem, and finally we propose a technique using adaptive smoothing - varying the smoothing parameter h across the window W.

4.6.1 Taylor Series Expansion of $\log(T_h(x))$

As mentioned in Section 4.5, the extent of the bias on the tensor field appears to increase as h increases, therefore we propose correcting the bias by estimating the tensor field with $h = 0$, namely $T_0(x)$. One way to estimate $T_0(x)$ is to use a Taylor series expansion.

Rather than calculate the Taylor series expansion that approximates $T_0(x)$, we work with $\log(T_0(x))$ which can be written as a weighted sum of tensors. The partial derivative $\frac{\partial}{\partial h} \log T_h(x)$ empirically estimated from data $y_1, ..., y_m$ is

$$\frac{\partial}{\partial h} \log T_h(x) = \frac{\partial}{\partial h} \left(\frac{\sum_i f_h(\text{dist}(x, y_i)) \log T_0(y_i)}{\sum_i f_h(\text{dist}(x, y_i))} \right). \tag{4.20}$$

By the quotient rule and substituting for $\log T_h(x)$ this evaluates to

$$\begin{aligned} &\frac{\sum_i \frac{\partial}{\partial h} f_h(\text{dist}(x, y_i)) \log T_0(y_i)}{\sum_i f_h(\text{dist}(x, y_i))} - \frac{\sum_i \frac{\partial}{\partial h} f_h(\text{dist}(x, y_i))}{\sum_i f_h(\text{dist}(x, y_i))} \log T_h(x) \\ &= \frac{\sum_i \text{dist}(x, y_i)^2 f_h(\text{dist}(x, y_i))(\log T_0(y_i) - \log T_h(x))}{\sum_i f_h(\text{dist}(x, y_i)) h^3} \end{aligned} \tag{4.21}$$

Recall that $f_h(\text{dist}(x, y_i))$ is a Gaussian function without normalising constant, and $T_0(y_i)$ is the initial tensor evaluated at data point y_i defined in Equation (4.5).

The theoretically unbiased tensor field $T_0(x)$ for $x \in W$ is estimated using a first order Taylor series approximation,

$$\begin{aligned} \widehat{\log T_0(x)} &= \log T_h(x) - \frac{\sum_i \text{dist}(x, y_i)^2 f_h(\text{dist}(x, y_i))(\log T_0(y_i) - \log T_h(x))}{\sum_i f_h(\text{dist}(x, y_i)) h^2} \\ &= \frac{\sum_i f_h(\text{dist}(x, y_i)) \left(\left(1 + \frac{\text{dist}(x, y_i)^2}{h^2} \right) \log T_0(y_i) - \frac{\text{dist}(x, y_i)^2}{h^2} \log T_h(x) \right)}{\sum_i f_h(\text{dist}(x, y_i))}. \end{aligned} \tag{4.22}$$

Effectively, the initial tensor $T_0(y_i)$ in Equation (4.14) has been replaced by the extrapolated tensor $\exp\left((1 + t) \log T_0(y_i) - t \log T_h(x)\right)$ where $t = \frac{\text{dist}(x, y_i)^2}{h^2}$. This approximation is improved by adding more terms to the Taylor series approximation.

Figures 4.12 and 4.13 show the principal eigenvector fields of the tensor fields calculated by the original method on fingerprint a005-05 from the NIST database (Wat-

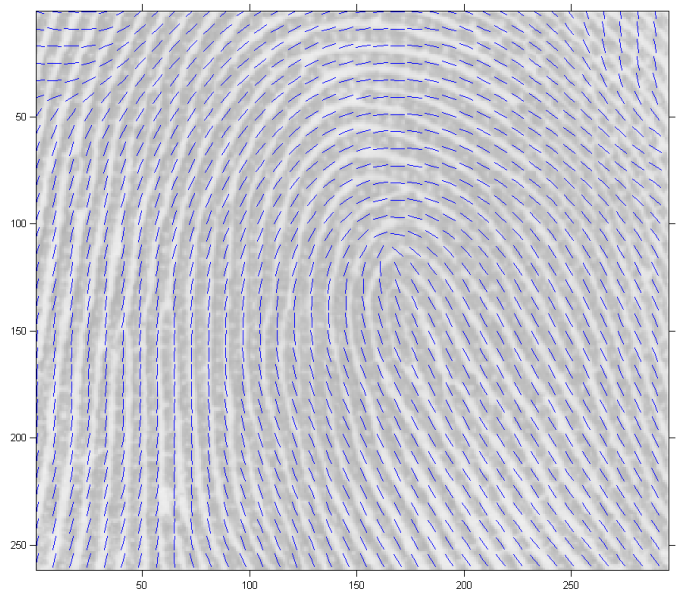


Figure 4.12: The original principal eigenvector field of the tensor field empirically estimated from the extracted pore data from fingerprint a005-05 from the NIST database (Watson, 2001) calculated with $h = 30$. Lines indicate the orientations of principal eigenvectors over a regular grid.

son, 2001) with $h = 30$, and bias corrected with a 2nd order Taylor series expansion. The 1st order Taylor series expansion produces similar results.

The Taylor series approximation appears to reduce the high curvature bias as there is less displacement of the central singularity. Unfortunately artefacts appear around the edges of the window where the point intensity is low and the initial tensor estimates inadequately reflect the ridge-line orientations; these are the areas where smoothing is necessary. These artefacts appear because, by estimating $T_0(x)$ we not only reduce the bias, but also the smoothness of the tensor field.

4.6.2 Extrapolation from Two Instances of the Tensor Field

An alternative approach to estimating the tensor field T_0 is to evaluate the tensor field for two values of h, say $h_2 > h_1$ and extrapolate back to $h = 0$.

Tensors are extrapolated using the log-Euclidean metric. Thus, $\log T_{h_0}(x)$, the logarithm of the tensor field evaluated at $x \in W$ with smoothing parameter $h = h_0$ is estimated by:

$$t(h_0) \log T_{h_1}(x) + (1 - t(h_0)) \log T_{h_2}(x), \tag{4.23}$$

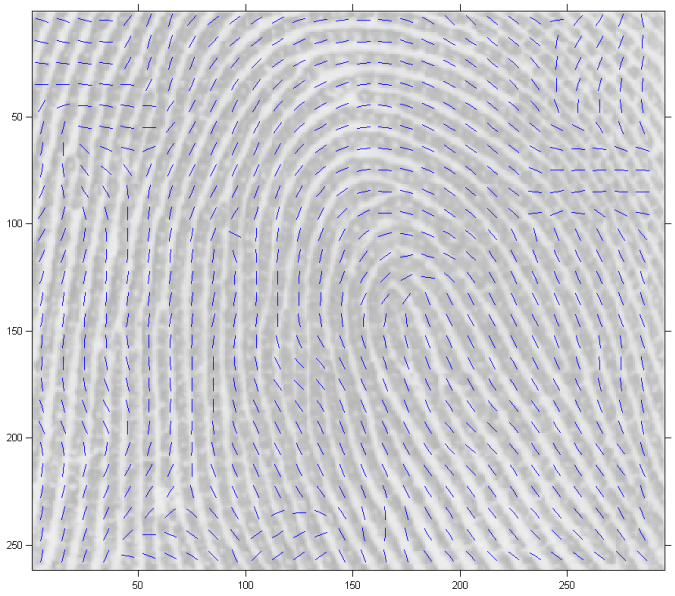


Figure 4.13: The principal eigenvector field of the bias-corrected tensor field calculated by estimating T_0 from T_h with a Taylor series of order 2. Lines indicate the orientations of principal eigenvectors over a regular grid.

where $t(h_0)$ is some function of the target smoothing parameter value, h_0. By rewriting Equation (4.23) in terms of weighted sums of the tensor $T_0(y_i)$ it is possible to estimate the function $t(h_0)$ satisfying

$$\log T_{h_0}(x) = \frac{\sum_i f_{h_0}(\mathrm{dist}(x, y_i)) \log(T_0(y_i))}{\sum_i f_{h_0}(\mathrm{dist}(x, y_i))} \tag{4.24}$$

$$= t(h_0) \frac{\sum_i f_{h_1}(\mathrm{dist}(x, y_i)) \log(T_0(y_i))}{\sum_i f_{h_1}(\mathrm{dist}(x, y_i))} + (1 - t(h_0)) \frac{\sum_i f_{h_2}(\mathrm{dist}(x, y_i)) \log(T_0(y_i))}{\sum_i f_{h_2}(\mathrm{dist}(x, y_i))}. \tag{4.25}$$

The denominator $\sum_i f_{h_j}(\mathrm{dist}(x, y_i))$ (for $j = 0, 1, 2$) can be approximated by the integral

$$\sum_i f_{h_j}(\mathrm{dist}(x, y_i)) \approx \int_{\mathbb{R}^2} \rho \exp\left(-\frac{(x - z)^2}{2h_j^2}\right) \mathrm{d}z$$

$$= 2\pi\rho h_j^2 \tag{4.26}$$

where ρ is an estimate of the density of points y_i.

Substituting the approximation of the denominator into Equation (4.25) gives:

$$\frac{\sum_i f_{h_0}(\mathrm{dist}(x,y_i))\log(T_0(y_i))}{h_0^2} \approx t(h_0)\frac{\sum_i f_{h_1}(\mathrm{dist}(x,y_i))\log(T_0(y_i))}{h_1^2} + (1-t(h_0))\frac{\sum_i f_{h_2}(\mathrm{dist}(x,y_i))\log(T_0(y_i))}{h_2^2}. \quad (4.27)$$

Ideally, we want to find a function $t(h_0)$ which is independent of the point process and the tensor fields. For this reason, we approximate the sum by

$$\sum_i f_{h_j}(\mathrm{dist}(x,y_i))\log(T_0(y_i)) \approx f_{h_j}(\mathrm{dist}(x,y_k))\log(T_k) \quad (4.28)$$

where y_k is the closest data point to x. We also approximate $f_{h_j}(\mathrm{dist}(x,y_i))$ by the upper bound 1. Hence Equation (4.27) becomes

$$\frac{\log(T_k)}{h_0^2} \approx t(h_0)\frac{\log(T_k)}{h_1^2} + (1-t(h_0))\frac{\log(T_k)}{h_2^2}. \quad (4.29)$$

Solving this equation for t_{h_0} suggests the function

$$t(h_0) = \frac{(h_2^2-h_0^2)h_1^2}{(h_2^2-h_1^2)h_0^2} \quad (4.30)$$

is appropriate for the extrapolation parameter. Now, as $h_0 \to 0$, $t(h_0) \to \infty$, so to estimate the unbiased tensor field $T_0(x)$ a large value of t_{h_0} should be chosen.

Examples are shown in Figures 4.14 and 4.15: two tensor fields with different smoothing parameters, shown in Figure 4.8 (h=60) and Figure 4.12 (h=30), are extrapolated with parameters $t = 3$ in Figure 4.14 and $t = 10$ in Figure 4.15. Like the Taylor series expansion in the previous section, the extrapolation appears to reduce the high curvature bias at the expense of introducing artefacts into the tensor field.

4.6.3 Adaptive Smoothing

Both of the previous approaches to bias correction reduce the smoothness of the tensor field. This can lead to the appearance of artefacts such as the singularities around the edges of Figure 4.14 and generally give a poorer estimation of the field of orientations around the edge of the window. This is undesirable as it can completely change the integral curves of the field of orientations, markedly affecting further analysis based on the orientation field estimate.

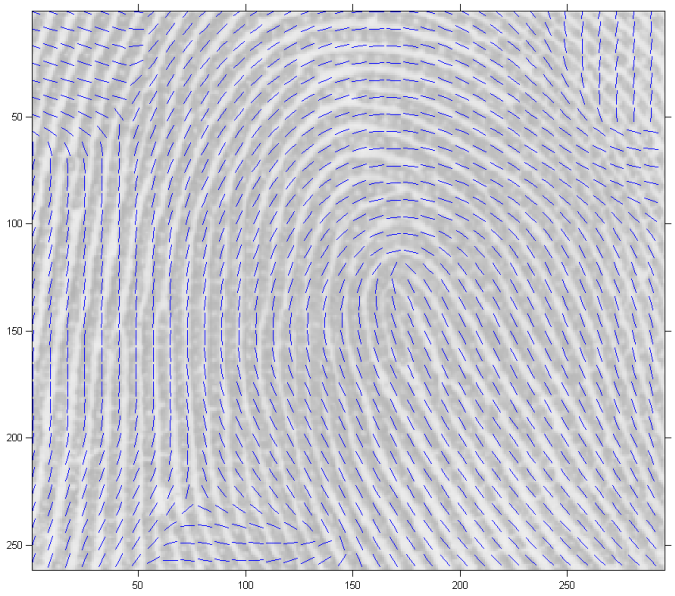


Figure 4.14: The principal eigenvector field of the extrapolated tensor field with parameters $h_1 = 30$, $h_2 = 60$, and $t = 3$. Lines indicate the orientations of principal eigenvectors over a regular grid. There is a noticeable improvement in the tensor field estimation around the central loop feature. However the extrapolation has had an undesirable effect in the top corners.

In this section we propose a method that allows the smoothness of the tensor field to vary across the window. The new approach is based on the intuition that in regions of W with high point density the point estimates of the local anisotropy (through the initial tensor) are more robust, and therefore less smoothing is necessary in these regions. A higher smoothing parameter h is used in areas of low point intensity so that the tensor estimate is based on information (initial tensor estimates) from a greater number of distant points. Conversely, the smoothing bias is reduced in areas of high point intensity by using a low smoothing parameter. This is achieved by replacing the fixed parameter h in Equation (4.13) by a function $h(x)$ where describes the sparsity of points in a neighbourhood of $x \in W$.

For example, in Figure 4.16 the map $h(x)$ is the Euclidean distance from the 10th nearest pore to x.

Choosing an appropriate function $h(x)$ is obviously important and likely to depend on the precise details of the point process. In another application an entirely different map h may be preferred. So despite the fairly good resultant tensor field, this method has the drawback that it requires prior knowledge of the relationship between points and fibres.

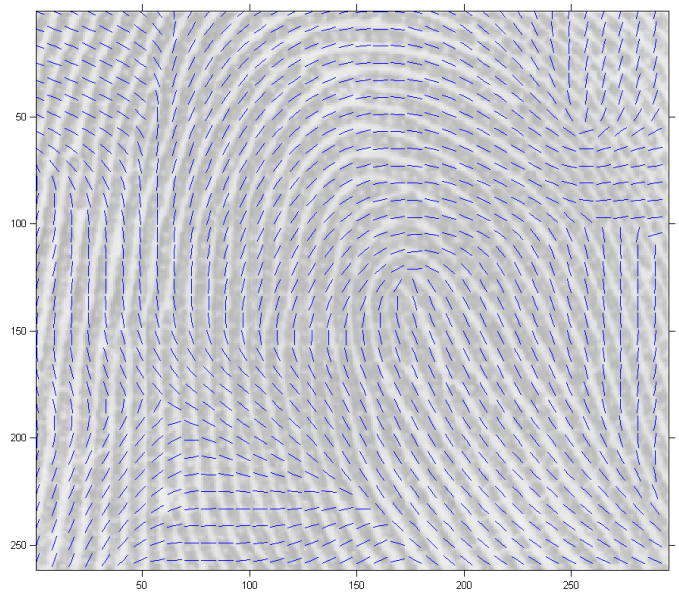


Figure 4.15: The principal eigenvector field of the extrapolated tensor field with parameters $h_1 = 30$, $h_2 = 60$, and $t = 10$. Lines indicate the orientations of principal eigenvectors over a regular grid. The tensor field shows some improvement on the extrapolated field with $t = 3$ around the loop structure, but outside the central region the gradient field has mostly lost its original structure.

While the examples focus on the fingerprint data, these three approaches to bias correction are equally applicable to any other fibre-generated point process exhibiting curvature bias.

4.7 Conclusions

In this chapter we described how an empirical Bayes technique can be used to construct a field of orientations estimator which is integrated by fibres that produce high likelihoods. Tensors (specifically, positive-definite symmetric matrices) are fundamental to the construction as they provide a convenient way to summarise directional data. An initial tensor is constructed at each data point summarising the local orientation of the cluster in which it lies. The tensors are then interpolated with a combined smoothing step, producing a field of tensors that assigns to each point a tensor estimator of the local cluster orientation. By choosing the dominant direction of these tensors - indicated by the principal eigenvector, an estimator for the field of orientations is constructed.

We have noted that the smoothing step introduces a bias into the tensor field.

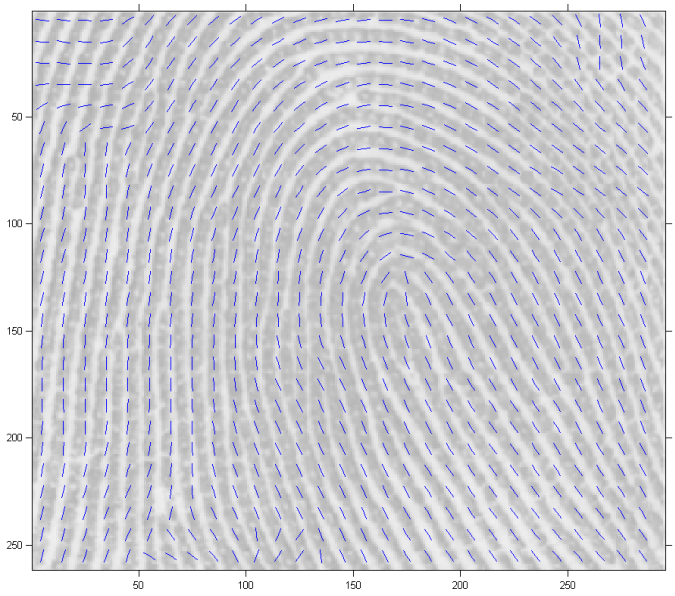


Figure 4.16: Field of orientations estimated using adaptive smoothing and evaluated over a finite grid. Specifically, the field of orientations is identified by the principal eigenvectors of the tensor field estimated using adaptive smoothing. The smoothing function $h(x)$ is equal to the distance from x of the 10th nearest pore. The bias around the central loop feature has significantly decreased and the orientation estimates at the edges of the image (particularly the top corners) benefit from increased smoothing.

The extent of this bias has been analysed by applying the smoothing step to two specific tensor fields. It appears that the bias-induced movement of a singularity is proportional to smoothing parameter h. To correct this bias three approaches have been suggested, although the field of orientations prior used in Chapters 5 and 6 is determined by the original tensor field of Equation (4.16).

There is plenty of scope for extending or adapting this estimator. A few ideas are suggested here.

An alternative transformation could be used in the calculation of initial tensors (see Equation (4.5)), indeed a different construction altogether could be implemented. Stevens [1978] estimates the local orientation of a point pattern by creating a histogram of orientations, rather like the empirical rose of directions (see Stoyan et al., 1995), of vectors connecting all combinations of pairs of points within a neighbourhood. This could provide the starting point for the construction of a smoother collection of sparsely located tensors.

Different kernels may be used for the kernel smoothing although, as already men-

tioned, this is unlikely to have a large effect on the resulting field of orientations.

Rather than find a point estimator for the field of orientations, we could construct a complete prior distribution. It is possible to create a distribution of fields of orientations directly from a tensor field by taking into consideration the anisotropy of the tensors rather than just the eigenvectors. However, sufficient care would be necessary to ensure that the smoothness of the field of orientations is controlled.

These ideas on extending the field of orientations estimate are further discussed in Chapter 8.

We proceed to the next chapter where details are given of a Markov chain Monte Carlo process designed to produce samples from the posterior distribution of fibres.

Chapter 5

Inference via Birth-Death Markov Chain Monte Carlo

In the previous two chapters we introduced a hierarchical Bayes model for point patterns exhibiting a filamentary structure, and gave details of the empirical Bayes estimation of the field of orientations. We would now like to make inferences on aspects of the fibre process given an instance of the point process. Typical attributes of interest include: the number of fibres, where they are located/orientated, which points arose from which fibre and which points arose from background noise.

Here we provide the details of a Markov chain Monte Carlo algorithm that draws samples from the posterior distribution of fibres. From these samples, quantities of interest can be calculated providing the basis of a Bayesian inference on the fibre process.

The mixing of a simple birth-death process of fibres is improved by including additional moves: updating signal probabilities; moving a fibre; adjusting the length of a fibre; updating the allocations of points to the noise and signal components; splitting and joining fibres; and updating the reference point of a fibre. These moves are summarised in Table 5.1. Other considerations relating to the implementation of this algorithm are outlined in Section 5.6.

5.1 Continuous-Time MCMC and Birth-Death MCMC

Direct inference from the model is hindered by the complexity of the hierarchical structure. Hence we choose to draw samples from the posterior distribution of the fibres and other variables using Markov chain Monte Carlo (MCMC) methods. Characteristics of interest can be estimated from these samples.

Table 5.1: Summary of moves in birth-death MCMC

Event	Variables Updated	Acceptance Probability	Frequency of Event (Rate)
birth	$k, \mathbf{F}, \mathbf{l}, \boldsymbol{\omega}, \mathbf{Z}, \mathbf{X}, \mathbf{p}$	1	β
death	$k, \mathbf{F}, \mathbf{l}, \boldsymbol{\omega}, \mathbf{Z}, \mathbf{X}, \mathbf{p}$	1	$\delta_1, ..., \delta_k$ Equation (5.4)
update $\boldsymbol{\varepsilon}$	$k, v_{\mathrm{FO}}, \mathbf{F}, \boldsymbol{\varepsilon}, \mathbf{X}, \mathbf{p}$	Equation (5.29)	$r_{\boldsymbol{\varepsilon}} (\approx 0.1)$
update $\mathbf{l}$	$\mathbf{F}, \mathbf{l}, \mathbf{Z}, \mathbf{X}, \mathbf{p}$	Equation (5.35)	$r_{\mathbf{l}} (\approx 1)$
move fibre	$\mathbf{F}, \boldsymbol{\omega}, \mathbf{X}, \mathbf{p}$	Equation (5.31)	$r_{\boldsymbol{\omega}} (\approx 1)$
update $\mathbf{Z}$	$\mathbf{Z}, \mathbf{X}, \mathbf{p}$	Equation (5.38)	$r_{\mathbf{Z}} (\approx 1)$
split fibre	$k, \mathbf{F}, \mathbf{l}, \boldsymbol{\omega}, \mathbf{X}, \mathbf{p}$	Equation (5.82)	$r_{\mathrm{S}} (\approx 2)$
join fibres	$k, \mathbf{F}, \mathbf{l}, \boldsymbol{\omega}, \mathbf{X}, \mathbf{p}$	Equation (5.83)	$r_{\mathrm{J}} (\approx 2)$
obtain sample	-	-	$r_{\mathrm{o}} (< 0.01)$

A basic knowledge of Markov chain Monte Carlo methodology is assumed, particularly the Metropolis-Hastings algorithm. Extensive background reading on MCMC's can be found in Gilks et al. [1995] and Brooks et al. [2011].

The starting point for our algorithm is a continuous time birth-death Markov chain Monte Carlo (BDMCMC) in which fibres are created and die at random times controlled by predetermined or calculated rates. This enables exploration of a wide range of models with different numbers of fibres; the question of how to compare and make inferences on models with different numbers of variables such as this, is called the variable dimension problem. See Møller and Waagepetersen [2004] (and also Preston, 1977) for an introduction to spatial Birth-Death processes for point processes and Stephens [2000a] for an application of BDMCMC to the mixture model context.

BDMCMC is a specific type of MCMC process where birth and death events occur at random times, with rates chosen to ensure that detailed balance holds. This type of MCMC, where events occur at random times rather than sequentially, is called continuous-time MCMC or CTMCMC - a term introduced by Cappé et al. [2003]. CTMCMC algorithms and examples of their implementation are presented in Huber [2011].

5.2 Details of the Birth-Death Markov Chain Monte Carlo

We choose to fix the birth rate, β, and calculate an appropriate death rate to maintain detailed balance; details of the calculation are given in Section 5.2.2. The

alternative is to propose births at a variable rate, depending on the current state of the chain. In principle, a birth rate may be defined so that the death rate is constant. However, this is difficult to achieve in practice with the complex model described here.

A variable t, initiated at 0, determines the algorithmic time of the BDMCMC; once it has passed a predetermined threshold, the algorithm is halted. At any stage in the BDMCMC, the waiting time until the next birth is exponentially distributed with rate equal to the birth rate, similarly, the waiting time until the death of the j-th fibre is exponentially distributed with rate δ_j. Any proposed move - birth or death, is automatically accepted.

The algorithm for this BDMCMC is as follows. Let $(\mathbf{F}, \boldsymbol{\theta})$ denote the current state of k fibres and associated parameters. The following steps are repeated until time t passes the predetermined threshold.

1. Calculate death rates $\delta_1, ..., \delta_k$ of the respective fibres.
2. Simulate $u_1, u_2 \sim \text{Uniform}(0, 1)$.
3. Set $t = t - \log(u_1) / (\beta + \delta_1 + ... + \delta_k)$ (where β is the birth rate).
4. If $u_2 < \frac{\beta}{\beta+\delta_1+...+\delta_k}$, a birth occurs, otherwise a death occurs.
    - Birth: Draw new fibre and parameters from birth density $b(F, \theta)$ and append F to $\mathbf{F}$, θ to $\boldsymbol{\theta}$. Return to step 1.
    - Death: Draw a fibre F_j from $F_1, ..., F_k$ with probabilities $\delta_1, ..., \delta_k$ respectively and remove fibre F_j from $\mathbf{F}$, leaving $\{F_1, ..., F_{j-1}, F_{j+1}, ..., F_k\}$. Similarly, remove θ_j from $\boldsymbol{\theta}$. Return to step 1.

5.2.1 Birth Density

The birth density $b(F_{k+1}, \omega_{k+1}, l_{k+1}, \mathbf{Z}', \mathbf{X}', \mathbf{p}')$ is constructed as follows. Recall that the parametrisation of fibres is described in Section 3.2.1.

Birth events occur randomly at rate β. Upon the occurrence of a birth, the number of fibres is updated from k to $k + 1$, and a new fibre is introduced by sampling a reference point ω_{k+1} and lengths $l_{k+1,1}, l_{k+1,2}$ from the prior distributions $P(\omega), P(l)$ respectively. Recall that the prior for the reference point is

$$P(\omega) = \frac{1}{|W|}, \tag{5.1}$$

and the prior for the lengths is

$$P(l_{k+1}|k+1,\lambda) = P(l_{k+1,1}|\lambda)P(l_{k+1,2}|\lambda) \qquad \text{where } l_{k+1,\cdot} \sim \text{Exp}(1/\lambda). \tag{5.2}$$

The new fibre F_{k+1} is then calculated by integrating the field of orientations according to these parameters, and the set of fibres $\mathbf{F} = \{F_1, ..., F_k\}$ is updated to $\mathbf{F}' = \{F_1, ..., F_k, F_{k+1}\}$. To ensure that the distribution of the lengths $l_{k+1,1}, l_{k+1,2}$ is independent of the respective directions in which the field of orientations is integrated, $l_{k+1,1}$ and $l_{k+1,2}$ are independently and identically distributed.

If $F_{k+1} \not\subseteq W$, which is to say the proposed fibre does not lie completely within W, then the fibre is re-sampled until $F_{k+1} \subset W$. This reduces edge effects as explained in Section 3.2.2.

Data points are assigned to the new fibre and anchor points $\mathbf{p}'$ are proposed with proposal densities $Q_{\text{sig,birth}}(\mathbf{Z}'|\cdot)$ and $Q_{\text{aux,birth}}(\mathbf{X}', \mathbf{p}'|\cdot)$.

First, the signal/noise allocation ($\mathbf{Z}$) is updated conditional on the new set of fibres and the current allocation of the points to fibres, then the other auxiliary variables $\mathbf{X}$ and $\mathbf{p}$ are resampled depending also on the updated signal/noise allocation ($\mathbf{Z}'$).

In full, the birth density of fibre F_{k+1} is proportional to

$$\begin{aligned} &b(F_{k+1}, \omega_{k+1}, l_{k+1}, \mathbf{Z}', \mathbf{X}', \mathbf{p}') \\ &= P(\omega_{k+1})P(l_{k+1})Q_{\text{sig,birth}}(\mathbf{Z}'|\mathbf{F}', \mathbf{l}', \boldsymbol{\varepsilon}, \mathbf{Z}, \mathbf{X}, \mathbf{y}, \sigma_{\text{disp}}) \\ &\times Q_{\text{aux,birth}}(\mathbf{X}', \mathbf{p}'|\mathbf{F}', \mathbf{Z}', \mathbf{X}, \mathbf{p}, \mathbf{y}, \sigma_{\text{disp}})\mathbf{1}_{[F_{k+1} \subset W]} \end{aligned} \tag{5.3}$$

where $P(\omega_{k+1})$ and $P(l_{k+1})$ are prior densities of reference point ω_{k+1} and lengths l_{k+1} respectively, and $\mathbf{1}_{[\cdot]}$ is the indicator function.

5.2.2 Death Rates

A death rate δ_j is calculated for each fibre to ensure detailed balance holds and the target distribution is the posterior density.

When the death of fibre F_j occurs, the following variables are updated :

$$\mathbf{F} \mapsto \mathbf{F} \backslash F_j$$

$$\mathbf{l} \mapsto \mathbf{l} \backslash l_j$$

$$\boldsymbol{\omega} \mapsto \boldsymbol{\omega} \backslash \omega_j$$

$$k \mapsto k - 1$$

$$\mathbf{Z} \mapsto \mathbf{Z}'$$

$$\mathbf{X} \mapsto \mathbf{X}'$$

$$\mathbf{p} \mapsto \mathbf{p}'.$$

The auxiliary variables $\mathbf{Z}, \mathbf{X}$, and $\mathbf{p}$ are proposed from densities $Q_{\text{sig,death}}(\cdot)$ and $Q_{\text{aux,death}}(\cdot)$, details of which are given in Section 5.2.3.

To maintain detailed balance the death rates δ_j of fibre F_j must satisfy

$$\begin{aligned}\delta_j = \frac{\beta}{k} & \frac{\pi(\mathbf{F}\backslash F_j, \mathbf{l}\backslash l_j, \boldsymbol{\omega}\backslash\omega_j, k-1, \upsilon_{\text{FO}}, \boldsymbol{\varepsilon}, \mathbf{Z}', \mathbf{X}', \mathbf{p}')}{\pi(\mathbf{F}, \mathbf{l}, \boldsymbol{\omega}, k, \upsilon_{\text{FO}}, \boldsymbol{\varepsilon}, \mathbf{Z}, \mathbf{X}, \mathbf{p})} \\ & \times \frac{b(F_j, \omega_j, l_j, \mathbf{X}, \mathbf{p})}{Q_{\text{sig,death}}(\mathbf{Z}'|\mathbf{F}', \mathbf{l}', \boldsymbol{\varepsilon}, \mathbf{Z}, \mathbf{X}, \mathbf{y}, \sigma_{\text{disp}}) Q_{\text{aux,death}}(\mathbf{X}', \mathbf{p}'|\mathbf{F}', \mathbf{Z}', \mathbf{X}, \mathbf{p}, \mathbf{y}, \sigma_{\text{disp}})} \end{aligned} \tag{5.4}$$

where $\pi(\cdot)$ is the posterior density.

Evaluation of the components gives

$$\begin{aligned}\delta_j = \frac{\beta}{k} & \frac{P(\mathbf{l}\backslash l_j|k-1,\lambda)}{P(\mathbf{l}|k,\lambda)} \frac{P(\boldsymbol{\omega}\backslash\omega_j|k-1)}{P(\boldsymbol{\omega}|k)} \frac{P(k-1|\kappa)}{P(k|\kappa)} \\ & \times \frac{P(m|\mathbf{l}\backslash l_j, \eta, \alpha_{\text{signal}}, \beta_{\text{signal}})}{P(m|\mathbf{l}, \eta, \alpha_{\text{signal}}, \beta_{\text{signal}})} \frac{P(\mathbf{X}'|\mathbf{Z}, \mathbf{l}\backslash l_j)}{P(\mathbf{X}|\mathbf{Z}, \mathbf{l})} \frac{P(\mathbf{Z}'|\boldsymbol{\varepsilon})}{P(\mathbf{Z}|\boldsymbol{\varepsilon})} \\ & \times \frac{P(\mathbf{p}'|\mathbf{F}\backslash F_j, \mathbf{X}', \alpha_{Dir})}{P(\mathbf{p}|\mathbf{F}, \mathbf{X}, \alpha_{Dir})} \frac{L(\mathbf{p}'|\sigma^2_{\text{disp}}, \mathbf{y})}{L(\mathbf{p}|\sigma^2_{\text{disp}}, \mathbf{y})} \\ & \times \frac{b(F_j, \omega_j, l_j, \mathbf{X}, \mathbf{p})}{Q_{\text{sig,death}}(\mathbf{Z}'|\mathbf{F}', \mathbf{l}', \boldsymbol{\varepsilon}, \mathbf{Z}, \mathbf{X}, \mathbf{y}, \sigma_{\text{disp}}) Q_{\text{aux,death}}(\mathbf{X}', \mathbf{p}'|\mathbf{F}', \mathbf{Z}', \mathbf{X}, \mathbf{p}, \mathbf{y}, \sigma_{\text{disp}})} \end{aligned} \tag{5.5}$$

$$= \frac{\beta}{\kappa} \exp\left(\eta \frac{\alpha_{\text{signal}} + \beta_{\text{signal}}}{\alpha_{\text{signal}}} l_{j,T}\right) \frac{D(\mathbf{q}'(\mathbf{p}'), \alpha_{\text{Dir}})}{D(\mathbf{q}(\mathbf{p}), \alpha_{\text{Dir}})} \frac{\prod_{i \text{ s.t. } Z_i'=0} l_T'}{\prod_{i \text{ s.t. } Z_i=0} l_T}$$

$$\times \prod_{i \text{ s.t. } X_i=j, Z_i'=1} \left(\frac{\exp(-\text{dist}(p_i', y_i)^2/(2\sigma_{\text{disp}}^2))}{\exp(-\text{dist}(p_i, y_i)^2/(2\sigma_{\text{disp}}^2))}\right)$$

$$\times \prod_{i \text{ s.t. } X_i=j, Z_i'=0} \left(\frac{1-\varepsilon_i}{\varepsilon_i} \frac{1/|W|}{\exp(-\text{dist}(p_i', y_i)^2/(2\sigma_{\text{disp}}^2)) \frac{1}{2\pi\sigma_{\text{disp}}^2}}\right)$$

$$\times \frac{Q_{\text{sig,birth}}(\mathbf{Z}|\mathbf{F}, \mathbf{l}, \boldsymbol{\varepsilon}, \mathbf{Z}', \mathbf{X}', \mathbf{y}, \sigma_{\text{disp}}) Q_{\text{aux,birth}}(\mathbf{X}, \mathbf{p}|\mathbf{F}, \mathbf{Z}, \mathbf{X}', \mathbf{p}', \mathbf{y}, \sigma_{\text{disp}})}{Q_{\text{sig,death}}(\mathbf{Z}'|\mathbf{F}', \mathbf{l}', \boldsymbol{\varepsilon}, \mathbf{Z}, \mathbf{X}, \mathbf{y}, \sigma_{\text{disp}}) Q_{\text{aux,death}}(\mathbf{X}', \mathbf{p}'|\mathbf{F}', \mathbf{Z}', \mathbf{X}, \mathbf{p}, \mathbf{y}, \sigma_{\text{disp}})}, \quad (5.6)$$

where l_T and l_T' are the total lengths of all fibres in the current and proposed fibre sets respectively. The proposal densities $Q_{\text{sig},\cdot}(\cdot)$ and $Q_{\text{aux},\cdot}(\cdot)$ are defined in the following section.

5.2.3 Updating Auxiliary Variables

Following the birth or death of a fibre we update auxiliary variables: $\mathbf{Z}$ to $\mathbf{Z}'$, the indicator of which component (signal/noise) each point is associated to; $\mathbf{X}$ to $\mathbf{X}'$, the indicator of which fibre each signal point is associated to; $\mathbf{p}$ to $\mathbf{p}'$, the vector $(p_1, ..., p_m)$ where p_i is the anchor point on the fibre that data point y_i is associated to.

First, the signal/noise indicator $\mathbf{Z}$ is drawn from proposal probability $Q_{\text{sig},\cdot}(\mathbf{Z}'|\mathbf{F}', \mathbf{l}, \boldsymbol{\varepsilon}, \mathbf{Z}, \mathbf{X}, \mathbf{y}, \sigma_{\text{disp}})$, where the '$\cdot$' denotes 'birth' or 'death' depending on which move is being proposed. This probability depends on the distance of each data point y_i to the fibres in $\mathbf{F}'$ and the prior probability that y_i is signal, denoted by ε_i.

Second, an anchor point is proposed for each signal point, drawing ${p_i}'$ and ${X_i}'$ from a density proportional to the likelihood $L(p_i'|y_i, \sigma_{\text{disp}}^2, Z_i' = 1)$. The general proposal density is

$$Q_{\text{aux}}(p_i', X_i'|\mathbf{F}', y_i, \sigma_{\text{disp}}^2, Z_i' = 1) = \frac{L(p_i'|y_i, \sigma_{\text{disp}}^2, Z_i' = 1)}{\sum_{j=1}^{k} \int_{F_j'} L(p|y_i, \sigma_{\text{disp}}^2, Z_i' = 1) \text{d}p}. \quad (5.7)$$

The normalising constant is written

$$N(y_i, \mathbf{F}', \sigma^2_{\text{disp}}, Z'_i = 1) := \sum_{j=1}^{k} \int_{F'_j} L(p|y_i, \sigma^2_{\text{disp}}, Z'_i = 1)\mathrm{d}p \tag{5.8}$$

$$= \frac{L(p'_i|y_i, \sigma^2_{\text{disp}}, Z'_i = 1)}{Q_{\text{aux}}(p'_i, X'_i|\mathbf{F}', y_i, \sigma^2_{\text{disp}}, Z'_i = 1)}, \tag{5.9}$$

a term that is also used in proposal probabilities for $\mathbf{Z}$. Note that this is not a sample from the full conditional of $\mathbf{p}$ and $\mathbf{X}$ as the proposal density does not involve the Dirichlet distribution of $\mathbf{p}$ or the prior probabilities of allocating points to fibres: $P(X_i = j|\mathbf{l}, Z_i = 1) \propto l_{j,T}$.

This is the general construction for the proposal of $\mathbf{Z}', \mathbf{X}'$ and $\mathbf{p}'$. Alterations are made for birth and death moves.

Updating p and X following a birth

Following a birth, the update of $\mathbf{p}$ and $\mathbf{X}$ is restricted, allowing only the reallocation of points to the new fibre F_j. Noise or signal points that lie near the new fibre are likely to be reallocated to it. If point y_i is allocated to the new fibre, its anchor point p_i is sampled from the set of all points on fibre F_j.

As before, the proposal density of p'_i is proportional to the likelihood $L(p'_i|y_i, \sigma^2_{\text{disp}}, Z'_i = 1)$, but the normalising constant is now

$$N_j(y_i, F'_j, \sigma^2_{\text{disp}}, X'_i = j) := \int_{F'_j} L(p|y_i, \sigma^2_{\text{disp}}, Z'_i = 1)\mathrm{d}p. \tag{5.10}$$

Once a new fibre F_j has been proposed, p_i, X_i and Z_i are updated according to the following rules:

1. If $Z_i = 0$, then $Z'_i = 1$ and $X'_i = j$ with probability

$$Q_{\text{sig,birth}}(Z'_i = 1|\mathbf{F}', \mathbf{l}, \boldsymbol{\varepsilon}, Z_i = 0, \sigma_{\text{disp}}) \tag{5.11}$$

$$= \frac{N_j(y_i, F'_j, \sigma^2_{\text{disp}}, X'_i = j)\varepsilon_i / l_{j,T}}{N_j(y_i, F'_j, \sigma^2_{\text{disp}}, X'_i = j)\varepsilon_i / l_{j,T} + (1 - \varepsilon_i)/|W|},$$

   otherwise $Z'_i = 0$.

2. If $Z_i = 1$, then $Z'_i = 1$ ($Q_{\text{sig,birth}}(Z'_i = 1|\mathbf{F}', \mathbf{l}, \boldsymbol{\varepsilon}, Z_i = 1, \sigma_{\text{disp}}) = 1$) and $X'_i = j$

with probability

$$\frac{N_j(y_i, F'_j, \sigma^2_{\text{disp}}, X'_i = j)/l_{j,T}}{N_j(y_i, F'_j, \sigma^2_{\text{disp}}, X'_i = j)/l_{j,T} + N_{X_i}(y_i, F'_{X_i}, \sigma^2_{\text{disp}}, X'_i = X_i)/l_{X_i,T}}, \tag{5.12}$$

otherwise $X'_i = X_i$.

3. If the data point has been assigned to the new fibre ($X'_i = j$), the anchor point p_i is sampled from the density

$$\frac{L(p'_i|y_i, \sigma^2_{\text{disp}}, Z'_i = 1)}{N_j(y_i, F'_j, \sigma^2_{\text{disp}}, X'_i = j)}. \tag{5.13}$$

The combined proposal density for the auxiliary variables following a birth is

$$Q_{\text{sig,birth}}(\mathbf{Z}'|\mathbf{F}', \mathbf{l}', \boldsymbol{\varepsilon}, \mathbf{Z}, \mathbf{X}, \mathbf{y}, \sigma_{\text{disp}}) Q_{\text{aux,birth}}(\mathbf{X}', \mathbf{p}'|\mathbf{F}', \mathbf{Z}', \mathbf{X}, \mathbf{p}, \mathbf{y}, \sigma_{\text{disp}}) = \tag{5.14}$$

$$\prod_{i=1}^{m} \left(\mathbf{1}_{[Z'_i=0]} \frac{\mathbf{1}_{[Z_i=1]} L(p'_i|y_i, \sigma^2_{\text{disp}}, Z'_i = 1)\varepsilon_i/l_{j,T} + \mathbf{1}_{[Z_i=0]}(1-\varepsilon_i)/|W|}{N_j(y_i, F'_j, \sigma^2_{\text{disp}}, X'_i = j)\varepsilon_i/l_{j,T} + (1-\varepsilon_i)/|W|} + \mathbf{1}_{[Z'_i=1]} \right. \tag{5.15}$$

$$\left. \times \frac{\mathbf{1}_{[X_i=j]} L(p'_i|y_i, \sigma^2_{\text{disp}}, Z'_i = 1)/l_{j,T} + \mathbf{1}_{[X_i \neq j]} N_{X_i}(y_i|F_{X_i}, \sigma^2_{\text{disp}}, X'_i = X_i)/l_{X_i,T}}{N_j(y_i, F'_j, \sigma^2_{\text{disp}}, X'_i = j)/l_{j,T} + N_{X_i}(y_i, F_{X_i}, \sigma^2_{\text{disp}}, X'_i = X_i)/l_{X_i,T}} \right).$$

Updating p and X following a death

If a fibre dies then any points allocated to that fibre must be reallocated. Each point is either allocated to another fibre or to the noise component.

In full, auxiliary variables $\mathbf{Z}'$, $\mathbf{X}'$ and $\mathbf{p}'$ are updated according to the following rules:

1. If $Z_i = 0$ then $Z'_i = Z_i$:

$$Q_{\text{sig,death}}(Z'_i|\mathbf{F}', \mathbf{l}', \boldsymbol{\varepsilon}, Z_i = 0, \mathbf{X}, \mathbf{y}, \sigma_{\text{disp}}) = \mathbf{1}_{[Z'_i=0]}. \tag{5.16}$$

2. If $Z_i = 1$ and $X_i \neq j$ then $Z'_i = Z_i$, $X'_i = X_i$ and $p'_i = p_i$:

$$Q_{\text{sig,death}}(Z'_i|\mathbf{F}', \mathbf{l}', \boldsymbol{\varepsilon}, Z_i = 1, X_i \neq j, \mathbf{y}, \sigma_{\text{disp}}) = \mathbf{1}_{[Z'_i=1]} \tag{5.17}$$

and

$$Q_{\text{aux,death}}(X'_i = X_i, p'_i = p_i|\mathbf{F}', Z_i = 1, X_i \neq j, \mathbf{p}, \mathbf{y}, \sigma_{\text{disp}}) = \mathbf{1}_{[Z'_i=1]}. \tag{5.18}$$

3. If $X_i = j$ the point remains in the signal component with probability

$$Q_{\text{sig,death}}(Z_i' = 1|\mathbf{F}', \mathbf{l}', \boldsymbol{\varepsilon}, Z_i = 1, X_i = j, \mathbf{y}, \sigma_{\text{disp}}) \tag{5.19}$$
$$= \frac{N(y_i, \mathbf{F}', \sigma^2_{\text{disp}}, Z_i' = 1)\varepsilon_i}{N(y_i, \mathbf{F}', \sigma^2_{\text{disp}}, Z_i' = 1)\varepsilon_i + (1 - \varepsilon_i)/|W|},$$

   otherwise $Z_i' = 0$, the point is allocated to noise.

4. If $X_i = j$ and the point remains in the signal component ($Z_i' = 1$), then it is reallocated to another fibre by sampling p_i' and X_i' from $Q_{\text{aux}}(p_i', X_i'|\cdot)$ as defined in Equation (5.7).

The combined proposal density of the auxiliary variables following the death of a fibre is

$$Q_{\text{sig,death}}(\mathbf{Z}'|\mathbf{F}', \mathbf{l}', \boldsymbol{\varepsilon}, \mathbf{Z}, \mathbf{X}, \mathbf{y}, \sigma_{\text{disp}}) \tag{5.20}$$
$$\times Q_{\text{aux,death}}(\mathbf{X}', \mathbf{p}'|\mathbf{F}', \mathbf{Z}', \mathbf{X}, \mathbf{p}, \mathbf{y}, \sigma_{\text{disp}})$$
$$= \prod_{i=1}^{m} \left(\mathbf{1}_{[X_i=j]} \frac{\mathbf{1}_{[Z_i'=1]} L(p_i', |y_i, \sigma^2_{\text{disp}}, Z_i' = 1)\varepsilon_i + \mathbf{1}_{[Z_i'=0]}(1 - \varepsilon_i)/|W|}{N(y_i, \mathbf{F}', \sigma^2_{\text{disp}}, Z_i' = 1)\varepsilon_i + (1 - \varepsilon_i)/|W|} \right).$$

The primary motivation for not reallocating all points following a birth or death is to localise the move. The consequence is that the birth or death of one fibre has little effect on the death rates of other fibres.

Practical Implementation of Updating Auxiliary Variables

When drawing X_i, p_i from $Q_{\text{aux}}(\cdot)$, the fibre X_i is sampled first, then the anchor point p_i.

Following from Equation (5.7), point y_i is allocated to fibre F_j with probability

$$\frac{\int_{F_j'} L(p|y_i, \sigma^2_{\text{disp}}, Z_i' = 1)\mathrm{d}p}{N(y_i, \mathbf{F}', \sigma^2_{\text{disp}}, Z_i' = 1)}. \tag{5.21}$$

The integral is estimated by summing over a discrete set of points, $\Xi_j = \{\xi_{j,1}, \xi_{j,2}...\}$ regularly spaced at unit-lengths along the fibre F_j:

$$\int_{F_j'} L(p|y_i, \sigma^2_{\text{disp}}, Z_i' = 1)\mathrm{d}p \approx \sum_{l=1,...} L(p_i = \xi_{j,l}|y_i, \sigma^2_{\text{disp}}, Z_i' = 1). \tag{5.22}$$

Storing fibres as a sequence of piece-wise linear segments of unit-length provides

a natural choice for the $\xi_{j,l}$: the midpoint of each segment. It also reduces the computational time required to recalculate fibres.

Anchor point p_i is proposed by drawing $\xi_{j,l}$ with probability

$$\frac{L(p_i = \xi_{j,l}|y_i, \sigma^2_{\text{disp}}, Z'_i = 1)}{\sum_{l=1,\ldots} L(p_i = \xi_{j,l}|y_i, \sigma^2_{\text{disp}}, Z'_i = 1)}, \tag{5.23}$$

and sampling p_i from the density proportional to $L(p_i = \xi_{j,l}|y_i, \sigma^2_{\text{disp}}, Z'_i = 1)$ conditional that it lies on the linear segment with midpoint $\xi_{j,l}$.

5.3 Additional Moves

It is highly desirable to add extra moves to the BDMCMC process to improve mixing. Some possible moves which were all utilised in the examples in Section 6 include

- Moving a fibre by a small amount (by perturbing the reference point),
- Resampling the lengths of a fibre (while keeping the reference point fixed),

Each of these events occur at some predefined rate, whence they are proposed and either accepted or rejected according to the Metropolis Hastings probability.

We may also wish to update other model variables, giving more flexibility and improving the algorithm's exploration of the sample space. The additional variable updates used in the examples in Section 6 include

- Proposing new signal-noise allocations of the data ($\mathbf{Z}$),
- Proposing new signal probabilities ($\boldsymbol{\varepsilon}$) - this move leads to an update in the prior for the field of orientations due to the empirical Bayes step, hence all fibres are resampled.

Hyperprior parameters, such as the constant of proportionality η in the prior for the Poisson-distributed number of points or σ_{disp} governing the deviation of points from fibres may also be updated. We have chosen not to update any hyperprior parameters in order to reduce complexity of the model.

Generally it is not feasible to draw from full conditional distributions, so we use Metropolis-Hastings updates.

Details of these moves are given in the following sections.

5.3.1 Updating Signal Probabilities

Updating the probability ε_i that point y_i is signal necessarily requires the recalculation of the field of orientations. This is because the empirical Bayes relationship between $\mathbf{y}$, $\boldsymbol{\varepsilon}$ and υ_{FO} is deterministic. Indeed, this provides a convenient means of updating the field of orientations.

The proposed update is accepted or rejected according to the Metropolis-Hastings acceptance probability.

Proposal Density

An update of $\boldsymbol{\varepsilon}$ is proposed from the conditional distribution $P(\boldsymbol{\varepsilon}|\mathbf{Z}, \alpha_{\mathrm{signal}}, \beta_{\mathrm{signal}}, m)$. For $i = 1, ..., m$,

$$\varepsilon_i' \sim \mathrm{Beta}(\alpha_{\mathrm{signal}} + \mathbf{1}_{[Z_i=1]}, \beta_{\mathrm{signal}} + \mathbf{1}_{[Z_i=0]}) \tag{5.24}$$

where $\mathbf{1}_{[Z_i=\cdot]}$ is the indicator function.

Given $\boldsymbol{\varepsilon}'$ the field of orientations υ_{FO}' is proposed deterministically through the empirical Bayes prior,

$$\upsilon_{\mathrm{FO}(\mathbf{y},\boldsymbol{\varepsilon},h_{\mathrm{FO}},\sigma_{\mathrm{FO}})} \mapsto \upsilon_{\mathrm{FO}}' = \upsilon_{\mathrm{FO}(\mathbf{y},\boldsymbol{\varepsilon}',h_{\mathrm{FO}},\sigma_{\mathrm{FO}})}', \tag{5.25}$$

where $\upsilon_{\mathrm{FO}(\mathbf{y},\boldsymbol{\varepsilon},h_{\mathrm{FO}},\sigma_{\mathrm{FO}})}$ indicates the orientation field as estimated in Section 4.4.2.

The set of fibres is also deterministically updated without altering reference points $\boldsymbol{\omega}$ or lengths $\mathbf{l}$:

$$\mathbf{F}(\boldsymbol{\omega}, \mathbf{l}, \upsilon_{\mathrm{FO}(\mathbf{y},\boldsymbol{\varepsilon},h_{\mathrm{FO}},\sigma_{\mathrm{FO}})}) \mapsto \mathbf{F}' = \mathbf{F}'(\boldsymbol{\omega}, \mathbf{l}, \upsilon_{\mathrm{FO}(\mathbf{y},\boldsymbol{\varepsilon}',h_{\mathrm{FO}},\sigma_{\mathrm{FO}})}'). \tag{5.26}$$

The notation corresponds with that of Section 3.2.2, so that each fibre in the set $\mathbf{F}(\boldsymbol{\omega}, \mathbf{l}, \upsilon_{\mathrm{FO}(\mathbf{y},\boldsymbol{\varepsilon},h_{\mathrm{FO}},\sigma_{\mathrm{FO}})})$ is determined by $F_j(\omega_j, l_j, \upsilon_{\mathrm{FO}(\mathbf{y},\boldsymbol{\varepsilon},h_{\mathrm{FO}},\sigma_{\mathrm{FO}})})$. If any fibre in $\mathbf{F}'$ does not completely lie within W the proposed move is rejected, as discussed in Section 3.2.2.

Auxiliary variables $\mathbf{X}'$ and $\mathbf{p}'$ are proposed from density $Q_{\mathrm{aux}}(p_i', X_i'|\mathbf{F}', y_i, \sigma^2_{\mathrm{disp}}, Z_i' = 1)$, as given in section 5.2.3. Note that we do not update $\mathbf{Z}$, the allocation of points to signal or noise, so $Z_i' = Z_i$.

Hence the proposal density is

$$
\begin{aligned}
&Q_{\varepsilon}(\boldsymbol{\varepsilon}', \upsilon'_{\mathrm{FO}}, \mathbf{F}', \mathbf{P}', \mathbf{X}') \\
&= \prod_{i=1}^{m} \frac{(\varepsilon'_i)^{\alpha_{\mathrm{signal}}+\mathbf{1}_{[Z_i=1]}-1}(1-\varepsilon'_i)^{\beta_{\mathrm{signal}}+\mathbf{1}_{[Z_i=0]}-1}}{\mathrm{B}(\alpha_{\mathrm{signal}}+\mathbf{1}_{[Z_i=1]}, \beta_{\mathrm{signal}}+\mathbf{1}_{[Z_i=0]})} \\
&\times \prod_{i \text{ s.t. } Z'_i=1} Q_{\mathrm{aux}}(p'_i, X'_i|\mathbf{F}', y_i, \sigma^2_{\mathrm{disp}}, Z'_i=1) \prod_{j=1}^{k} \mathbf{1}_{[F_j \subset W]}
\end{aligned}
\tag{5.27}
$$

where $\mathrm{B}(\cdot,\cdot)$ is the Beta function.

Acceptance Probability

The proposed state is accepted with probability

$$
\min\left\{1, \frac{\pi(\mathbf{F}', \mathbf{l}, \boldsymbol{\omega}, k, \upsilon'_{\mathrm{FO}}, \boldsymbol{\varepsilon}', \mathbf{Z}, \mathbf{X}', \mathbf{p}')}{\pi(\mathbf{F}, \mathbf{l}, \boldsymbol{\omega}, k, \upsilon_{\mathrm{FO}}, \boldsymbol{\varepsilon}, \mathbf{Z}, \mathbf{X}, \mathbf{p})} \frac{Q_{\varepsilon}(\boldsymbol{\varepsilon}, \upsilon_{\mathrm{FO}}, \mathbf{F}, \mathbf{P}, \mathbf{X})}{Q_{\varepsilon}(\boldsymbol{\varepsilon}', \upsilon'_{\mathrm{FO}}, \mathbf{F}', \mathbf{P}', \mathbf{X}')}\right\}. \tag{5.28}
$$

The product of density ratios evaluates to

$$
\begin{aligned}
&\frac{P(\mathbf{F}'|\boldsymbol{\omega}, l, \upsilon'_{\mathrm{FO}})}{P(\mathbf{F}|\boldsymbol{\omega}, l, \upsilon_{\mathrm{FO}})} \frac{P(\boldsymbol{\varepsilon}'|m, \alpha_{\mathrm{signal}}, \beta_{\mathrm{signal}})}{P(\boldsymbol{\varepsilon}|m, \alpha_{\mathrm{signal}}, \beta_{\mathrm{signal}})} \frac{P(\mathbf{Z}|\boldsymbol{\varepsilon}')}{P(\mathbf{Z}|\boldsymbol{\varepsilon})} \frac{P(\mathbf{X}'|\mathbf{Z}, \mathbf{l})}{P(\mathbf{X}|\mathbf{Z}, \mathbf{l})} \frac{P(\mathbf{p}'|\mathbf{F}', \mathbf{X}', \alpha_{Dir})}{P(\mathbf{p}|\mathbf{F}, \mathbf{X}, \alpha_{Dir})} \\
&\times \frac{L(\mathbf{p}'|\sigma^2_{\mathrm{disp}}, \mathbf{y})}{L(\mathbf{p}|\sigma^2_{\mathrm{disp}}, \mathbf{y})} \prod_{i=1}^{m} \mathbf{1}_{[Z_i=Z'_i]} \frac{\varepsilon_i^{\alpha_{\mathrm{signal}}+\mathbf{1}_{[Z_i=1]}-1}(1-\varepsilon_i)^{\beta_{\mathrm{signal}}+\mathbf{1}_{[Z_i=0]}-1}}{\mathrm{B}(\alpha_{\mathrm{signal}}+\mathbf{1}_{[Z_i=1]}, \beta_{\mathrm{signal}}+\mathbf{1}_{[Z_i=0]})} \\
&\times \prod_{i=1}^{m} \frac{\mathrm{B}(\alpha_{\mathrm{signal}}+\mathbf{1}_{[Z_i=1]}, \beta_{\mathrm{signal}}+\mathbf{1}_{[Z_i=0]})}{(\varepsilon'_i)^{\alpha_{\mathrm{signal}}+\mathbf{1}_{[Z_i=1]}-1}(1-\varepsilon'_i)^{\beta_{\mathrm{signal}}+\mathbf{1}_{[Z_i=0]}-1}} \\
&\times \prod_{i \text{ s.t. } Z_i=1} \frac{L(p'_i|y_i, \sigma^2_{\mathrm{disp}}, Z'_i=1)}{N(y_i, \mathbf{F}', \sigma^2_{\mathrm{disp}}, Z'_i=1)} \frac{N(y_i, \mathbf{F}, \sigma^2_{\mathrm{disp}}, Z_i=1)}{L(p_i|y_i, \sigma^2_{\mathrm{disp}}, Z_i=1)} \prod_{j=1}^{k} \mathbf{1}_{[F_j \subset W]} \\
&= \frac{D(\mathbf{q}'(\mathbf{p}'), \alpha_{\mathrm{Dir}})}{D(\mathbf{q}(\mathbf{p}), \alpha_{\mathrm{Dir}})} \prod_{i \text{ s.t. } Z_i=1} \mathbf{1}_{[Z_i=Z'_i]} \frac{N(y_i, \mathbf{F}, \sigma^2_{\mathrm{disp}}, Z_i=1)}{N(y_i, \mathbf{F}', \sigma^2_{\mathrm{disp}}, Z'_i=1)}.
\end{aligned}
\tag{5.29}
$$

The term $N(y_i, \mathbf{F}, \sigma^2_{\mathrm{disp}}, Z_i=1)$ is defined in Equation (5.8), and $D(\mathbf{q}(\mathbf{p}), \alpha_{\mathrm{Dir}})$ is given in Section 3.2.2.

This move can be adapted to optimise the acceptance rate. Each ε_i is updated with probability ρ_{ε}, otherwise it keeps its value $\varepsilon'_i = \varepsilon_i$. Tests show that choosing $\rho_{\varepsilon} = \min\{1/2, 10/m\}$, where m is the expected number of data points, gives a reasonable acceptance rate.

Updating the field of orientations is one of the computationally slowest steps in the algorithm. This is due in part to the Fast-Fourier Transform calculation on large matrices (used to evaluate the kernel-smoothed tensor field), and also because every tensor must be eigen-decomposed. Therefore a trade-off is made between infrequent proposals which ensure that the BDMCMC is not too slow, and proposing the move sufficiently often that the chain mixes well. In the examples of Chapter 6, a proposal rate of 10% of the birth rate was used.

5.3.2 Moving a Fibre: Perturbation of the Reference Point

Making small adjustments to the location of a fibre allows faster movement between similar fibres than a pure birth-death process. This move is an extension of the *shift* move described in Huber [2011], where an element of a Poisson point process is chosen at random and replaced by a new element.

A single fibre F_j is moved by proposing a perturbation of the reference point ω_j while preserving the lengths l_j. The new fibre, together with the remaining $k-1$ fibres are accepted with the appropriate Metropolis-Hastings probability.

Proposal Density

The fibre to be moved is picked uniformly at random from the set of fibres $\{F_1, ..., F_k\}$. The proposed reference point ω_j', is drawn from an isotropic bivariate normal distribution centred at the current reference point ω_j with variance σ^2_{move}. The auxiliary variables $\mathbf{X}'$ and $\mathbf{p}'$ are proposed from proposal density $Q_{\text{aux}}(p_i', X_i' | \mathbf{F}', y_i, \sigma^2_{\text{disp}}, Z_i' = 1)$ as given in Section 5.2.3. We choose not to update $\mathbf{Z}$, the allocation of points to signal or noise.

In summary, changes to variables are proposed as follows:

$$\omega_j \mapsto \omega_j' \sim \text{MVN}(\omega_j, \sigma^2_{\text{move}}\mathbf{I}_2)$$

$$\boldsymbol{\omega} \mapsto \boldsymbol{\omega}' = \{\omega_1, ..., \omega_{j-1}, \omega_j', \omega_{j+1}, ..., \omega_k\}$$

$$\mathbf{F} \mapsto \mathbf{F}' = \mathbf{F}'(\upsilon_{\text{FO}}, \mathbf{l}, \boldsymbol{\omega}')$$

$$\mathbf{X} \mapsto \mathbf{X}'$$

$$\mathbf{p} \mapsto \mathbf{p}'.$$

If $\omega_j' \notin W$ or if $F_j \not\subset W$ the proposed move is automatically rejected. We make the approximation that the normalising constant in the forward proposal (of proposing

F') is equal to the normalising constant of the backwards proposal (or proposing F). The slight error, arising from the edge effects induced by rejecting fibres and reference points that lie outside W is moderated by proposing small changes in the reference point, i.e. using a small value of σ_{move}. Implementation of the BDMCMC has shown no evidence of this approximation affecting the chain's convergence or target density.

Acceptance Probability

The Metropolis-Hastings acceptance probability of this move is given by the minimum of 1 and

$$\begin{aligned}
&\frac{\pi(\mathbf{F}', \mathbf{l}, \boldsymbol{\omega}', k, \upsilon_{\text{FO}}, \boldsymbol{\varepsilon}, \mathbf{Z}, \mathbf{X}', \mathbf{p}')}{\pi(\mathbf{F}, \mathbf{l}, \boldsymbol{\omega}, k, \upsilon_{\text{FO}}, \boldsymbol{\varepsilon}, \mathbf{Z}, \mathbf{X}, \mathbf{p})} \frac{Q_{\text{move}}(\boldsymbol{\omega}, \mathbf{F}, \mathbf{p}, \mathbf{X})}{Q_{\text{move}}(\boldsymbol{\omega}', \mathbf{F}', \mathbf{p}', \mathbf{X}')} \\
&= \frac{P(\mathbf{F}'|\boldsymbol{\omega}', \mathbf{l}, \upsilon_{\text{FO}}) P(\boldsymbol{\omega}'|k) P(\mathbf{X}'|\mathbf{Z}, \mathbf{l}) P(\mathbf{p}'|\mathbf{F}', \mathbf{X}', \alpha_{Dir}) L(\mathbf{p}'|\sigma^2_{\text{disp}}, \mathbf{y})}{P(\mathbf{F}|\boldsymbol{\omega}, \mathbf{l}, \upsilon_{\text{FO}}) P(\boldsymbol{\omega}|k) P(\mathbf{X}|\mathbf{Z}, \mathbf{l}) P(\mathbf{p}|\mathbf{F}, \mathbf{X}, \alpha_{Dir}) L(\mathbf{p}|\sigma^2_{\text{disp}}, \mathbf{y})} \qquad (5.30) \\
&\times \frac{k}{k} \frac{\phi_{\omega_j, \sigma^2_{\text{move}}\mathbf{I}_2}(\omega'_j)}{\phi_{\omega'_j, \sigma^2_{\text{move}}\mathbf{I}_2}(\omega_j)} \prod_{i \text{ s.t. } Z_i=1} \frac{L(p_i|y_i, \sigma^2_{\text{disp}}, Z_i = 1)}{L(p'_i|y_i, \sigma^2_{\text{disp}}, Z'_i = 1)} \frac{N(y_i, \mathbf{F}', \sigma^2_{\text{disp}}, Z_i = 1)}{N(y_i, \mathbf{F}, \sigma^2_{\text{disp}}, Z_i = 1)} \mathbf{1}_{[F_j \subset W]} \\
&= \frac{D(\mathbf{q}'(\mathbf{p}'), \alpha_{\text{Dir}})}{D(\mathbf{q}(\mathbf{p}), \alpha_{\text{Dir}})} \prod_{i \text{ s.t. } Z_i=1} \frac{N(y_i, \mathbf{F}', \sigma^2_{\text{disp}}, Z_i = 1)}{N(y_i, \mathbf{F}, \sigma^2_{\text{disp}}, Z_i = 1)} \qquad (5.31)
\end{aligned}$$

where $\phi_{\mu, \sigma^2\mathbf{I}_2}$ is the bivariate normal density function with mean μ and covariance matrix $\sigma^2\mathbf{I}_2$, and $N(y_i, \ldots)$ is the normalising constant given in Equation (5.8).

The acceptance rate can be controlled by changing the variance σ^2_{move} in the proposal of ω_j. Tests indicate that a value of $\sigma_{\text{move}} \approx \frac{\sigma_{\text{disp}}}{2}$ performs well.

5.3.3 Updating Fibre Lengths

Like the previous move, updating the length of a fibre allows the chain to explore a number of possible fibres without waiting for the birth of a fibre in the same location.

Proposal Density

An update to the lengths $l_j = (l_{j,1}, l_{j,2})$ is proposed by choosing either $l_{j,1}$ or $l_{j,2}$ and adding a normally distributed perturbation. The choices of which fibre to update and which of $l_{j,1}$ or $l_{j,2}$ to perturb are drawn uniformly at random. In

summary,

$$j \sim \text{Uniform}(\{1, ..., k\})$$

$$u \sim \text{Norm}(0, \sigma^2_{\text{length}})$$

$$l_j \mapsto l_j{}' = \begin{cases} (l_{j,1} + u, l_{j,2}) & \text{with probability } 1/2 \\ (l_{j,1}, l_{j,2} + u) & \text{with probability } 1/2. \end{cases}$$

If either $l'_{j,1}$ or $l'_{j,2}$ are negative, the move is immediately rejected. Otherwise, we propose a new fibre $F_j{}'$

$$F_j \mapsto F_j{}' = F_j{}'(\upsilon_{\text{FO}}, l'_j, \omega_j).$$

If the change in length has resulted in $F'_j \not\subset W$ the proposed move is rejected.

All points are reallocated to either noise or signal. The proposal probability mass function for $\mathbf{Z}$ is

$$Q_{\text{sig,length}}(\mathbf{Z}'|\mathbf{F}', \boldsymbol{\varepsilon}, \mathbf{y}, \sigma^2_{\text{disp}}), \tag{5.32}$$

where the probability a point is allocated to signal, regardless of its current state and independent of other points, is

$$\frac{\varepsilon_i N(y_i, \mathbf{F}', \sigma^2_{\text{disp}}, Z'_i = 1)/l_T}{\varepsilon_i N(y_i, \mathbf{F}', \sigma^2_{\text{disp}}, Z'_i = 1)/l_T + (1 - \varepsilon_i)/|W|}. \tag{5.33}$$

The term $N(y_i, \mathbf{F}', \sigma^2_{\text{disp}}, Z'_i = 1)$ is defined in Equation (5.8), and l_T is the total length of all fibres. Otherwise the point is allocated to noise.

Finally X_i and p_i are updated for all data points in the signal component by drawing them from $Q_{\text{aux}}(p'_i, X'_i|\mathbf{F}', y_i, \sigma^2_{\text{disp}}, Z'_i = 1)$ given in Equation (5.7).

Acceptance Probability

Without loss of generality assume $l_{j,1}$ is the length updated. The acceptance probability is given by

$$\min\left\{1, \frac{\pi(\mathbf{F}', \mathbf{l}', \boldsymbol{\omega}, k, \upsilon_{\mathrm{FO}}, \boldsymbol{\varepsilon}, \mathbf{Z}, \mathbf{X}', \mathbf{p}')}{\pi(\mathbf{F}, \mathbf{l}, \boldsymbol{\omega}, k, \upsilon_{\mathrm{FO}}, \boldsymbol{\varepsilon}, \mathbf{Z}, \mathbf{X}, \mathbf{p})} \frac{Q_{\mathbf{l},\mathbf{F},\mathbf{p},\mathbf{X}}}{Q_{\mathbf{l}',\mathbf{F}',\mathbf{p}',\mathbf{X}'}}\right\} \tag{5.34}$$

$$= \frac{P(\mathbf{l}'|k,\lambda)P(\mathbf{F}'|\boldsymbol{\omega},\mathbf{l}',\upsilon_{\mathrm{FO}})P(m|\mathbf{l}',\eta)P(\mathbf{Z}'|\boldsymbol{\varepsilon})P(\mathbf{X}'|\mathbf{Z},\mathbf{l}')P(\mathbf{p}'|\mathbf{F}',\mathbf{X}',\alpha_{Dir})}{P(\mathbf{l}|k,\lambda)P(\mathbf{F}|\boldsymbol{\omega},\mathbf{l},\upsilon_{\mathrm{FO}})P(m|\mathbf{l},\eta)P(\mathbf{Z}|\boldsymbol{\varepsilon})P(\mathbf{X}|\mathbf{Z},\mathbf{l})P(\mathbf{p}|\mathbf{F},\mathbf{X},\alpha_{Dir})}$$

$$\times \frac{L(\mathbf{p}'|\mathbf{y},\sigma^2_{\mathrm{disp}})}{L(\mathbf{p}|\mathbf{y},\sigma^2_{\mathrm{disp}})} \frac{\phi_{l'_{j,1},\sigma^2_{\mathrm{length}}}(l_{j,1})}{\phi_{l_{j,1},\sigma^2_{\mathrm{length}}}(l'_{j,1})} \frac{2k}{2k} \frac{Q_{\mathrm{sig,length}}(\mathbf{Z}|\mathbf{F},\boldsymbol{\varepsilon},\mathbf{y},\sigma^2_{\mathrm{disp}})}{Q_{\mathrm{sig,length}}(\mathbf{Z}'|\mathbf{F}',\boldsymbol{\varepsilon},\mathbf{y},\sigma^2_{\mathrm{disp}})}$$

$$\times \prod_{i \text{ s.t. } Z'_i=1}^{m} \frac{N(y_i,\mathbf{F}',\sigma^2_{\mathrm{disp}},Z'_i=1)}{L(p'_i,|y_i,\sigma^2_{\mathrm{disp}},Z'_i=1)} \prod_{i \text{ s.t. } Z_i=1}^{m} \frac{L(p_i,|y_i,\sigma^2_{\mathrm{disp}},Z_i=1)}{N(y_i,\mathbf{F},\sigma^2_{\mathrm{disp}},Z_i=1)} \mathbf{1}_{[\mathbf{F}\subset W]}$$

$$= \exp\left((l_{j,T}-l'_{j,T})\left(1/\lambda+\eta\frac{\alpha_{\mathrm{signal}}+\beta_{\mathrm{signal}}}{\alpha_{\mathrm{signal}}}\right)\right)\frac{D(\mathbf{q}'(\mathbf{p}'),\alpha_{\mathrm{Dir}})}{D(\mathbf{q}(\mathbf{p}),\alpha_{\mathrm{Dir}})}$$

$$\times \prod_{i=1}^{m} \frac{\varepsilon_i N(y_i,\mathbf{F}',\sigma^2_{\mathrm{disp}},Z'_i=1)+(1-\varepsilon_i)l'_T/|W|}{\varepsilon_i N(y_i,\mathbf{F},\sigma^2_{\mathrm{disp}},Z_i=1)+(1-\varepsilon_i)l_T/|W|} \tag{5.35}$$

where ϕ_{μ,σ^2} is a univariate normal density function with mean μ and variance σ^2.

The rate of acceptance is controlled by varying $\sigma^2_{\mathrm{length}}$. Tests indicate that $\sigma_{\mathrm{length}} \approx \lambda/20$ gives a reasonable acceptance rate.

5.3.4 Updating Allocation of Points to Noise/Signal

It is beneficial to the mixing of the BDMCMC to include a move which updates the noise and signal allocations without also changing properties of the fibres.

Proposal Density

Each point y_i is allocated to the signal or noise component independent of its current state Z_i. Auxiliary variables X_i and p_i are then updated for points that have been allocated to signal.

The allocation of points to noise ($Z_i = 0$) or signal ($Z_i = 1$) is updated as fol-

lows,

$$Z_i \mapsto Z_i' = \begin{cases} 1 & \text{with prob.} \quad \frac{\varepsilon_i N(y_i,\mathbf{F},\sigma^2_{\text{disp}},Z_i=1)/l_T}{\varepsilon_i N(y_i,\mathbf{F},\sigma^2_{\text{disp}},Z_i=1)/l_T+(1-\varepsilon_i)/|W|} \\ 0 & \text{with prob.} \quad \frac{(1-\varepsilon_i)/|W|}{\varepsilon_i N(y_i,\mathbf{F},\sigma^2_{\text{disp}},Z_i=1)/l_T+(1-\varepsilon_i)/|W|} \end{cases} \tag{5.36}$$

where $l_T = \sum_{j=1}^{k} l_{j,T}$ is the total length of all fibres and $N(y_i, \mathbf{F}, \sigma^2_{\text{disp}}, Z_i = 1)$ is defined in Equation (5.8).

Auxiliary variables $p_i{}'$ and X_i' are drawn from $Q_{\text{aux}}(p_i', X_i'|\mathbf{F}', y_i, \sigma^2_{\text{disp}}, Z_i' = 1)$ given in Equation (5.7).

Acceptance Probability

The Metropolis-Hastings acceptance probability is equal to the minimum of 1 and

$$\begin{aligned} &\frac{\pi(\mathbf{F},\mathbf{l},\boldsymbol{\omega},k,\upsilon_{\text{FO}},\boldsymbol{\varepsilon},\mathbf{Z}',\mathbf{X}',\mathbf{p}')}{\pi(\mathbf{F},\mathbf{l},\boldsymbol{\omega},k,\upsilon_{\text{FO}},\boldsymbol{\varepsilon},\mathbf{Z},\mathbf{X},\mathbf{p})} \frac{Q_Z(\mathbf{Z},\mathbf{X},\mathbf{p})}{Q_Z(\mathbf{Z}',\mathbf{X}',\mathbf{p}')} \\ &= \prod_i \left(\frac{P(Z_i'|\varepsilon_i)P(X_i'|Z_i',\mathbf{L})P(p_i'|X_i',\mathbf{F})L(p_i'|y_i,\sigma^2_{\text{disp}})}{P(Z_i|\varepsilon_i)P(X_i|Z_i,\mathbf{L})P(p_i|X_i,\mathbf{F})L(p_i|y_i,\sigma^2_{\text{disp}})} \right. \\ &\left. \times \frac{\varepsilon_i N(y_i,\mathbf{F},\sigma^2_{\text{disp}},Z_i=1)\mathbf{1}_{[Z_i=1]}/l_T + \frac{1-\varepsilon_i}{|W|}\mathbf{1}_{[Z_i=0]}}{\varepsilon_i N(y_i,\mathbf{F},\sigma^2_{\text{disp}},Z_i'=1)\mathbf{1}_{[Z_i'=1]}/l_T + \frac{1-\varepsilon_i}{|W|}\mathbf{1}_{[Z_i'=0]}} \right) \qquad (5.37) \\ &\times \prod_{i \text{ s.t. } Z_i=1} \frac{L(p_i,|y_i,\sigma^2_{\text{disp}},Z_i=1)}{N(y_i,\mathbf{F},\sigma^2_{\text{disp}},Z_i=1)} \prod_{i \text{ s.t. } Z_i'=1} \frac{N(y_i,\mathbf{F},\sigma^2_{\text{disp}},Z_i'=1)}{L(p_i',|y_i,\sigma^2_{\text{disp}},Z_i'=1)} \\ &= \frac{D(\mathbf{q}'(\mathbf{p}'),\alpha_{\text{Dir}})}{D(\mathbf{q}(\mathbf{p}),\alpha_{\text{Dir}})}. \qquad (5.38) \end{aligned}$$

This choice of proposal density produces a reasonable acceptance rate if α_{Dir} is not too large.

5.3.5 Split and Join Moves

To improve the mixing of the MCMC, a second pair of reversible moves - *splits* and *joins*, are implemented. A split move is proposed by selecting a fibre from $\{F_1, ..., F_k\}$, splitting the fibre at a random point along its length, and perturbing the resultant fibres so that they do not touch. The reverse (a join move) is also introduced in order to maintain the reversibility of the BDMCMC.

Proposal Density for a Split Move

Henceforth, we simplify the notation of this section to help clarify the acceptance probability calculations. In particular, we consider the case of a single fibre splitting to two fibres, although the work in this section applies to any general set of fibres. The initial fibre F has lengths l_1, l_2 and reference point $\underline{\omega}$. The proposed fibres F_1 and F_2 have lengths $l_{1,1}, l_{1,2}$ (resp. $l_{2,1}, l_{2,2}$) and reference point $\underline{\omega}_1$ (resp. $\underline{\omega}_2$). The total length of fibre F is denoted $l_F = l_1 + l_2$ and similarly for $l_{1,F}$ (the length of F_1) and $l_{2,F}$ (the length of F_2). Underlined variables such as $\underline{\omega}$ indicate locations in the window W. As the field of orientations is fixed for join and split moves, it is omitted from the notation; the fibre F is now written $F = F(\underline{\omega}, l_1, l_2)$.

The field of orientations maps each point in W to an orientation (or undirected direction); for each orientation there are two possible directions of integration. The assumption is made that the field of orientations is locally smooth so that a direction of 'positive' orientation which is continuous in a neighbourhood of F can be identified. We choose the direction of integration corresponding to the length l_2 from reference point ω as the positive direction (the direction corresponding to length l_1 is thus the negative direction).

Hereafter, for $j = 1, 2$, $F(\underline{\omega}_j, l_{j,1}, l_{j,2})$ defines the fibre obtained by integrating the field of orientations from $\underline{\omega}_j$ to a length $l_{j,2}$ in the positive direction, and a length $l_{j,1}$ in the negative direction. The matter of choosing a positive direction of orientation is not encountered in birth or death moves as the lengths $l_{j,1}$ and $l_{j,2}$ are identically and independently distributed.

We introduce a function $\varphi : W \times (-\infty, \infty) \mapsto \mathbb{R}^2$ where $\varphi(\underline{x}, d)$ is the point reached after integrating the field of orientations in the positive direction from $\underline{x}$ to a distance d. For example $\varphi(\underline{x}, 0) = \underline{x}$, and $\varphi(\underline{\omega}, -l_1)$ and $\varphi(\underline{\omega}, l_2)$ identify the end points of fibre $F = F(\underline{\omega}, l_1, l_2)$. The fibre F can now be written $F(\underline{\omega}, l_1, l_2) = \{\varphi(\underline{\omega}, d) : -l_1 \leq d \leq l_2\}$.

In the general case, where there is more than one fibre in the initial state, fibre $F_j \in \mathbf{F}$ is chosen by drawing j from a proposal distribution over $\{1, ..., k\}$. For simplicity we use a uniform distribution over the discrete set; each fibre is chosen with probability $1/k$.

A random number $u_{\text{split}} \sim \text{Beta}(\alpha_{\text{split},p}, \beta_{\text{split},p})$ is generated, and the point at which the fibre splits is given by

$$\underline{p}_{\text{split}} = \varphi(\underline{\omega}, -l_1 + u_{\text{split}} l_F). \tag{5.39}$$

Fibres are independent of the choice of positive direction of orientation; to enforce

symmetry in the proposal distribution we choose $\alpha_{\text{split},p} = \beta_{\text{split},p}$.

Two new *end points* $\underline{e}_1, \underline{e}_2$ are proposed, both independently perturbed from $\underline{p}_{\text{split}}$. They are drawn from a bivariate normal distribution with covariance matrix $\sigma^2_{\text{split}}\mathbf{I}_2$, centred at $\underline{p}_{\text{split}}$. The proposed fibres are

$$F_1 = F(\underline{e}_1, l_F u_{\text{split}}, 0) \tag{5.40}$$

$$F_2 = F(\underline{e}_2, 0, l_F(1 - u_{\text{split}})). \tag{5.41}$$

Reference points $\underline{\omega}_1$ and $\underline{\omega}_2$ are sampled for the two new fibres: random numbers $u_{1,1}$ and $u_{2,1}$ are generated from a Beta$(\alpha_{\text{split},\omega}, \beta_{\text{split},\omega})$ distribution and the proposed reference points are

$$\underline{\omega}_1 := \varphi(\underline{e}_1, -(1 - u_{1,1})u_{\text{split}} l_F) \tag{5.42}$$

$$\underline{\omega}_2 := \varphi(\underline{e}_2, u_{2,1}(1 - u_{\text{split}})l_F). \tag{5.43}$$

The values of $\underline{\omega}_1, \underline{\omega}_2$ and fibres F_1, F_2 determine the proposed lengths $l_{1,1}, l_{1,2}, l_{2,1}, l_{2,2}$. We choose $\alpha_{\text{split},\omega} = \beta_{\text{split},\omega}$ for symmetry. If either $F_1 \not\subset W$ or $F_2 \not\subset W$ the proposed split move is rejected.

Finally, auxiliary variables X_i and p_i are drawn from $Q_{\text{aux}}(p_i', X_i' | \mathbf{F}', y_i, \sigma^2_{\text{disp}}, Z_i' = 1)$ as defined in Equation (5.7) for all signal points y_i. Note that we do not reallocate points to noise or signal, hence $\mathbf{Z}$ remains fixed.

In summary, the changes in variables are proposed as follows:

$$k \mapsto k + 1 \tag{5.44}$$

$$u_{\text{split}} \sim \text{Beta}(\alpha_{\text{split},p}, \beta_{\text{split},p}) \tag{5.45}$$

$$u_{1,1}, u_{2,1} \sim \text{Beta}(\alpha_{\text{split},\omega}, \beta_{\text{split},\omega}) \tag{5.46}$$

$$l_{1,1} := u_{1,1} u_{\text{split}} l_F \tag{5.47}$$

$$l_{1,2} := (1 - u_{1,1}) u_{\text{split}} l_F \tag{5.48}$$

$$l_{2,1} := u_{2,1}(1 - u_{\text{split}}) l_F \tag{5.49}$$

$$l_{2,2} := (1 - u_{2,1})(1 - u_{\text{split}}) l_F \tag{5.50}$$

$$\underline{p}_{\text{split}} := \varphi(\underline{\omega}, -l_1 + u_{\text{split}} l_F) \tag{5.51}$$

$$\underline{e}_1 \sim \text{MVN}(\underline{p}_{\text{split}}, \sigma^2_{\text{split}} \mathbf{I}_2) \tag{5.52}$$

$$\underline{e}_2 \sim \text{MVN}(\underline{p}_{\text{split}}, \sigma^2_{\text{split}} \mathbf{I}_2) \tag{5.53}$$

$$F_1 := F(\underline{e}_1, u_{\text{split}} l_F, 0) \tag{5.54}$$

$$F_2 := F(\underline{e}_2, 0, (1 - u_{\text{split}}) l_F) \tag{5.55}$$

$$\underline{\omega}_1 := \varphi(\underline{e}_1, -(1 - u_{1,1}) u_{\text{split}} l_F) \tag{5.56}$$

$$\underline{\omega}_2 := \varphi(\underline{e}_2, u_{2,1}(1 - u_{\text{split}}) l_F) \tag{5.57}$$

$$\mathbf{X} \mapsto \mathbf{X}' \tag{5.58}$$

$$\mathbf{p} \mapsto \mathbf{p}'. \tag{5.59}$$

Proposal Density for Join Move

The join move is simply the reverse of a split move. From an initial state with fibres F_1, F_2 a new fibre F is proposed with the same associated lengths and reference points as the corresponding fibres in the split move.

For an initial state containing k fibres, there are $4k(k-1)$ possible pairs of end points of different fibres. Most of these pairs are too far apart to result from a split so we identify a set of plausible pairs of end points, and choose a pair from this set. This reduces the number of rejected proposals. If the distance between a pair of fibre end-points is small enough that they could have arisen from a split move with non-negligible probability, they are considered a possible *join-pair*. Explicitly, a threshold t_d on the maximum distance between end points is chosen, and all pairs closer than this threshold are added to the set of possible join-pairs.

The implication is that a proposed split move will be rejected if the end points $\underline{e}_1$ and $\underline{e}_2$ are further apart than t_d.

This approximation results in a small error in the normalising constants of the proposal distribution of $\underline{e}_1, \underline{e}_2$ and $\underline{p}_{\text{split}}$. However, by choosing a sufficiently large value of t_d, we can make this error arbitrarily small. For example, with a variance parameter of $\sigma^2_{\text{split}} = 1$ and a threshold of $t_d = 3$, the error in the normalising constant, approximated by 1, is around 0.0018. In order to control the low error, the threshold must be increased proportionally to σ_{split}.

From the list of possible join-pairs, one is selected uniformly at random, represented

by a pair of end points $\underline{e}_1$ and $\underline{e}_2$. The split point $\underline{p}_{\text{split}}$ is drawn from a bivariate normal distribution centred at the Euclidean mean of these end points.

The proposed fibre is $F := F(\underline{p}_{\text{split}}, l_{1,F}, l_{2,F})$, and the new reference point is $\underline{\omega} = \varphi(\underline{p}_{\text{split}}, -(l_{1,F}) + v(l_{1,F} + l_{2,F}))$ where $v \sim \text{Beta}(\alpha_{\text{split},\omega}, \beta_{\text{split},\omega})$. Auxiliary variables X_i and p_i are updated from the proposal density $Q_{\text{aux}}(p_i, X_i | \mathbf{F}, y_i, \sigma^2_{\text{disp}}, Z_i = 1)$, given in Section 5.2.3. Noise and signal allocations are not resampled.

The proposed state is summarised as follows

$$\underline{p}_{\text{split}} \sim \text{MVN}((\underline{e}_1 + \underline{e}_2)/2, \frac{\sigma^2_{\text{split}}}{4}\mathbf{I}_2) \tag{5.60}$$

$$v \sim \text{Beta}(\alpha_{\text{join},\omega}, \beta_{\text{join},\omega}) \tag{5.61}$$

$$l_1 := v(l_{1,F} + l_{2,F}) \tag{5.62}$$

$$l_2 := (1 - v)(l_{1,F} + l_{2,F}) \tag{5.63}$$

$$F := F(\underline{p}_{\text{split}}, l_{1,F}, l_{2,F}) \tag{5.64}$$

$$\underline{\omega} = \varphi(\underline{p}_{\text{split}}, -(l_{1,F}) + v(l_{1,F} + l_{2,F})) \tag{5.65}$$

$$\mathbf{X}' \mapsto \mathbf{X} \tag{5.66}$$

$$\mathbf{p}' \mapsto \mathbf{p}. \tag{5.67}$$

The two states, initial and proposed, for the reversible pair of moves, splits and joins, are depicted graphically in Figure 5.3.5.

Acceptance Probability of a Split Move

Join and split moves, like birth and death moves, change the dimension of the model. Details of the calculation of acceptance probabilities for such moves are outlined in Green [1995].

The dimensionality of each move is equal to the sum of the total dimension of the variables in the initial state and the total dimension of the random variables generated in the proposal step. The split and join moves must have equal dimensionality for reversibility to hold.

Excluding the variables that do not change from the initial state and the proposed

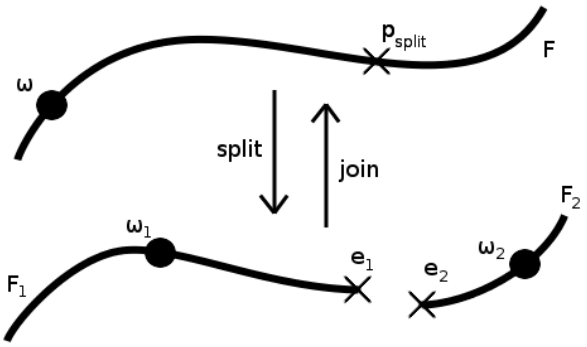


Figure 5.1: The two states involved in a split/join move, with end points and reference points indicated.

state, we have:

$$(\boldsymbol{\theta}; \mathbf{u}) \xrightarrow{\text{split}} (\boldsymbol{\theta}'; \mathbf{u}') \tag{5.68}$$

$$\begin{aligned}(\underline{\omega}, l_1, l_2, \mathbf{p}, \mathbf{X}; u_{\text{split}}, u_{1,1}, u_{2,1}, \underline{e}_1, \underline{e}_2, \mathbf{u}_{\mathbf{p}',\mathbf{X}'}) \\ \xrightarrow{\text{split}} (\underline{\omega}_1, \underline{\omega}_2, l_{1,1}, l_{1,2}, l_{2,1}, l_{2,2}, \mathbf{p}, {}' \mathbf{X}'; \underline{p}_{\text{split}}, v, \mathbf{v}_{\mathbf{p},\mathbf{X}})\end{aligned} \tag{5.69}$$

where $\mathbf{v}_{\mathbf{p},\mathbf{X}}$ and $\mathbf{u}_{\mathbf{p}',\mathbf{X}'}$ are the random numbers generated during the proposals of $\mathbf{p}, \mathbf{X}$ and $\mathbf{p}', \mathbf{X}'$ respectively.

The acceptance probability of the split move is

$$\min\left\{1, \frac{\pi(\mathbf{F}', (l_{1,1}, l_{1,2}), (l_{2,1}, l_{2,2}), \boldsymbol{\omega}', k+1, \mathbf{X}', \mathbf{p}') Q_{\text{join}}(\underline{p}_{\text{split}}, v)}{\pi(\mathbf{F}, (l_1, l_2), \boldsymbol{\omega}, k, \mathbf{X}, \mathbf{p}) Q_{\text{split}}(u_{\text{split}}, u_{1,1}, u_{2,1}, \underline{e}_1, \underline{e}_2)} \left| \frac{\partial(\boldsymbol{\theta}'; \mathbf{u}')}{\partial(\boldsymbol{\theta}; \mathbf{u})} \right| \right\} \tag{5.70}$$

where $Q_{\text{split}}(\cdot)$ and $Q_{\text{join}}(\cdot)$ are proposal densities and $\left| \frac{\partial(\boldsymbol{\theta}'; \mathbf{u}')}{\partial(\boldsymbol{\theta}; \mathbf{u})} \right|$ is the Jacobian determinant of the bijective function mapping $(\boldsymbol{\theta}; \mathbf{u})$ to $(\boldsymbol{\theta}'; \mathbf{u}')$.

Let n_e be the number of possible end point configurations for a join move in the proposed resultant state. The split move acceptance probability is the minimum of

1 and

$$\frac{\pi(\mathbf{F}', (l_{1,1}, l_{1,2}), (l_{2,1}, l_{2,2}), \boldsymbol{\omega}', k+1, \mathbf{X}', \mathbf{p}')}{\pi(\mathbf{F}, (l_1, l_2), \boldsymbol{\omega}, k, \mathbf{X}, \mathbf{p})} \frac{Q(\underline{p}_{\text{split}}, v)}{Q(u_{\text{split}}, u_{1,1}, u_{2,1}, \underline{e}_1, \underline{e}_2)}$$

$$\times \frac{Q_{\text{aux}}(p_i, X_i | \mathbf{F}, y_i, \sigma^2_{\text{disp}}, Z_i = 1)}{Q_{\text{aux}}(p'_i, X'_i | \mathbf{F}', y_i, \sigma^2_{\text{disp}}, Z'_i = 1)} \left| \frac{\partial(\boldsymbol{\theta}'; \mathbf{u}')}{\partial(\boldsymbol{\theta}; \mathbf{u})} \right|$$

$$= \frac{\kappa}{k+1} \frac{1}{|W|\lambda^2} \frac{D(\mathbf{q}'(\mathbf{p}'), \alpha_{\text{Dir}})}{D(\mathbf{q}(\mathbf{p}), \alpha_{\text{Dir}})} \frac{r_J}{r_S} \frac{k}{n_e} \prod_{i \text{ s.t. } Z_i = 1} \left(\frac{l_{X'_i, F}}{l_{X_i, F}} \frac{N(y_i, \mathbf{F}, \sigma^2_{\text{disp}}, Z_i = 1)}{N(y_i, \mathbf{F}', \sigma^2_{\text{disp}}, Z'_i = 1)} \right)$$

$$\times 8\pi\sigma^2_{\text{split}} \exp\left(\frac{(\underline{e}_1 - \underline{p}_{\text{split}})^2 + (\underline{e}_1 - \underline{p}_{\text{split}})^2 - 4((\underline{e}_1 + \underline{e}_2)/2 - \underline{p}_{\text{split}})^2}{2\sigma^2_{\text{split}}} \right)$$

$$\times \frac{f_{\text{Beta}}(v, \alpha_{\text{join},\omega}, \beta_{\text{join},\omega})}{f_{\text{Beta}}(u_{\text{split}}, \alpha_{\text{split},p}, \beta_{\text{split},p}) f_{\text{Beta}}(u_{1,1}, \alpha_{\text{split},\omega}, \beta_{\text{split},\omega}) f_{\text{Beta}}(u_{2,1}, \alpha_{\text{split},\omega}, \beta_{\text{split},\omega})}$$

$$\times \left| \frac{\partial(\boldsymbol{\theta}'; \mathbf{u}')}{\partial(\boldsymbol{\theta}; \mathbf{u})} \right| \tag{5.71}$$

where r_J, r_S are respectively the rates of proposals of join and split moves, $f_{\text{Beta}}(\cdot, \alpha, \beta)$ is the Beta density function with parameters α and β, and $N(y_i, \mathbf{F}, \sigma^2_{\text{disp}}, Z_i = 1)$ is defined in Equation (5.8). The variable v, used in the reverse (join) move to sample the location of start point $\underline{\omega}$ along the initial fibre, is calculated directly from $\underline{\omega}$ and F. Similarly p_{split} is determined from fibre F and u_{split}.

The Jacobian matrix under this parametrisation,

$$\left| \frac{\partial(\underline{\omega}_1, \underline{\omega}_2, l_{1,1}, l_{1,2}, l_{2,1}, l_{2,2}, \mathbf{p}', \mathbf{X}'; \underline{p}_{\text{split}}, v, \mathbf{v}'_{\mathbf{p},\mathbf{X}})}{\partial(\underline{\omega}, l_1, l_2, \mathbf{p}, \mathbf{X}; u_{\text{split}}, u_{1,1}, u_{2,1}, \underline{e}_1, \underline{e}_2, \mathbf{u}_{\mathbf{p}',\mathbf{X}'})} \right| \tag{5.72}$$

is very large. It is not written in full here. Most partial derivatives are calculated directly from the split move summary in equation (5.44). The non-trivial partial derivatives are given below.

The following partial derivatives involve φ:

$$\frac{\partial(\underline{\omega}_j)}{\partial(l_1)}, \frac{\partial(\underline{\omega}_j)}{\partial(l_2)}, \frac{\partial(\underline{\omega}_j)}{\partial(u_{\text{split}})} \text{ and } \frac{\partial(\underline{\omega}_j)}{\partial(u_{\cdot,1})} \tag{5.73}$$

for $j = 1, 2$. These terms disappear in the Jacobian determinant so no further evaluation is necessary.

The partial derivatives

$$\frac{\partial(\underline{\omega}_1)}{\partial(\underline{e}_1)} = \frac{\partial\varphi(\underline{e}_1, u_{\text{split}} l_F u_{1,1})}{\partial(\underline{e}_1)} \tag{5.74}$$

and

$$\frac{\partial(\underline{\omega}_2)}{\partial(\underline{e}_2)} = \frac{\partial\varphi(\underline{e}_2, (1-u_{\text{split}})(l_F)u_{2,1})}{\partial(\underline{e}_2)} \tag{5.75}$$

do not cancel in the calculation of the Jacobian determinant. For a field of constant orientation, a translation in $\underline{e}_1$ (respectively $\underline{e}_2$) will directly result in a translation the location of $\underline{\omega}_1$ (respectively $\underline{\omega}_2$), and these partial derivatives are identity matrices. However, for a general field of orientations the effect of $\underline{e}_1$ on $\underline{\omega}_1$ is not so trivially calculated.

By definition, $\varphi(\underline{x}, l)$ is the solution $f = \varphi$ to the differential equation

$$\frac{\partial f(\underline{x}, l)}{\partial l} = \upsilon_{FO}(f(\underline{x}, l)) \tag{5.76}$$

for $\underline{x} \in W$, $f : W \times (-\infty, \infty) \mapsto W$. Hence $\varphi(\underline{x}, l)$ can be approximated using the Euler method: $\varphi(\underline{x}, l) \approx \underline{x}_n$ where $\underline{x}_n$ is iteratively defined by

$$\underline{x}_0 := \underline{x} \tag{5.77}$$

$$\underline{x}_i = g(\underline{x}_{i-1}) := \underline{x}_{i-1} + \frac{l}{n}\upsilon_{FO}(\underline{x}_{i-1}) \qquad \text{for } i \in \{1, ..., n\}$$

for some choice of $n > 0$. By the chain rule, the Jacobian $\frac{\partial\varphi(\underline{x},l)}{\partial\underline{x}}$ is

$$J_{\underline{x}}(\varphi) \approx J_{\underline{x}_{n-1}}(g) J_{\underline{x}_{n-2}}(g) \dots J_{\underline{x}_0}(g) \tag{5.78}$$

where $J_{\underline{x}_i}(g)$ is the Jacobian $\frac{\partial g(\underline{x},l)}{\partial\underline{x}}$ evaluated at x_i. The Jacobian matrices evaluate to

$$J_{\underline{x}_i}(g) = \mathbf{I}_2 + \frac{l}{n}\frac{\partial}{\partial\underline{x}}\upsilon_{FO}(\underline{x}_i). \tag{5.79}$$

The derivative of the field of orientations can be estimated directly.

Partial derivatives $\frac{\partial(\underline{\omega}_2)}{\partial(\underline{e}_2)}$ and $\frac{\partial(p_{\text{split}})}{\partial(\underline{\omega})}$ are evaluated similarly.

In practice, for most smooth fields of orientations, the contribution of $\frac{\partial(\underline{\omega}_1)}{\partial(\underline{e}_1)}$ is negligible, in spite of the additional computational time required for its evaluation. For this reason we choose to omit these terms in the implementation of the algorithm.

The other non-trivial terms are the partial derivatives involving $\mathbf{v}_{\mathbf{p},\mathbf{X}}$, $\mathbf{u}_{\mathbf{p}',\mathbf{X}'}$, $\mathbf{p}$ and $\mathbf{X}$. The remaining variables in $\boldsymbol{\theta}'$ and $\mathbf{u}'$ are independent of $\mathbf{v}_{\mathbf{p},\mathbf{X}}$, $\mathbf{u}_{\mathbf{p}',\mathbf{X}'}$, $\mathbf{p}$ and $\mathbf{X}$ and, as variables X_i and p_i are drawn identically and independently of X_j, p_j for $i \neq j$ it suffices to calculate $\frac{\partial(X_i,p_i,u_{\mathbf{p}',\mathbf{X}',i})}{\partial(X_i',p_i',v_{\mathbf{p},\mathbf{X},i})}$. The terms $v_{\mathbf{p},\mathbf{X},i}$ and $, u_{\mathbf{p}',\mathbf{X}',i}$ are the random numbers used in the proposals of p_i, X_i and p_i', X_i' respectively.

The partial derivatives $\frac{\partial(X_i,p_i)}{\partial(X_i',p_i')}$ and $\frac{\partial(u_{\mathbf{p}',\mathbf{X}',i})}{\partial(v_{\mathbf{p},\mathbf{X},i})}$ are both zero, and

$$\frac{\partial(X_i,p_i)}{\partial(v_{\mathbf{p},\mathbf{X},i})} = \frac{\partial(X_i',p_i')}{\partial(u_{\mathbf{p}',\mathbf{X}',i})} \tag{5.80}$$

so the absolute value of the determinant of $\frac{\partial(X_i,p_i,u_{\mathbf{p}',\mathbf{X}',i})}{\partial(X_i',p_i',v_{\mathbf{p},\mathbf{X},i})}$ is 1.

In the calculation of the determinant of this Jacobian matrix, most terms cancel, leaving

$$\left|\frac{\partial(\boldsymbol{\theta}'\mathbf{u}')}{\partial(\boldsymbol{\theta};\mathbf{u})}\right| = \left| l_1 l_2 \frac{\partial\underline{\omega}_1}{\partial\underline{e}_1}\frac{\partial\underline{\omega}_2}{\partial\underline{e}_2}\frac{\partial\underline{p}_{\text{split}}}{\partial\underline{\omega}}\right|. \tag{5.81}$$

Following the notation of previous sections, the acceptance probability in Equation (5.71) is the minimum of 1 and

$$\begin{aligned}
&\frac{\kappa}{k+1}\frac{1}{|W|\lambda^2}\prod_{i \text{ s.t. } Z_i=1}\frac{l_{X_i',T}}{l_{X_i,T}}\frac{D(\mathbf{q}'(\mathbf{p}'),\alpha_{\text{Dir}})}{D(\mathbf{q}(\mathbf{p}),\alpha_{\text{Dir}})}\frac{r_J}{r_S}\frac{k}{n_e}\frac{P(y_i|\mathbf{F}',\sigma^2_{\text{disp}},Z_i'=1)}{P(y_i|\mathbf{F},\sigma^2_{\text{disp}},Z_i=1)}\\
&\times 8\pi\sigma^2_{\text{split}}\exp\left(\frac{(\underline{e}_1-\underline{p}_{\text{split}})^2+(\underline{e}_1-\underline{p}_{\text{split}})^2-4((\underline{e}_1+\underline{e}_2)/2-\underline{p}_{\text{split}})^2}{2\sigma^2_{\text{split}}}\right)\\
&\times\frac{f_{\text{Beta}}(v,\alpha_{\text{join},\omega},\beta_{\text{join},\omega})}{f_{\text{Beta}}(u_{\text{split}},\alpha_{\text{split},p},\beta_{\text{split},p})f_{\text{Beta}}(u_{1,1},\alpha_{\text{split},\omega},\beta_{\text{split},\omega})f_{\text{Beta}}(u_{2,1},\alpha_{\text{split},\omega},\beta_{\text{split},\omega})}\\
&\times l_1 l_2\left|\frac{\partial\underline{\omega}_1}{\partial\underline{e}_1}\frac{\partial\underline{\omega}_2}{\partial\underline{e}_2}\frac{\partial\underline{p}_{\text{split}}}{\partial\underline{\omega}}\right|.
\end{aligned} \tag{5.82}$$

Acceptance Probability of a Join Move

The acceptance probability of the join move from F_1, F_2 to F, is the minimum of 1 and the reciprocal of the split acceptance probability in Equation (5.82). If the variables are re-labeled so that the starting state is the fibre set $\mathbf{F}$ containing k fibres, with lengths $\mathbf{l}$, etc. and the proposed state consists of fibre set $\mathbf{F}'$ with $k-1$

fibres, the acceptance probability is the minimum of 1 and

$$\frac{k}{\kappa}|W|\lambda^2 \prod_{i \text{ s.t. } Z_i=1} \frac{l_{X_i',T}}{l_{X_i,T}} \frac{D(\mathbf{q}'(\mathbf{p}'),\alpha_{\text{Dir}})}{D(\mathbf{q}(\mathbf{p}),\alpha_{\text{Dir}})} \frac{r_S}{r_J} \frac{n_e}{k-1} \frac{P(y_i|\mathbf{F}',\sigma^2_{\text{disp}},Z_i'=1)}{P(y_i|\mathbf{F},\sigma^2_{\text{disp}},Z_i=1)}$$

$$\times \frac{1}{8\pi\sigma^2_{\text{split}}} \exp\left(\frac{-(\underline{e}_1-\underline{p}_{\text{split}})^2-(\underline{e}_1-\underline{p}_{\text{split}})^2+4((\underline{e}_1+\underline{e}_2)/2-\underline{p}_{\text{split}})^2}{2\sigma^2_{\text{split}}}\right)$$

$$\times \frac{f_{\text{Beta}}(u_{\text{split}},\alpha_{\text{split},p},\beta_{\text{split},p}) f_{\text{Beta}}(u_{1,1},\alpha_{\text{split},\omega},\beta_{\text{split},\omega}) f_{\text{Beta}}(u_{2,1},\alpha_{\text{split},\omega},\beta_{\text{split},\omega})}{f_{\text{Beta}}(v,\alpha_{\text{join},\omega},\beta_{\text{join},\omega})}$$

$$\times l_1 l_2 \left|\frac{\partial\underline{\omega}_1}{\partial\underline{e}_1}\frac{\partial\underline{\omega}_2}{\partial\underline{e}_2}\frac{\partial\underline{p}_{\text{split}}}{\partial\underline{\omega}}\right|^{-1}. \tag{5.83}$$

The acceptance rates of split and join moves can be adjusted by varying σ_{split}. Trans-dimensional moves, such as the split and join, generally have low acceptance probabilities (compared with fixed-dimension Metropolis Hastings) so we aim to choose the value for σ_{split} that maximises the number of accepted splits and joins. Tests indicate that a values of $\sigma_{\text{split}} \in [\sigma_{\text{disp}}/2, \sigma_{\text{disp}}]$ perform well. For example, acceptance rates of around 9.5% were recorded when the MCMC was applied to the fingerprint data (see Section 6.2.4), where $\sigma_{\text{disp}} = 2$ and $\sigma_{\text{split}} = 1$. See Brooks et al. [2003] for a further discussion on acceptance rates of trans-dimensional moves.

5.3.6 Updating the Reference Point of a Fibre

The final move is the update of the reference point of a fibre.

We have already introduced a move which resamples the lengths of fibres while fixing the reference point ω_j, therefore it is useful to include a complementary move that resamples ω_j without changing the fibre. No variables other than ω_j and l_j need to be updated as the total length and location of fibre F_j are not altered. Reference point ω_j is sampled from its full conditional distribution.

Having chosen a fibre F_j, reference point ω_j is proposed by sampling a random number u from the conditional density $P(l_{j,1}/l_{j,T}|l_{j,T})$. As $l_{j,1}$ and $l_{j,2}$ are both exponentially distributed with equal mean, this density has a Beta$(1,1)$ density.

Variables are updates as follows

$$u \sim \text{Beta}(1,1)$$

$$l_{j,1} \mapsto l'_{j,1} = u l_{j,T}$$

$$l_{j,2} \mapsto l'_{j,2} = (1-u) l_{j,T}$$

$$\omega_j \mapsto \omega'_j = \varphi(\omega_j, u(l_{j,1} + l_{j,2}) - l_{j,1}).$$

As ω_j is drawn from the full conditional distribution the acceptance probability is 1.

5.4 Implementation of Additional Moves

The implemented algorithm runs on a continuous time scale. Additional moves are proposed at random times governed by fixed rates. Relative values for the rates of each move are suggested in Table 5.1. The units for the rate of an event are 'per unit of algorithm time'. The BDMCMC is then allowed to run for a large number of time units and samples are taken at random times (at some fixed rate). In section 5.6 an appropriate burn-in and sampling interval are suggested. Of course the relationship of algorithm time to actual processing time depends on hardware and implementation details.

Death rates are recalculated following any accepted move.

5.5 Algorithm Validation: A Simple Data-Independent Model

The implementation of a BDMCMC with all the additional moves suggested here will lead to a computer program that is long and intricate, and therefore prone to errors. To fully understand the BDMCMC and check it is drawing samples from the posterior distribution as expected, it is essential to validate the algorithm. This is achieved by limiting the number of moves with non-zero rates, and by using simplified data sets with predictable output.

It is particularly interesting to monitor how the reversible pairs of moves: birth-death and split-join, interact. Here, a simplified model for the fibre process is designed and the dependency of the data on the fibres is removed, leading to easily interpreted death rates and acceptance rates.

A substantial simplification of the model is obtained by omitting the dependency of the data on the fibres. The posterior distribution of the fibres is then independent of the data and the likelihood is constant. Effectively the BDMCMC produces samples from the prior distribution of fibres. Auxiliary variables $\mathbf{Z}$, $\mathbf{X}$ and $\mathbf{p}$ are now superfluous, as are the signal probabilities $\boldsymbol{\varepsilon}$ and the number of data points m.

The only parameters remaining in the model are the fibres $\mathbf{F}$, lengths $\mathbf{l}$, reference points $\boldsymbol{\omega}$, field of orientations Υ_{FO}, and associated hyperparameters.

The model is further simplified by choosing $W = \mathbb{R}^2$, and ignoring the spatial distribution of the fibres, i.e. each fibre is described only in terms of its length, without a reference point.

The new model consists of a vector of lengths $l_{1,T}, ..., l_{k,T}$, where $k \sim \text{Poisson}(\kappa)$ and $l_{j,T} \sim \text{Gamma}(n, 1/\lambda)$. Taking $n = 2$ allows a direct comparison with the full model, but it is also interesting to see the effect of using a general length distribution.

The equivalent birth density for this model is the inclusion of a new 'fibre' $l_{k+1,T}$ with length proposed from the prior distribution ($l_{k+1,T} \sim \text{Gamma}(n, 1/\lambda)$). Births are proposed at a rate β. In a state consisting of a set of k fibres, the total death rate satisfying detailed balance is simply $\beta k/\kappa$. Each fibre is equally likely to die with individual death rate β/κ, compare with Equation (5.6). At equilibrium, the number of fibres k is Poisson distributed with mean κ and each length $l_{j,T}$ is $\text{Gamma}(n, 1/\lambda)$ distributed.

Split and join moves are now introduced. A split is proposed by first sampling a random vector (j_1, j_2, u) where j_1 is a random integer from $\{1, ..., k\}$, j_2 is a random integer from $\{1, ..., k+1\}$, and $u \sim \text{Beta}(\alpha_u, \beta_u)$. A bijective mapping of the current state and this random vector determines the resulting family of fibres: fibre F_{j_1} is split into two fibres, F'_{j_1} and $F'_{j,2}$ of lengths $ul_{j_1,T}$ and $(1-u)l_{j_1,T}$ respectively. The second fibre is then inserted into position j_2 in the vector of fibres with the index of subsequent fibres is incremented:

$$\mathbf{F}' = \{F'_1, ..., F'_{k+1}\} = \{F_1, ..., F_{j_2-1}, F'_{j_2}, F_{j_2}, F_{j_2+1}, ..., F_{j_1-1}, F'_{j_1}, F_{j_1+1}, ..., F_k\}. \tag{5.84}$$

The join move is the reverse of the split move: a random integer i_1 is randomly sampled from $\text{Uniform}(\{1, ..., k+1\})$ and then a second, i_2, is drawn from $\text{Uniform}(\{1, ..., k+1\} \backslash \{i_1\})$. The i_1-th fibre is replaced by a fibre of length $l_{i_1,T} + l_{i_2,T}$ and the i_2-th fibre is removed. Fibre indices greater than i_2 are decremented.

For a set $\mathbf{F}$ of k fibres splitting to a set $\mathbf{F}'$ of $k+1$ fibres the Metropolis-Hastings acceptance probability, corresponding to the full model acceptance probability in

Equation (5.82), is

$$\begin{aligned}
&\min\left\{1, \frac{\pi(\mathbf{F}')}{\pi(\mathbf{F})}\frac{Q_{\text{join}}(\mathbf{F})}{Q_{\text{split}}(\mathbf{F}')}\right\}\left|\frac{\partial(\theta', \mathbf{u}')}{\partial(\theta, \mathbf{u})}\right| \\
&= \min\left\{1, \frac{r_J\kappa}{r_S k}\frac{(u(1-u)l_{j_1,T})^{n-1}}{\Gamma(n)\lambda^n}\frac{1}{f_{\text{Beta}}(u, \alpha_u, \beta_u)}l_{j_1,T}\right\} \\
&= \min\left\{1, \frac{r_J\kappa}{r_S k}\left(\frac{l_{j_1,T}}{\lambda}\right)^n\frac{(u(1-u))^{n-1}}{f_{\text{Beta}}(u, \alpha_u, \beta_u)}\frac{1}{\Gamma(n)}\right\} \qquad (5.85)
\end{aligned}$$

where r_J and r_S are the proposal rates of joins and splits, $\Gamma(\cdot)$ is the gamma function, and $f_{\text{Beta}}(\cdot, \alpha_u, \beta_u)$ is the Beta density function with parameters α_u and β_u. Here $\left|\frac{\partial(\theta', \mathbf{u}')}{\partial(\theta, \mathbf{u})}\right|$ is the Jacobian determinant corresponding to the Jacobian determinant of the full model, given in Equation (5.70).

In the case $n = 2$, if $\alpha_u = \beta_u = 1$ the acceptance probabilities become

$$\min\left\{1, \frac{r_J\kappa}{r_S k}\left(\frac{l_{j_1,T}}{\lambda}\right)^2\right\} \qquad (5.86)$$

for a split move, and

$$\min\left\{1, \frac{r_S k}{r_J\kappa}\left(\frac{\lambda}{l_{j_1,T}}\right)^2\right\} \qquad (5.87)$$

for a join move.

Note the similarity between the acceptance probability of a join move and the individual death rate of a fibre $\beta k/\kappa$ when $\beta = r_J = r_S = 1$. As expected, if $k > \kappa$ then a death or a join is preferred over a birth or a death, but if the length of a typical fibre is larger than λ, a split move is more likely to be accepted than a join.

Calculation of the acceptance probabilities for a simplified model such as this, is both informative and highly beneficial when testing computer code for errors. This is particularly true here, as our model contains a large number of variables and can lead to complicated acceptance probabilities. This analysis could potentially be extended to fibre processes which lie entirely within a window W, providing a foundation for further investigations into the full implications of edge effects.

5.6 Output Analysis

The samples collected from the BDMCMC are summarised by estimating properties of the marginal distributions of variables, for example: the number of fibres, the number of noise points, or the lengths of fibres. As a large range of models is ex-

plored, it is usually more informative to consider conditional empirical distributions, for example the number of noise points conditional on there being k fibres.

It would be interesting to investigate properties of the individual fibres. In general, however, it is not possible to identify corresponding fibres across samples as typically fibres will die and be replaced between samples. This is known as the label-switching problem, briefly discussed in Section 2.3.3 and also in Stephens [2000b], with possible solutions suggested in Richardson and Green [1997], Celeux et al. [1995] and Stephens [1997]. In principle, fibre categorisation could be achieved by constructing a distance function between fibres, and distinguishing two fibres if the distance between them is sufficiently large. It may then be possible to classify fibres across samples, and measure how each fibre varies as the chain progresses. Such a distance function would also identify a variance statistic for fibres. This notion of comparing two curves is not a trivial problem, and careful consideration of the type of variation of interest (e.g. curvature, length, location) is necessary.

In this thesis, output analysis is generally restricted to summary statistics of marginal and conditional empirical distributions of the model variables. However, through an empirical estimate of the signal point density, we can visually discern general characteristics of the fibres.

5.6.1 Burn-In Time

A heuristic lower bound on a suitable burn-in time is motivated by considering aspects of the priors derived after inspection of the data (e.g. $\sigma_{\text{disp}}, \lambda, \kappa$ - see Section 6.1.1), and estimating the number of fibre births that must occur before a fibre has been created around each potential fibre cluster. We approximate the lower bound by considering only the number of fibre births required for a fibre to be born around the *smallest* fibre cluster.

A lower bound on half the length of the shortest fibre cluster is derived from the 10% quantile of an exponentially distributed random variable of rate κ/λ. Then the probability a point chosen at random from W lies in a region corresponding to an actual fibre of this length (up to $2\sigma_{\text{disp}}$ from the fibre) is approximated by

$$\frac{8\lambda \log(10/9)\sigma_{\text{disp}}}{\kappa|W|}. \tag{5.88}$$

It follows that, with probability 0.99, a fibre will be proposed in the region corre-

sponding to the shortest fibre within the first

$$\frac{\log(0.01)}{\log\left(1-\frac{8\lambda\log(10/9)\sigma_{\text{disp}}}{\kappa|W|}\right)} \tag{5.89}$$

births. Hence we choose a minimum burn-in time of

$$T_{\text{burn}} = \max\left\{1500, \frac{\log(0.01)}{\beta\log\left(1-\frac{8\lambda\log(10/9)\sigma_{\text{disp}}}{\kappa|W|}\right)}\right\}. \tag{5.90}$$

The lower bound 1500 is chosen to ensure the burn-in time remains substantial, and yet the algorithm can run to 1500 algorithmic time units in reasonable time on most modern computers (the specifications of the computers used are detailed in the following chapter). Note that this is a minimum burn-in time as inspection of output variables may suggest a longer burn-in is required.

5.6.2 Thinning/Sampling Rate

Outputs of different variables are sampled at random times at some constant rate. The rate of this sampling (effectively the reciprocal of the thinning of the Monte Carlo process) is chosen such that there is a low probability that any of the fibres remain unchanged between samples. The inclusion of the extra moves designed to improve mixing also helps to decrease the thinning required. The thinning is chosen approximately proportional to the number of fibres (estimated based on aspects of priors derived following inspection of the data).

5.6.3 Number of Iterations

Following the burn-in time, the BDMCMC is run until sufficiently many samples have been taken. Clearly, a greater number of samples will reduce the error in the estimation of summary statistics such as quantiles or means. However, even with a high specification machine (full details are given in Chapter 6), $30,000$ units of BDMCMC algorithm time, run on a 200 point dataset clustered around 5 fibres, on a window of size $|W| \approx 1000$, may take up to 2 days to run as the algorithm has not been expertly optimised. There is certainly scope for a more efficient implementation of the algorithm; Section 6.1.2 lists some steps taken towards a more efficient implementation, further improvements are suggested in Section 8.2.4. Experiments indicate that a minimum of around $150 - 200$ samples are required for analysis of the posterior distribution.

5.6.4 Convergence Diagnostics

Convergence is assessed by considering variables such as the number of fibres k, the number of noise points, or the log of the posterior (calculable up to normalising constant) and using Geweke's spectral density diagnostic, see Geweke [1991]. Convergence of a sequence of n samples is rejected if the mean value of the variable in the first $n/10$ samples is not sufficiently similar to the mean value over the last $n/2$ samples.

An alternative approach to identifying whether the BDMCMC has converged is to run multiple chains from dispersed starting points and use Gelman and Rubin's statistic (Gelman and Rubin, 1992), which compares variation within and between chains. Brooks and Giudici [1998] and Castelloe and Zimmerman [2002] show how this can be extended to the variable dimension problem by considering the variation between models as well as between chains. Over-dispersed initial states can be constructed by varying $\boldsymbol{\varepsilon}$ and subsequently the field of orientations, or simply by starting one chain with a large number of fibres and a second chain with one fibre or none at all. While neither form of Gelman and Rubin's statistic have been used formally, multiple simulations from varying initial states are run to check the chains mix sufficiently.

5.7 Conclusions

In this chapter we have provided the full details of a BDMCMC algorithm, including a number additional moves which are designed to improve the chain's mixing properties. The BDMCMC allows us to draw samples from the posterior distribution of fibres given an instance of a fibre-process generated point process. Examples of samples of fibre sets, along with summary statistics of the posterior distribution for some example data sets are presented in the following chapter.

Chapter 6

Examples: Earthquakes, Fingerprints (and briefly Galaxies)

This chapter illustrates the application of the BDMCMC described in Chapter 5 to four different data sets. The first data set is our own simulated example, which is followed by a simulated data set from Stanford and Raftery [2000], and finally two real data sets: earthquake epicentres and fingerprint pores. Preliminary results on two 3-dimensional data sets are also presented.

First, we briefly discuss some implementation considerations including how hyperparameters should be chosen and how to improve the efficiency of the implemented algorithm.

6.1 Implementation Considerations

6.1.1 Hyperparameters

As a rough guideline, hyperparameters can be chosen as follows.

The prior mean number of fibres κ and the prior mean length of fibres λ can be estimated from any prior knowledge or expectations of the fibres. The deviation of points from fibres σ^2_{disp} can be estimated using prior knowledge of fibre widths (for example, the average width of a ridge in the fingerprint data) and the approximation that 95% of points should lie within $2\sigma_{\text{disp}}$ of the centre of a fibre. The density of signal points per unit length of fibre η can be estimated similarly.

Parameters for the field of orientation, h_{FO} and σ_{FO}, should be chosen to ensure the field of orientations is smooth. These can be estimated by evaluating the field of orientations for different selections of h_{FO}, σ_{FO} and choosing from this set. If the proportion of noise points is approximately known then the hyperparameters α_{Signal} and β_{Signal} can be suitably estimated, however we suggest choosing the parameters such that $\alpha_{\text{Signal}}, \beta_{\text{Signal}} > 1$ to ensure good mixing properties of the Markov chain Monte Carlo sampling algorithm. Otherwise the noise hyperparameters can be set equal to 1 indicating no prior knowledge.

Alternatively, if little prior information is known about the nature of the latent curvilinear structure, then it is feasible to extend the empirical Bayes step to include the estimation of further prior parameters.

In the examples of this chapter, hyperparameters were all chosen as suggested here.

6.1.2 Efficiency and Run-Times

All four 2-dimensional examples were run on the cluster owned by the Statistics Department in the University of Warwick using a Dell PowerEdge 1950 server with a 3.16GHz Intel Xeon Harpertown (X5460) processor, 2GB fully-buffered RAM. The algorithm was implemented in Octave version 3.2.4. The total run-times on the cluster ranged from 39.6 hours for the fingerprint pore data (40000 units of algorithm time) to 83.6 hours for the earthquake data set (60000 units of algorithm time).

As the BDMCMC explores a multi-dimensional model over a large window, a computationally slow algorithm is anticipated. For the implementation of the algorithm, the following considerations were made, and assisted in reducing the overall run-time of the BDMCMC:

- Fibres are approximated by piecewise-linear curves.
- Variables that require calculation, such as distances, are saved in order to avoid recalculation.
- The additional moves implemented vastly improve the chain's mixing properties, reducing the length of chain required to draw a reasonable number of samples with low auto-correlation.
- Death rates are updated only after the acceptance of a move that may cause the death rates to change.

- Updates which involve resampling the field of orientations are proposed infrequently, as the calculations involved are slow.
- Fast-Fourier Transforms are used in the calculation of the field of orientations.

Further optimisation is possible, both in the implementation of the algorithm, and through carefully constructing proposal densities (or birth densities) which are less likely to be rejected. The latter is discussed in Section 8.2.4.

Implementing the algorithm in a more powerful language, such as C++, would likely result in much faster run times although it is not yet possible to quantify the expected increase in speed. A further improvement would likely be found by implementing the code on parallel processors, however due to the limitations of the current version of Octave, the benefits of true parallelism have not yet been explored.

More immediate improvements could be made by fully vectorising all calculations and investigating ideas from morphology (Vincent, 1999) to reduce the time required for distance calculations. Alternatively choices of data structures may also improve algorithmic speeds; here we used a combination of matrices and cell arrays.

6.1.3 Other Considerations

The field of orientations is estimated over a square grid of points, each one unit length from its four nearest neighbours. The total size of this grid is given by the dimensions of the window W.

Burn-in times were chosen by consideration of output graphs and the heuristic lower bound calculated in Section 5.6.1. Geweke's spectral density diagnostic (Geweke, 1991) was applied to both the number of fibres k, the total number of signal points, $\sum_i Z_i$, and also to summary statistics such as the total length of the fibres, in order to test for a lack of convergence in the chain.

6.2 Two-Dimensional Examples

In each of the following examples births are proposed at unit rate, as are the following moves: moving a fibre, adjusting the lengths of a fibre and updating signal-noise allocation. Split and join moves are proposed at twice the unit rate to account for the typically low acceptance probabilities. Signal probability (ε) updates are proposed at a rate of 0.1 per unit of time. The rate at which samples are taken varies for different data sets.

6.2.1 Simulated Example

Figure 6.1(a) shows a simulated data set of clusters of points around three fibres with additional background noise. The signal data was generated by perturbing 117 points that were randomly distributed along 3 fibres. A further 63 points of homogeneous background noise were superimposed on the signal points.

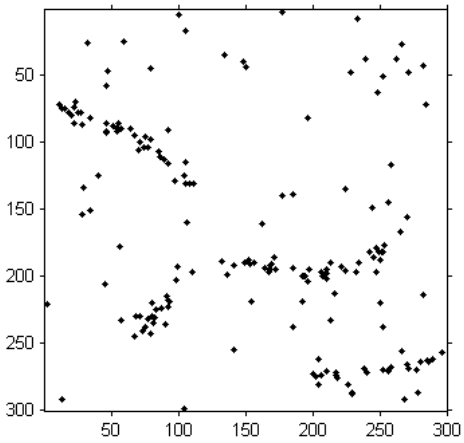
On the above timescale the BDMCMC was run for 40,000 units of algorithm time, the first 21,500 of which were discarded. Samples were taken at a rate of 0.025 per unit of algorithm time. The initial state was a set of $\kappa = 3$ fibres drawn from the prior distribution of fibres. Other hyperparameters were chosen as follows: dispersion parameter $\sigma_{\text{disp}} = 4$; signal probability hyperparameters $\alpha_{\text{signal}} = 2$ and $\beta_{\text{signal}} = 1$; density parameter $\eta = 0.3$; mean half-fibre length $\lambda = 55$; and the Dirichlet parameter $\alpha_{\text{Dir}} = 1$.

Figure 6.1(b) shows a typical sample from the output. It is evident that the main clusters of signal points have been correctly identified, however the long cluster is split into 3 fibres clusters rather than one. This subdivision of the long cluster is also discernible in Figure 6.1(c), which depicts an empirical estimate of the density of the signal point process estimated from the samples of fibres. As mentioned at the end of Section 4.5, the orientation field can suffer from a bias in areas of high curvature. This leads to a tendency for the fitted fibres to be less curved than the original fibres (due to the smoothness of the field of orientations). A consequence is that a curvilinear cluster may often be approximated by several shorter fibres rather than a single long fibre. The bias should be borne in mind when considering the real data examples.

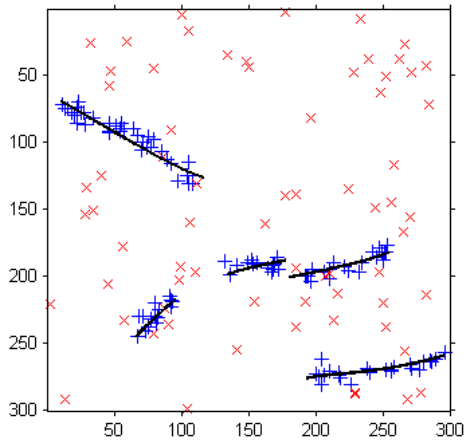
Figure 6.1(d) displays the clustering of points estimated by assigning any two points to the same cluster if they are associated with the same fibre in more than 50% of samples. In cluster analysis terms, we use an agglomerative hierarchical algorithm to determine the clusters, calculating the distance between a pair of clusters by the nearest neighbour method. The measure of similarity between two points is the proportion of samples with both points in the same fibre cluster. This is a very basic approach to data clustering, and is used only as a summary of the clustering across the samples. The book by Everitt et al. [2011] provides a thorough background in cluster analysis including a number of alternative approaches.

Table 6.1 gives the posterior probabilities of the number of fibres and the means and highest posterior density intervals of a variety of properties conditional on the number of fibres.

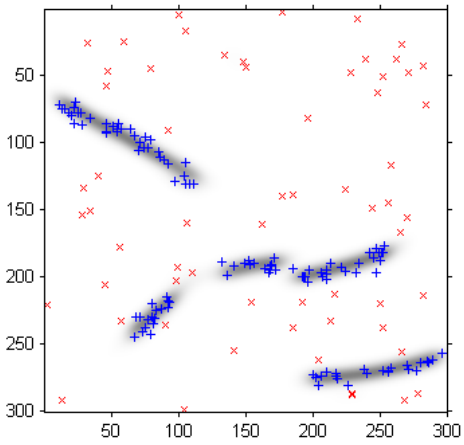
The number of fibres is simply a count of the fibres present in each sample; in this



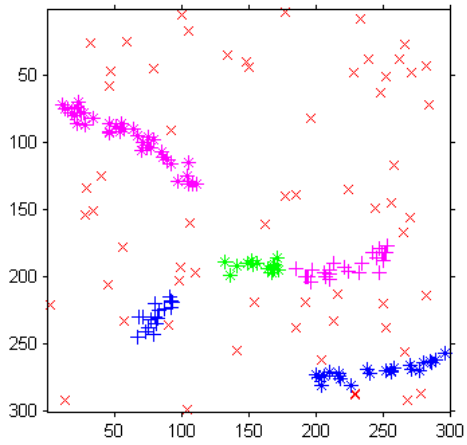
(a) Simulated data clustered around 3 fibres.



(b) A random sample from the BDMCMC output. Fibres are represented by curves, pluses indicate points allocated to signal and crosses indicate points allocated to noise in this sample.



(c) Estimate of the clustering of the signal points - different clusters are distinguished by varying symbol, crosses indicate noise. Estimated by considering how often pairs of points are associated with the same fibre across a number of samples.



(d) Estimate of the clustering of the signal points - different symbols indicate different clusters, crosses indicate noise. Estimated by considering how often pairs of points are associated with the same fibre across a number of samples.

Figure 6.1: Simulated Example.

Posterior Probabilities for Number of Fibres				
Number of Fibres	5	6	7	8
Posterior Probability	0.20	0.44	0.33	0.03

Other Properties Conditioned on the Number of Fibres				
	Number of Fibres	Posterior Mean	50% HPD Interval	95% HPD Interval
Number of Noise Points	5	65.32	[63,65]	[62,70]
	6	67.60	[65,68]	[62,72]
	7	67.07	[66,69]	[60,72]
95th Percentile of the Distances from Signal Points to Fibres	5	9.94	[10.07,10.83]	[9.17,10.83]
	6	9.95	[9.60,10.44]	[8.82,11.15]
	7	9.78	[9.38,10.27]	[8.60,11.18]
Total Length of Fibres	5	404.71	[379,403]	[365,435]
	6	391.51	[376,395]	[362,438]
	7	412.24	[402,429]	[379,453]

Table 6.1: Results for Simulated Example: First sub-table gives posterior probabilities on the number of fibres, while the second gives posterior means and 50% and 95% HPD (highest posterior density) intervals for a selection of properties of the posterior distribution conditional on the number of fibres. The simulated data consists of 117 signal points and 63 noise points on a 300×300 window, based on a family of three fibres and using a dispersion parameter of $\sigma_{\text{disp}} = 4$ and with prior mean probability that a point is noise equal to 0.33. Posterior probabilities only given if non-zero to rounding error.

example we expect it to be around 3, however as the long cluster is split into a number of components, the posterior number of fibres is somewhat higher.

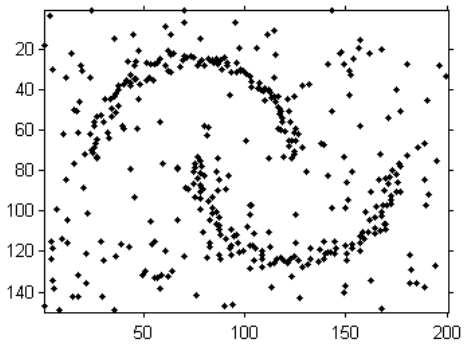
We might expect the number of points assigned to the noise component to be negatively correlated with the number of fibres, as with more fibres comes a greater chance of there being a fibre close to a given point and hence a greater chance that it is a signal point. However, it appears that the number of noise points is not closely correlated with the number of fibres, supporting the hypothesis that increases in the number of fibres arise from splitting a single fibre into two. In the construction of this simulated data set 63 noise points were superimposed on 117 signal points. The evidence in Table 6.1 is that our approach models the allocation of points to noise and signal well.

We take the 95th percentile of the distances of signal points to anchor points for each sample. This summarises the dispersion of points from the fibres and is comparable to $2.45\sigma_{\text{disp}}$; the dispersion parameter σ_{disp} is set to 4 in this example. We note that this value of 9.8 tends to lie comfortably within the HPD intervals, confirming that the model for the relationship between points and fibres fits the data well in this instance.

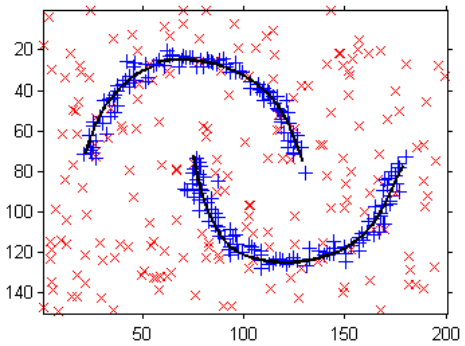
6.2.2 Stanford and Raftery's Simulated Example

A simulated data set used in Stanford and Raftery [2000] is shown in Figure 6.2(a). We include it here to facilitate comparison with the methods proposed by Stanford and Raftery. The data consist of 200 signal points and 200 noise points and is based on a family of two fibres each of length 157. The original data set, consisting of points over a $[-1.5, 2.5] \times [-1.5, 1.5]$ window, were scaled and translated to lie in a 200×150 window.

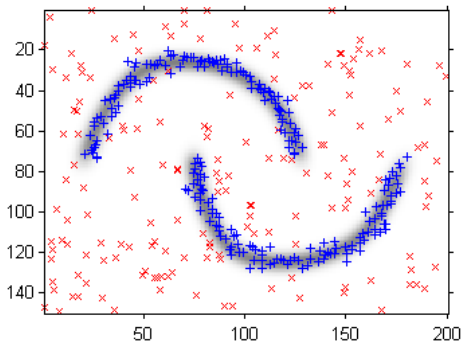
The BDMCMC was run for 60,000 units of algorithm time, the first 30,000 of which



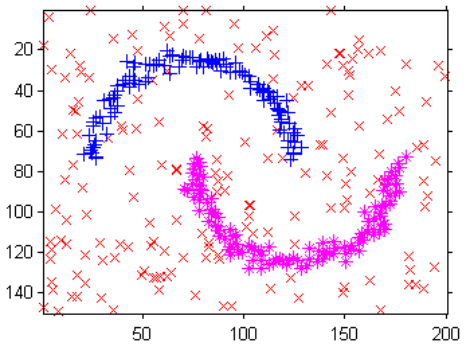
(a) Simulated data.



(b) A random sample from the BDMCMC output. Fibres are represented by curves, pluses indicate points allocated to signal and crosses indicate points allocated to noise in this sample.



(c) Estimate of the density of signal points found by smoothing a series of samples of fibres (darker areas indicate higher densities). Pluses indicate points allocated to signal and crosses indicate points allocated to noise in at least 50% of samples. The size of points representing the data has been reduced to enhance the clarity of the density estimate.



(d) Estimate of the clustering of the signal points - different clusters are distinguished by varying symbol, crosses indicate noise. Estimated by considering how often pairs of points are associated with the same fibre across a number of samples.

Figure 6.2: Simulated Example from Stanford and Raftery [2000].

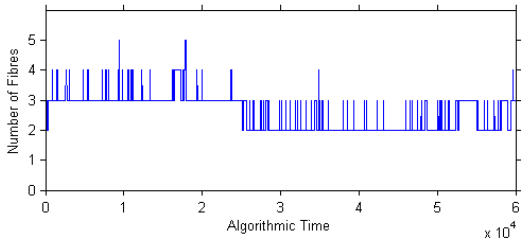


Figure 6.3: Trace plot of the number of fibres against algorithmic time. Following consideration of this plot it was decided that the first 30,000 algorithmic seconds should be discarded.

were discarded. This rather long burn-in was chosen following consideration of the trace plot of the number of fibres, see Figure 6.3. This plot may suggest that insufficient mixing is occurring as the chain appears to traverse a region of the posterior distribution with 3-4 fibres, and then after 25,000 algorithmic seconds, move to a region favouring 2-3 fibres. However, from consideration of fibre samples and also of the posterior density, we have concluded that the chain, if run for longer, would not return to the original region favouring 3-4 fibres. Trace plots of other statistics such as the total length of the fibres and the total number of noise points show no evidence of a lack of convergence or poor mixing.

Samples were taken at a rate of 0.033 per unit of time. The initial state was a randomly sampled set of $\kappa = 2$ fibres. Other hyperparameters were chosen as follows: dispersion parameter $\sigma_{\text{disp}} = 3$; signal probability hyperparameters $\alpha_{\text{signal}} = 1$ and $\beta_{\text{signal}} = 1$; density parameter $\eta = 0.64$; mean half-fibre length $\lambda = 78.5$; and the Dirichlet parameter $\alpha_{\text{Dir}} = 1$.

Figures 6.2(b) to 6.2(d) show that our model fits the data very well. The two fibres in the sample in Figure 6.2(b) compare favourably with the principal curves fitted in Stanford and Raftery [2000].

Naturally, noise points that lie near a fibre will frequently be associated with the signal component during the course of the BDMCMC. Most points sufficiently close to a fibre are associated to signal in at least 50% of samples as is evident in Figure 6.2(c). However, in an individual sample a random subset of these points will be noise (see Figure 6.2(b)), reflecting how noise is included in the model - as the superposition of a homogeneous Poisson process. This is in contrast to the work of Stanford and Raftery [2000] where the emphasis is on fitting a principal curve to the points, for this reason they use the approximation that all points that belong to a dense cluster are signal points.

As in the previous example, we estimate properties of the posterior distribution. Posterior means and highest posterior density intervals are given in Table 6.2.

Posterior Probabilities for Number of Fibres		
Number of Fibres	2	3
Posterior Probability	0.78	0.22

Other Properties Conditioned on the Number of Fibres				
	Number of Fibres	Posterior Mean	50% HPD Interval	95% HPD Interval
Number of Noise Points	2	195.81	[194,202]	[181,209]
	3	195.09	[191,199]	[180,206]
95th Percentile of Point to Fibre Distances	2	7.46	[7.11,7.65]	[6.60,8.47]
	3	7.48	[7.08,7.53]	[6.80,8.22]
Total Length of Fibres	2	317.44	[312,320]	[306,331]
	3	319.02	[309,318]	[305,337]

Table 6.2: Results for Stanford and Raftery's Simulated Example: First sub-table gives posterior probabilities on the number of fibres, while the second gives posterior means and 50% and 95% HPD (highest posterior density) intervals for a selection of properties of the posterior distribution conditional on the number of fibres. The simulated data consists of 200 signal points and 200 noise points over a 200×150 window, and is based on a family of two fibres each of length 157. The dispersion parameter σ_{disp} is set to 3 and the prior mean probability that a point is noise is 0.5. Posterior probabilities only given if non-zero to rounding error.

In this example, more points are associated to signal than expected. This is partly due to the high intensity of noise points, and also explained by a slight bias in the length of the fibres. The posterior statistics on the lengths of the fibres arguably suggest that the extension of fibres beyond their known length (of 157) is supported by the high intensity of noise points. This extrapolation is sometimes beneficial, particularly for fibre reconstruction in areas of missing data. Here the extrapolation is less desirable as it suggests there is evidence for fibres in the background noise.

The extrapolation of fibres into less dense regions of points can be reduced by choosing a higher Dirichlet parameter α_{Dir} for the distribution of anchor points along the fibres. This decreases the posterior density of fibres lying through point clusters of non-constant intensity. The drawback of increasing α_{Dir} is that proposed moves are more frequently rejected. This is because (A) no proposals account for the distribution of anchor points along fibres, and (B) a large value of α_{Dir} leads to a multimodal anchor point distribution with most of the probability weighted around the modes. Hence, the proposal of a state with low posterior density is more likely.

6.2.3 Application: Earthquakes on the New Madrid Fault-line

The epicentres of earthquakes along seismic faults are a good example of point data clustered around a system of fibres with additional background noise. Here the fibres are the unknown fault-lines. Stanford and Raftery [2000] consider the structure of the data set of earthquakes around the New Madrid fault line in central USA. We use data on earthquakes in the New Madrid region between 1st Jan 2006 and 3rd Aug 2008 (inclusive) taken from the full earthquake catalogue found at CERI (Center for Earthquake Research and Information), `http://www.ceri.memphis.edu/seismic/catalogs/cat_nm.html`.

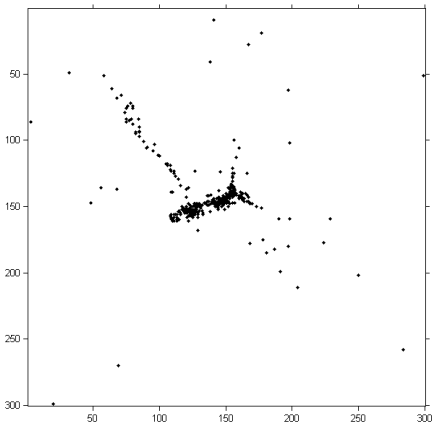
The BDMCMC was run for 40,000 units of algorithm time, the first 10,000 of which were discarded. The burn in time was chosen following consideration of trace plots of various statistics of the output samples. In particular, the trace plot of the total length of fibres, shown in Figure 6.5, suggested that a burn in of $10,000$ was sufficient. Samples were taken at a rate of 0.0167 per unit of time. The initial state was a randomly sampled set of $\kappa = 4$ fibres. Other hyperparameters were chosen as follows: dispersion parameter $\sigma_{\text{disp}} = 2$; signal probability hyperparameters $\alpha_{\text{signal}} = 4$ and $\beta_{\text{signal}} = 1$; density parameter $\eta = 1.06$; mean half-fibre length $\lambda = 30$; and the Dirichlet parameter $\alpha_{\text{Dir}} = 1$.

Our method has the advantage over Stanford and Raftery [2000], in that it does not try to over fit the fibres where there is less data. Rather it uses information from surrounding data to extrapolate fibres as required.

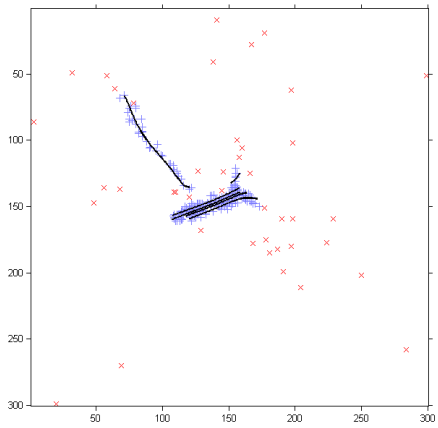
Table 6.3 gives some numerical properties of the posterior distribution of fibres.

A limitation of our model is that every fibre is assumed to share a number of properties. In particular the displacement of points from fibres (effectively the width of influence of a fibre) and the intensity of signal points per unit length of fibre are assumed to be constant, independent of the fibre. These assumptions are not reasonable for this data as the 'thickness' and density of points varies considerably. This is apparent in Figure 6.4(b) where the central dense cluster is described by multiple parallel fibres. The dispersion parameter σ_{disp} was chosen by considering the apparent 'width' of the longer thinner fibre, hence points around the shorter, wider fibre effectively increase the 95th percentile of the point to fibre distances, as given in Table 6.3. The solution to this problem is to extend the model to allow different hyperparameters for each fibre.

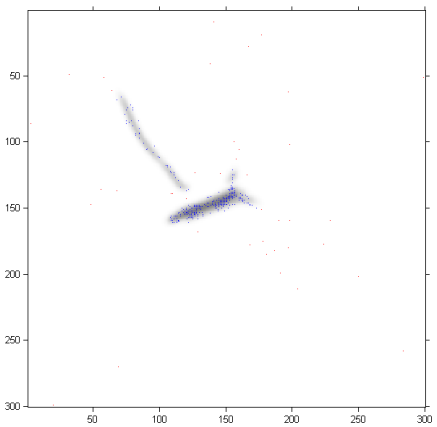
While multiple fibres in the central cluster is a common feature in samples from this BDMCMC, Figure 6.4(d) indicates that the agglomerative clustering algorithm identifies the points as arising from the same cluster.



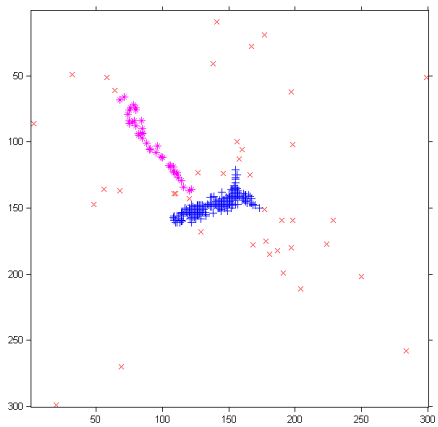
(a) Earthquake data



(b) A random sample from the BDMCMC output. Fibres are represented by curves, pluses indicate points allocated to signal and crosses indicate points allocated to noise in this sample.



(c) Estimate of the density of signal points found by smoothing a series of samples of fibres (darker areas indicate higher densities). Pluses indicate points allocated to signal in at least 50% of samples. The size of points representing the data has been reduced to enhance the clarity of the density estimate.



(d) Estimate of the clustering of the signal points - different clusters are distinguished by varying symbol, crosses indicate noise. Estimated by considering how often pairs of points are associated with the same fibre across a number of samples.

Figure 6.4: New Madrid Fault Earthquake Data.

Posterior Probabilities for Number of Fibres			
Number of Fibres	6	7	8
Posterior Probability	0.52	0.37	0.10

Other Properties Conditioned on the Number of Fibres				
	Number of Fibres	Posterior Mean	50% HPD Interval	95% HPD Interval
Number of Noise Points	6	43.07	[41,43]	[40,47]
	7	42.25	[40,42]	[36,46]
	8	41.80	[39,41]	[39,45]
95th Percentile of the Distances from Signal Points to Fibres	6	4.94	[4.84,5.10]	[4.56,5.30]
	7	4.90	[4.75,4.98]	[4.55,5.27]
	8	4.97	[4.80,5.06]	[4.61,5.78]
Total Length of Fibres	6	293.56	[275,302]	[264,332]
	7	292.82	[292,315]	[258,321]
	8	296.15	[292,305]	[266,315]

Table 6.3: Results for Earthquake Data: First sub-table gives posterior probabilities on the number of fibres, while the second gives posterior means and 50% and 95% HPD (highest posterior density) intervals for a selection of properties of the posterior distribution conditional on the number of fibres. The data are all the recorded earthquakes in the New Madrid region between 1st Jan 2006 and 3rd Aug 2008; the data were acquired from the CERI (Center for Earthquake Research and Information) found at `http://www.ceri.memphis.edu/seismic/catalogs/cat_nm.html`. In total there are 317 points in a 300×300 window, the dispersion parameter σ_{disp} is set to 2 and the prior mean probability that a point is noise is 0.2. Posterior probabilities only given if non-zero to rounding error.

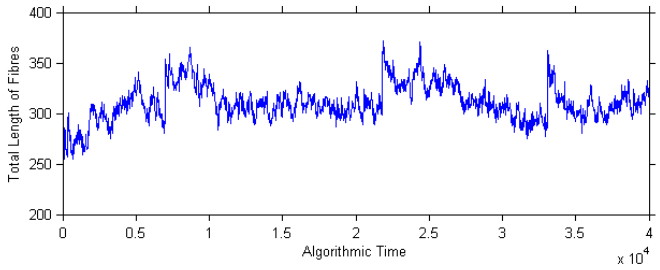


Figure 6.5: Trace plot of the total length of fibres in samples from the BDMCMC. Following consideration of this plot it was decided that the first 10,000 algorithmic seconds should be discarded.

Interestingly the total length of fibres does not appear to be positively correlated to the number of fibres, suggesting that the additional fibres arise from splitting a fibre into multiple parts while preserving the total fibre length.

6.2.4 Application: Fingerprint Data

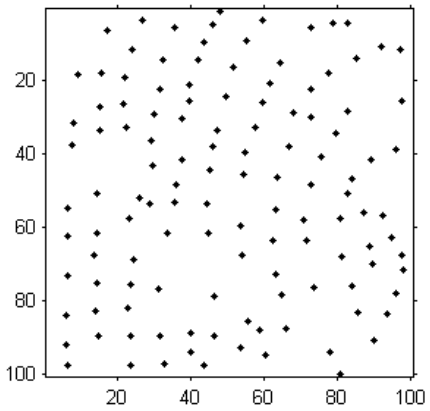
The second application we consider is that of pores aligned along ridge lines in fingerprints. Fingerprint pore data is considered in some depth in Su et al. [2008] and Su [2009].

We use a portion of the data set extracted from fingerprint a002-05 from the *NIST (National Institute of Standards and Technology) Special Database 30* (Watson, 2001). The procedure for extracting the pore locations from the fingerprint image is described in Su et al. [2008].

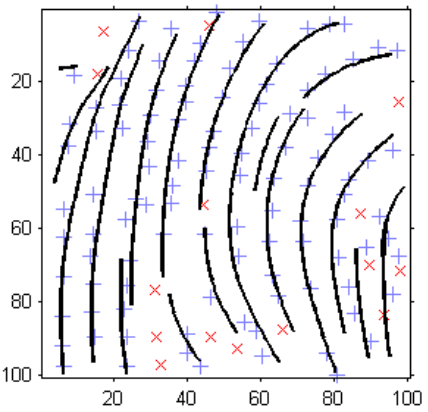
The BDMCMC was run for 40,000 units of algorithm time, the first 8,000 of which were discarded. Samples were taken at a rate of 0.007 per unit of time. The initial state was a randomly sampled set of $\kappa = 10$ fibres. Other hyperparameters were chosen as follows: dispersion parameter $\sigma_{\text{disp}} = 1.5$; signal probability hyperparameters $\alpha_{\text{signal}} = 15$ and $\beta_{\text{signal}} = 1$; density parameter $\eta = 0.13$; mean half-fibre length $\lambda = 45$; and the Dirichlet parameter $\alpha_{\text{Dir}} = 1.5$.

The fingerprint pore data will typically cause nearest neighbour clustering methods to breakdown. This is because, whilst the filamentary structure of the point pattern is clear when viewing the global picture, it is not so apparent on a small scale. This phenomenon is partly due to the apparent inter-ridge alignment of points (from left to right in Figure 6.2(a)). By way of contrast, our field of orientations model takes any information available on a small scale and uses it across the window thanks to the smoothing step in the field of orientations estimation.

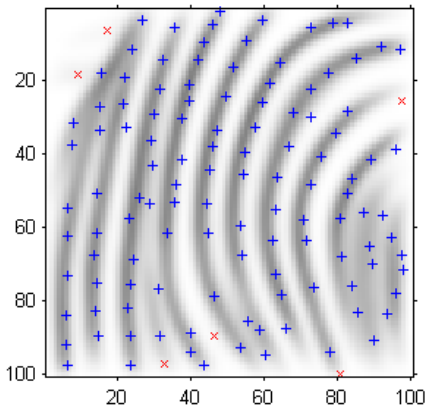
As Figure 6.6 shows, our model succeeds in fitting many of the fibres (or fingerprint



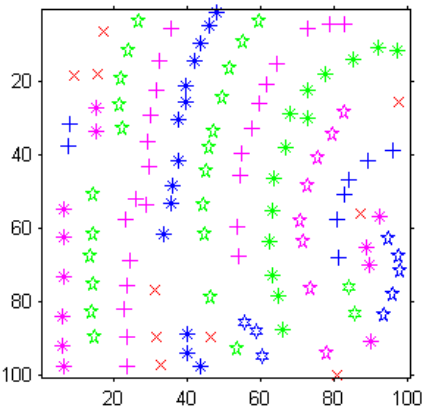
(a) Pore data



(b) A random sample from the BDMCMC output. Fibres are represented by curves, pluses indicate points allocated to signal and crosses indicate points allocated to noise in this sample.



(c) Estimate of the density of signal points found by smoothing a series of samples of fibres (darker areas indicate higher densities).



(d) Estimate of the clustering of the signal points - different clusters are distinguished by varying symbol, crosses indicate noise. Estimated by considering how often pairs of points are associated with the same fibre across a number of samples.

Figure 6.6: Pores from portion of fingerprint a002-05 from the *NIST Special Database 30* (Watson, 2001).

Posterior Probabilities for Number of Fibres							
Number of Fibres	17	18	19	20	21	22	23
Posterior Probability	0.07	0.10	0.13	0.16	0.15	0.13	0.09

Other Properties Conditioned on the Number of Fibres				
	Number of Fibres k	Posterior Mean	50% HPD Interval	95% HPD Interval
Number of Noise Points	19	17.00	[14,18]	[11,22]
	20	15.86	[15,17]	[12,19]
	21	15.25	[15,17]	[10,19]
	22	15.95	[13,15]	[9,21]
95th Percentile of the Distances from Signal Points to Fibres	19	3.54	[3.42,3.63]	[3.07,4.14]
	20	3.65	[3.50,3.56]	[3.50,3.93]
	21	3.70	[3.71,3.95]	[3.34,4.10]
	22	3.64	[3.56,3.84]	[3.16,3.93]
Total Length of Fibres	19	989.65	[891,974]	[891,1079]
	20	966.14	[945,969]	[913,1043]
	21	969.81	[983,1027]	[840,1062]
	22	985.95	[956,1008]	[927,1089]

Table 6.4: Fingerprint Pore Data Set: Posterior means and 50% and 95% credible intervals of a selection of properties of the posterior distribution conditional on the number of fibres. The data was extracted from a portion of fingerprint a002-05 from the *NIST (National Institute of Standards and Technology) Special Database 30* (Watson, 2001). It consists of 123 points on a 100×100 window. A dispersion parameter of $\sigma_{\text{disp}} = 1.5$ is used, and the mean prior probability a point is noise is 0.091. Posterior probabilities only given if non-zero to rounding error.

ridges) to the pore data. Figure 6.6(c) indicates a few areas of doubt in the fibre locations where the shading is lighter near the edges of the window, showing that fibre samples were more dispersed.

This data set is an ideal candidate for the reconstruction of missing data. We work under the assumptions that pores lie at fairly regularly intervals along ridges, but some are not identified during the pore extraction process. Our method uses information from nearby ridges to complete fibres where data is missing. In this example the missing data is particularly evident in the region below the centre of the window. Knowledge of the posterior distribution of fibres could lead to a 'filling in the gaps' approach to reconstructing the missing pore data.

Table 6.4 gives some numerical properties of the posterior distribution of fibres.

6.3 Three-Dimensional Examples

This section presents some preliminary results for 3-dimensional data. Most of the model extends easily to 3 dimensions; naturally the distribution of the displacement of data point y_i from anchor point p_i is now a 3-dimensional multivariate normal distribution. However, the increase in the total run-time of a BDMCMC on a 3-dimensional data set is quite substantial. The actual execution time taken to calculate the field of orientations is of order $|W|$. Calculation of the field of orientation over a 100×100 grid in 2 dimensions takes takes approximately the same time as the same calculation over a $20 \times 20 \times 25$ cube in 3 dimensions. It is therefore necessary to reduce the resolution of the grid over which we approximate of the field of orientations.

As Baddeley et al. [1993] discuss, edge effects are increasingly apparent in higher dimensions. Indeed, for a 2-dimensional square of area A, the ratio of edges to area is $4/A^{1/2}$, while for the equivalent cube of volume V, the ratio of edges (or faces) to volume is $6/V^{1/3}$, indicating that a larger increase in volume (than in area) is required to reduce the relative edge effects.

There are also visualisation issues with 3-dimensional data. Both on screen, and in print, any image of the data is a projection onto a 2-dimensional surface, therefore depth is imperceptible. The problem is heightened when viewing a point pattern, as it is more difficult for the brain to perceive the pattern's structure and therefore it cannot 'guess' the depth. On the screen the problem can be solved by creating a video or an interactive plot of points and fibres which allows 3-dimensional rotations. In the first simulated example, all figures include plots viewed from 3 slightly different angles to clarify the 3-dimensional structure. The data in the second example are sparser, and as the window of observation is a thin slice of the total volume additional plots are not deemed necessary.

In the following two examples births are proposed at unit rate. Additionally, moves that adjust lengths or alter the location of a fibre by perturbing the reference point, are proposed at unit rate. No other moves have been implemented. Omission of these additional moves slows the mixing of the BDMCMC. The result of not including the signal probability $\boldsymbol{\epsilon}$ update is that samples are drawn from the posterior distribution of fibres conditional on a single estimate of the field of orientations as described in Section 4.4.1.

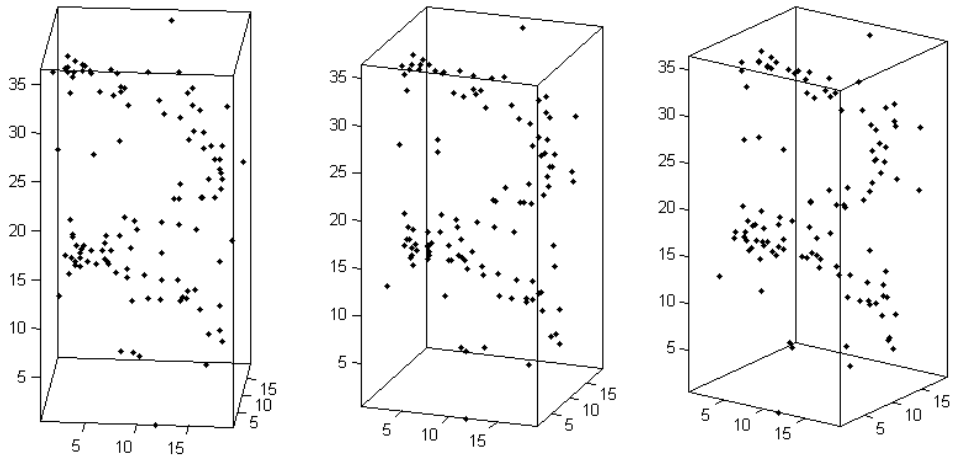


Figure 6.7: Simulated helix data, viewed from 3 different angles.

6.3.1 Simulated Example: Helix

The first 3-dimensional example is a simulated data set consisting of 97 signal points clustered around a circular helix of radius 7.5 and arc length 80.8. A further 20 noise points uniformly distributed across the $20 \times 20 \times 40$ window are superimposed on the signal points. The variance of the multivariate normal perturbation of the points from anchor points regularly spaced along the helix, is 1. Figure 6.7 shows this data. Plots for this data set include views from 3 different angles to clarify the structure of the 3-dimensional data.

The BDMCMC was run for 33,000 units of algorithm time, the first 10,000 of which were discarded. Samples were taken at a rate of 0.05 per unit of algorithm time. The initial state was a randomly sampled set of $\kappa = 1$ fibre. Other hyperparameters were chosen as follows: dispersion parameter $\sigma_{\text{disp}} = 1$; signal probability hyperparameters $\alpha_{\text{signal}} = 10$ and $\beta_{\text{signal}} = 1$; density parameter $\eta = 1.55$; mean half-fibre length $\lambda = 30$; and the Dirichlet parameter $\alpha_{\text{Dir}} = 1$.

A sample from the posterior distribution of fibres is displayed in Figure 6.8, different symbols indicate different clusters. It is clear that the model fits the data reasonably well as the fibres approximately lie along the helix . However, it is rather striking that there are 7 fibres, rather than 1. This segmentation of the helix into shorter fibres is also seen in Figure 6.9, which depicts an empirical estimate of the density of the signal point process. A single fibre along the helix would produce a higher posterior density than multiple short fibres, as the prior distribution for the number of fibres is Poisson with mean 1. This suggests that not implementing split and join moves has resulted in a BDMCMC with poor mixing properties.

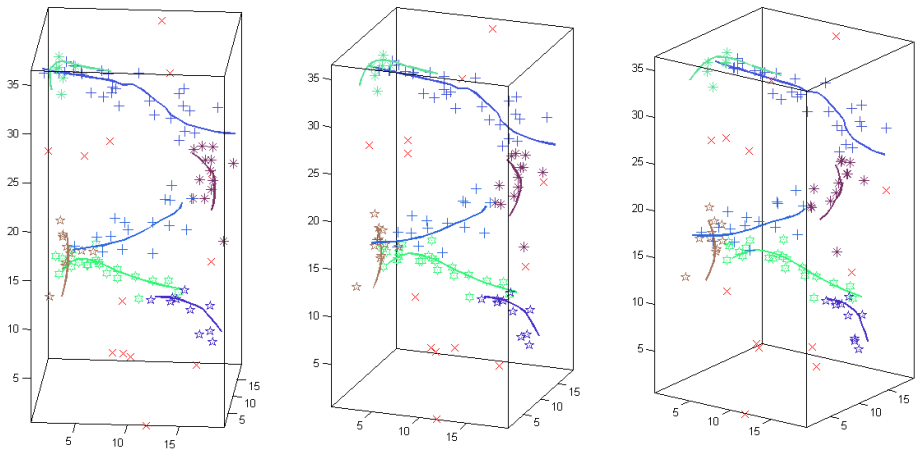


Figure 6.8: Three different views (the same angles as in Figure 6.7) of a sample from the posterior distribution of fibres. Curves represent fibres, and the clustering of points to fibres is highlighted by marking them with different symbols. Noise points in this sample are represented by crosses.

The samples are drawn from the conditional distribution of fibres given the field of orientations. However, the estimate of the orientation field does not consistently agree with the line tangent to the helix. This is exemplified by the short fibre located at a height of 15 in Figure 6.8; the fibre is nearly vertical in orientation - approximately perpendicular to the orientation of the true helix at that point. Integral curves are sensitive to errors in the orientation field estimate. The sensitivity is heightened by the relatively high curvature (rapidly varying orientation) of the field of orientations compared to the resolution of the grid over which the orientation field is evaluated, and the discretisation of the fibres into linear segments. Where it is not possible to sample a fibre, conditional on the field of orientations, that fits the data (produces a high likelihood) the Bayesian hierarchical model supports splitting the data into smaller fibre clusters.

Numerical statistics of samples from the BDMCMC are given in Table 6.5. A low prior mean for the probability that each point is noise is reflected by the low posterior mean number of noise points. Despite the segmentation of the cluster into short clusters, the posterior total length (an estimate of the arc length of the helix) is very similar to the known length, 80.8.

6.3.2 Application: Galaxies

The second 3-dimensional data set is the locations of galaxies as analysed in Stoica et al. [2007]. The original data set is mapped in the 2dF Galaxy Redshift Survey

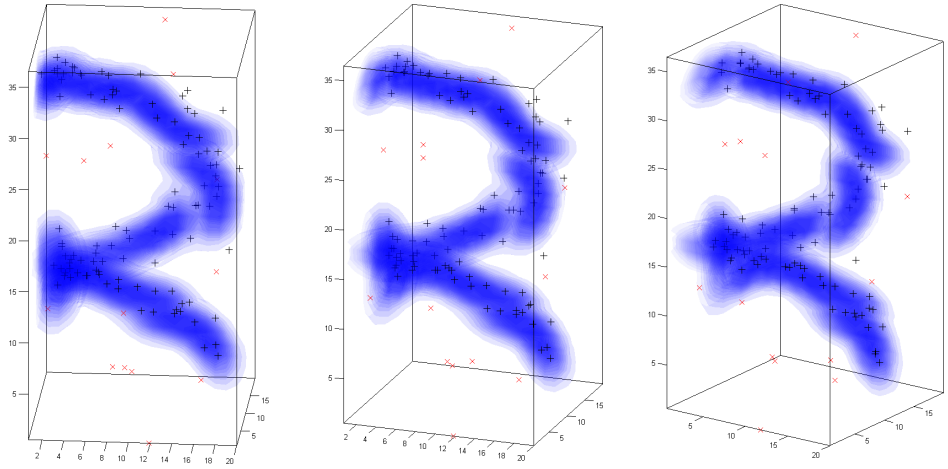


Figure 6.9: Empirical estimate of the signal point process density viewed from 3 angles. Points associated with noise in more than 50% of samples are indicated by $\times$, other points are denoted by $+$.

Posterior Probabilities for Number of Fibres	
Number of Fibres	7
Posterior Probability	0.99

Other Properties Conditioned on the Number of Fibres				
	Number of Fibres	Posterior Mean	50% HPD Interval	95% HPD Interval
Number of Noise Points	7	13.97	[13,14]	[13,15]
95th Percentile of the Distances from Signal Points to Fibres	7	2.56	[2.43,2.65]	[2.25,2.90]
Total Length of Fibres	7	88.66	[80,86]	[72,113]

Table 6.5: Results for Simulated Helix Data: First sub-table gives posterior probabilities on the number of fibres, while the second gives posterior means and 50% and 95% HPD (highest posterior density) intervals for a selection of properties of the posterior distribution conditional on the number of fibres. The data consists of 97 points perturbed by a multivariate normal distribution with variance 1, from a helix of length 80.8. Twenty noise points, uniformly distributed over the $20 \times 20 \times 40$ window, were superimposed on the signal point pattern. The dispersion parameter σ_{disp} is set to 1 and the prior mean probability that a point is noise is 0.091. Posterior probabilities only given if non-zero to rounding error.

(Colless et al., 2001), a 3-dimensional map of 221,000 galaxies. However, the analysis here is restricted to a subset (124 galaxies) of the scope of galaxies in the database. This is because running a BDMCMC on the full data set would take a long time - an estimated 40 weeks for the same 40,000 units of algorithm time. Stoica et al. [2007] identified three cuboidal samples or 'bricks' in the 2dF Galaxy Redshift Survey, each with an approximately constant intensity of galaxies. We use a portion of the first brick (NGP150), specifically $W = [0, 40] \times [30, 60] \times [0, 10]$, where galaxy positions are given in respect to the lower left corner of the brick. One reason for choosing this particular subset of the data is that it does not exhibit 2-dimensional *walls* of galaxies, or dense spherical cluster of galaxies. These structures do appear in maps of galaxies, but such objects are not included in our model. It is anticipated that cosmic walls, or 2-dimensional surfaces embedded in the 3-dimensional space, would be challenging mathematical objects to model. It is unclear exactly how to identify a random surface from a field of orientations, or a field of vectors normal to a random surface, as linear integration techniques (used to identify fibres) do not naturally extend to 2-dimensional surfaces. For more information on the various cosmic structures see Martínez and Saar [2002].

The first 13,000 units of algorithm time (of a total 40,000) were discarded. Samples were taken at a rate of 0.013 units of algorithm time. The initial state was a randomly sampled set of $\kappa = 6$ fibres. Other hyperparameters were chosen as follows: dispersion parameter $\sigma_{\text{disp}} = 2$; signal probability hyperparameters $\alpha_{\text{signal}} = 3$ and $\beta_{\text{signal}} = 1$; density parameter $\eta = 1.88$; mean half-fibre length $\lambda = 10$; and the Dirichlet parameter $\alpha_{\text{Dir}} = 1$.

The data are presented in Figure 6.10.

An example of the clustering of points based on one sample is shown in Figure 6.11. It is evident that the data are located over approximately 6 clusters of points. However, from the empirical estimate of the density of signal points (see Figure 6.12) the relative proximity of these fibre clusters is clearer, and it appears that 5 is a better estimate for the number of fibres.

Implementation of all the additional moves described in Section 5.3 in 3 dimensions would improve the mixing properties of the BDMCMC. However these early results indicate that our model extends well to 3 dimensions.

6.4 Conclusions

This chapter demonstrates our approach to making inferences on the underlying curvilinear structure of point patterns through application to four planar data sets.

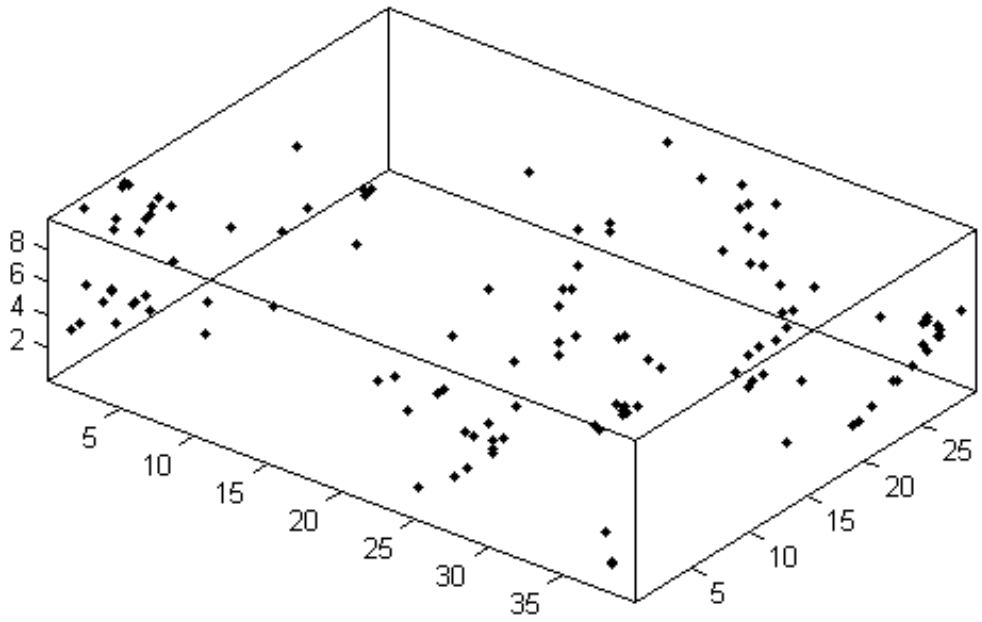


Figure 6.10: Subset of galaxy data taken from the 2dF Galaxy Redshift Survey (Colless et al., 2001). Specifically galaxies located in the window $[0, 40] \times [30, 60] \times [0, 10]$ of the brick of galaxies NGP150, identified by Stoica et al. [2007].

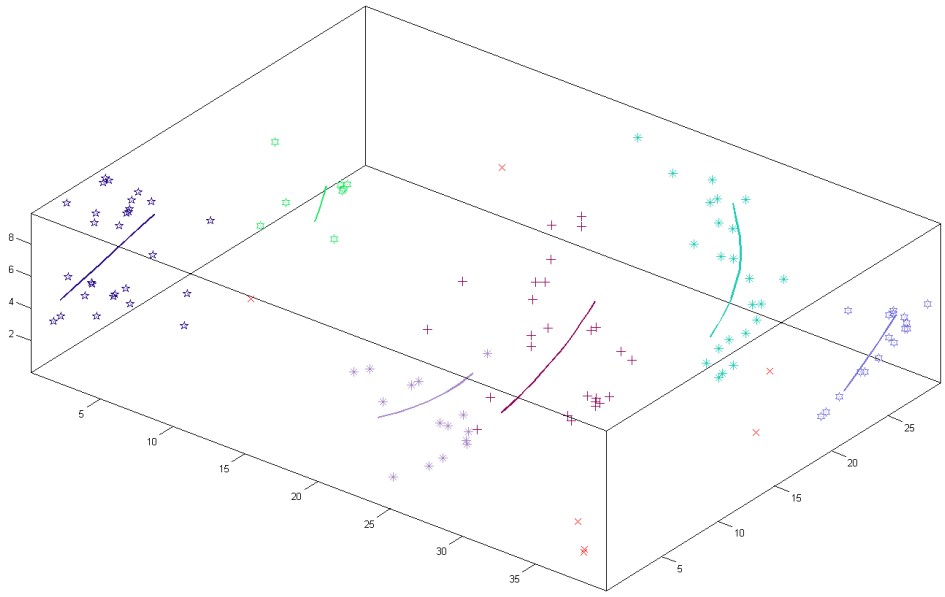


Figure 6.11: Clustering of points in one sample. Curves represent fibres and different symbols represent different clusters

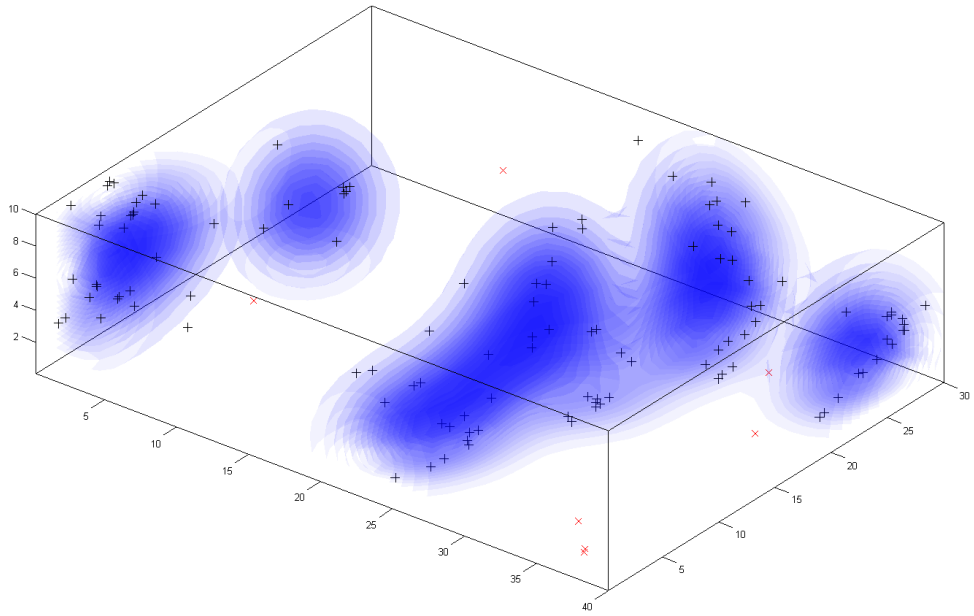


Figure 6.12: Empirical estimate of the density of signal points, darker areas indicate higher densities. Points allocated to noise in more than 50% of samples are represented by $\times$, while other points are represented by $+$.

Posterior Probabilities for Number of Fibres				
Number of Fibres	4	5	6	7
Posterior Probability	0.01	0.32	0.57	0.10

Other Properties Conditioned on the Number of Fibres				
	Number of Fibres	Posterior Mean	50% HPD Interval	95% HPD Interval
Number of Noise Points	5	8.05	[7,8]	[6,15]
	6	7.32	[6,7]	[5,9]
	7	6.09	[5,6]	[4,9]
95th Percentile of the Distances from Signal Points to Fibres	5	5.39	[5.06,5.50]	[4.69,6.15]
	6	5.07	[4.87,5.29]	[4.43,5.94]
	7	5.12	[4.81,5.25]	[4.67,6.04]
Total Length of Fibres	5	51.20	[45,51]	[45,59]
	6	54.62	[50,56]	[46,62]
	7	56.37	[55,60]	[48,65]

Table 6.6: Results for Galaxy Data: The first sub-table gives posterior probabilities on the number of fibres, while the second gives posterior means and 50% and 95% HPD (highest posterior density) intervals for a selection of properties of the posterior distribution conditional on the number of fibres. The data are the galaxies located in the window $[0, 40] \times [30, 60] \times [0, 10]$ of the brick of galaxies NGP150, identified by Stoica et al. [2007]. There are 124 points in a $40 \times 30 \times 10$ window, the dispersion parameter σ_{disp} is set to 2 and the prior mean probability that a point is noise is 0.25. Posterior probabilities are only given if non-zero to rounding error.

Also presented are preliminary results on two 3-dimensional data sets.

Following consideration of these examples, it is apparent that the curvature bias in the field of orientations on fibre samples can affect the number of fibres by causing the model to favour short fibre segments over a long single fibre. This is most apparent in the first simulated example. In order to estimate the number of fibres generating a point pattern from the posterior distribution, a bias-corrected estimator, weighted towards the lower end of the posterior distribution of the number of fibres, may be appropriate.

The examples in this chapter provide evidence that our approach can be applied to data exhibiting various types of fibre structure. For example, fingerprint pores lie close to the centre of ridge lines which, in turn, lie almost parallel to each other on the fingertip, yet the flexible model can be fitted to both this data and the densely clustered data of earthquake epicentres.

In the following chapter, we return to the discussion of tensors, analysing the robustness of the local orientation estimate provided by the tensor method, and introducing a new measure of anisotropy (the degree to which a tensor deviates from isotropy).

Chapter 7

Measures of Anisotropy and Tensor Robustness

This chapter presents a collection of analyses on the tensor method, used to estimate local orientations in a point pattern (see Section 4.3.2), together with some more general remarks on the properties of tensors. The tensor is fundamental to the estimation of the field of orientations, and it is therefore important to gain some understanding of when the tensor method provides a good orientation estimate, and when it is less reliable.

The first section considers the result of applying the tensor method to a homogeneous Poisson process, providing an isotropic model for comparison. We also identify the mean tensor calculated on two other point processes - a homogeneous Poisson process conditional on a specific point, and a Poisson process with a cosine intensity function. These basic point process models provide a reference point for further analysis.

Section 7.2 describes how the tensor can be summarised in terms of three meaningful quantities: orientation, magnitude and anisotropy. While the relation between a tensor and its dominant orientation is clear (indicated by the principal eigenvector), the measures of magnitude and anisotropy have no single definition. A number of different examples from the literature are mentioned, and a new measure of anisotropy, motivated by the tensor method construction, is proposed. Their comparative behaviour for the mean tensor of a homogeneous Poisson process conditional on a single point is presented.

A robustness analysis of the tensor method is presented in Section 7.3, the effect of noise on tensors is estimated by approximating the fibre-generated point process by three basic models. Finally, Section 7.4 describes how knowledge of the anisotropy

of a point pattern could be used to extend the model described in Chapter 3.

Throughout this chapter, the parameter of the tensor method σ_{FO} will be written as σ for the purpose of clarity.

7.1 The Tensor Method Applied to Specific Point Processes

It is not possible to derive simple formulae for the result of the tensor method on a general point process. However the mean tensor can be calculated for some Poisson processes. This section presents the mean tensors for three Poisson processes: an isotropic homogeneous Poisson process, a homogeneous Poisson process conditional on there being a point at a specific location, and an inhomogeneous Poisson process with a cosine-intensity. The first two are useful for understanding the effect of homogeneous background noise on the tensor, the last provides an example of an anisotropic point process.

7.1.1 Homogeneous Poisson Process

In order to study the tensor method and understand what the constructed tensors represent, it is useful to work with a very simple point pattern. While the homogeneous Poisson process is not a definitive benchmark for an isotropic point process it proves a useful starting point, and is extended in the following section by conditioning on a single point. Considering the results of the tensor method on a Poisson process could potentially aid the classification of point patterns into those with underlying directional properties (anisotropic) and those without (isotropic).

We define Π to be a Poisson process on $\mathbb{R}^2$ with constant intensity $\rho > 0$. Without loss of generality we choose to evaluate the tensor created at point q, and use polar coordinates (r, θ) centred at q, with an arbitrary but fixed choice of orientation for $\theta = 0$, to identify the remaining points in Π.

Let T denote the tensor evaluated at q under these Poisson process assumptions,

$$T = \sum_{(r,\theta)\in\Pi} \exp\left(-\frac{r^2}{2\sigma^2}\right) (\cos\theta, \sin\theta)^{\mathrm{T}} (\cos\theta, \sin\theta) \tag{7.1}$$

for some $\sigma \in \mathbb{R}_+$ (see Section 4.3.2, Su, 2009 or Su et al., 2008 for further details).

Now let $A \subset \mathbb{R}^2$ be a Borel set, and define $T(A)$ to be the tensor calculated over points in $A \cap \Pi$:

$$T(A) = \sum_{(r,\theta)\in A\cap\Pi} \exp\left(-\frac{r^2}{2\sigma^2}\right)(\cos\theta, \sin\theta)^{\mathrm{T}}(\cos\theta, \sin\theta). \tag{7.2}$$

Consider an increasing sequence of Borel sets $A_1 \subseteq A_2 \subseteq ... \subseteq A_n \subseteq ... \subset \mathbb{R}^2$, such that $\lim_{n\to\infty} A_n = \mathbb{R}^2$. We will show that $T(A_n) \to T$ and hence calculate $\mathbb{E}(T)$.

Theorem 3. *Let $A_1 \subseteq A_2 \subseteq ... \subseteq A_n \subseteq ... \subset \mathbb{R}^2$ be an increasing sequence of Borel sets such that $\lim_{n\to\infty} A_n = \mathbb{R}^2$. Then*

- $\lim_{n\to\infty} T(A_n) = T$
- $\lim_{n\to\infty} \mathbb{E}(T(A_n)) = \mathbb{E}(T) = \pi\rho\sigma^2\mathbf{I}_2$

where $\mathbf{I}_2$ is the 2×2 identity matrix.

Proof. The first result follows from dominated convergence, as

$$\left|\exp\left(\frac{-r^2}{2\sigma^2}\right)\begin{pmatrix}\cos^2\theta & \sin\theta\cos\theta\\ \sin\theta\cos\theta & \sin^2\theta\end{pmatrix}\right| < \exp\left(\frac{-r^2}{2\sigma^2}\right)\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix}. \tag{7.3}$$

The sum $\sum_{(r,\theta)\in\Pi}\exp\left(\frac{-r^2}{2\sigma^2}\right)$ can be shown to have finite expectation for Poisson process Π. Hence $T(A_n)$ converges almost surely to T.

For the second result we see that, by the monotone convergence theorem,

$$\mathbb{E}\left[\sum_{(r,\theta)\in\Pi\cap A_n}\exp\left(\frac{-r^2}{2\sigma^2}\right)\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix}\right] \tag{7.4}$$

converges, and hence by the dominated convergence theorem, $\lim_{n\to\infty}\mathbb{E}(T(A_n)) = \mathbb{E}(T)$.

By the isotropy (directional invariance) of the Poisson process, we know that the off-diagonal elements of $\mathbb{E}[T]$ are 0, and the diagonal elements are equal. So it suffices to consider

$$\mathbb{E}\left[\sum_{(r,\theta)\in\Pi}\exp\left(-\frac{r^2}{2\sigma^2}\right)\cos^2\theta\right]. \tag{7.5}$$

As the expectation is calculated with respect to the intensity of the Poisson process,

this is written

$$\mathbb{E}[T] = \rho \int_0^{\infty} \int_0^{2\pi} r \exp\left(-\frac{r^2}{2\sigma^2}\right) \cos^2\theta \, \mathrm{d}\theta \, \mathrm{d}r \tag{7.6}$$

$$= \rho\pi\sigma^2, \tag{7.7}$$

hence $\mathbb{E}[T] = \rho\pi\sigma^2\mathbf{I}_2$. □

In fact, if we choose $A_t := B(q,t)$, a series of discs centred at q of radius $t > 0$, then

$$M_t := \begin{cases} \frac{T(A_t)}{1-\exp\left(\frac{-t^2}{2\sigma^2}\right)} & \text{for } t > 0 \\ \rho\pi\sigma^2\mathbf{I}_2 & \text{for } t = 0 \end{cases} \tag{7.8}$$

is a martingale, i.e. a sequence of random variables such that $\mathbb{E}[|M_t|] < \infty$, and $\mathbb{E}[M_{t+1}|M_t, ..., M_1] = M_t$, see Brémaud [1981] for further details. Correspondingly, we can write

$$\mathbb{E}\left[T(A_t)\right] = \left(1 - \exp\left(\frac{-t^2}{2\sigma^2}\right)\right) \rho\pi\sigma^2\mathbf{I}_2. \tag{7.9}$$

So, if it is believed that a point process is approximately distributed as a Poisson process outside a given radius t of point q, then this provides an estimate of the error in calculating the tensor at q omitting all points outside the radius t. As $t \to \infty$, the error reduces rapidly to 0.

7.1.2 Homogeneous Poisson Process Conditional on a Point

Having identified the tensor on a homogeneous Poisson process, we now condition on there being a point at a given location and consider the Palm distribution of the tensor conditional on this point. A Palm distribution is simply a distribution conditional on the occurrence of a specific event; in a point process this is often taken to be the event that there is a point lying at a specific location (see Stoyan et al., 1995). As the events in the Poisson process occur independently, conditioning on a single point does not affect the distribution of the remaining points. This conditional point process is used to compare measures of anisotropy in Section 7.2.4.

The mean tensor $\mathbb{E}[T]$ was calculated at an arbitrary point in a homogeneous Poisson process with rate ρ in the previous section. Now consider $\mathbb{E}[T|\Pi \cap (r_0, \theta_0) \neq \emptyset]$, the conditional expectation of the tensor given there is a point at (r_0, θ_0). As a Poisson process conditional on a point is still a Poisson process with the same rate, we

have

$$\mathbb{E}[T|\Pi \cap (r_0, \theta_0) \neq \emptyset] = \rho\pi\sigma^2 \mathbf{I}_2 + \exp\left(\frac{-r_0^2}{2\sigma^2}\right) \begin{pmatrix} \cos^2(\theta_0) & \cos(\theta_0)\sin(\theta_0) \\ \cos(\theta_0)\sin(\theta_0) & \sin^2(\theta_0) \end{pmatrix}. \tag{7.10}$$

Without loss of generality we can set $(\theta_0 = 0)$ to get

$$\mathbb{E}[T|\Pi \cap (r_0, \theta_0) \neq \emptyset] = \begin{pmatrix} \rho\pi\sigma^2 + \exp\left(\frac{-r_0^2}{2\sigma^2}\right) & 0 \\ 0 & \rho\pi\sigma^2 \end{pmatrix}, \tag{7.11}$$

giving a tensor of the form $\begin{pmatrix} \alpha+\beta & 0 \\ 0 & \alpha \end{pmatrix}$, for $\alpha, \beta > 0$. As $\exp\left(\frac{-r_0^2}{2\sigma^2}\right) < 1$, the knowledge of the location of one point in the Poisson process does not have a large effect on the mean tensor, unless the intensity ρ is particularly small.

7.1.3 Cosine Poisson Process

As an alternative to the homogeneous Poisson process, this section considers the behaviour of the tensor over a Poisson process which is anisotropic around the location the tensor is evaluated at. The Poisson process described here is loosely based on the fingerprint data where sweat pores lie along fibres that are reasonably approximated by a collection of parallel lines.

Specifically, we consider the mean tensor evaluated at $(0, 0)$ of the Poisson process with intensity $\tau(x, y) = \frac{\gamma}{2\pi}(1 + \cos(x))$, for some constant γ, using cartesian coordinates $(x, y) \in \mathbb{R}^2$. The mean tensor calculated by the tensor method is

$$\mathbb{E}[T] = \int_0^{2\pi} \int_0^{\infty} \frac{\gamma}{2\pi}(1 + \cos(r\cos\theta)) \exp\left(-\frac{r^2}{2\sigma^2}\right) r \begin{pmatrix} \cos^2\theta & \cos\theta\sin\theta \\ \cos\theta\sin\theta & \sin^2\theta \end{pmatrix} \mathrm{d}r\,\mathrm{d}\theta \tag{7.12}$$

$$= \gamma \left(\frac{\sigma^2}{2}\mathbf{I}_2 + \begin{pmatrix} (1+\sigma^2)\exp\left(-\frac{\sigma^2}{2}\right) - 1 & 0 \\ 0 & 1 - \exp\left(-\frac{\sigma^2}{2}\right) \end{pmatrix} \right).$$

7.2 Tensor Decomposition

Recall the construction of the initial tensor at data point q, calculated from data $y_1, ..., y_m$, as given in Chapter 4, Section 4.3.2. The vector $v^i = (v_1^i, v_2^i)$ from point

q to point y_i is transformed to

$$\tilde{v}^i \quad = \quad (\tilde{v}_1^i, \tilde{v}_2^i) = \frac{\exp\left(-\left((v_1^i)^2 + (v_2^i)^2\right)/4\sigma^2\right)}{\sqrt{(v_1^i)^2 + (v_2^i)^2}}(v_1^i, v_2^i) \tag{7.13}$$

where σ is a scaling parameter. If we use polar coordinates centred at q, $v^i = (r_i \cos\theta_i, r_i \sin\theta_i)$ (as in the previous section), then the initial tensor at point q is

$$\sum_i \frac{\exp\left(-r_i^2/2\sigma^2\right)}{r_i^2}(r_i \cos\theta_i, r_i \sin\theta_i)^{\mathrm{T}}(r_i \cos\theta_i, r_i \sin\theta_i). \tag{7.14}$$

This tensor

$$T = \begin{pmatrix} A & B \\ B & C \end{pmatrix} = \sum_i \exp\left(-\frac{r_i^2}{2\sigma^2}\right) \begin{pmatrix} \cos^2\theta_i & \cos\theta_i \sin\theta_i \\ \cos\theta_i \sin\theta_i & \sin^2\theta_i \end{pmatrix}, \tag{7.15}$$

where $AC > B^2$, is a 3-dimensional structure, but the meaning of the values A, B and C is not intuitive. This section is concerned with how these three variables can be transformed to variables that are more informative about the properties of the tensor.

7.2.1 Orientation

The natural variable to consider is the orientation of the tensor. This is expressed as

$$\phi = \arctan\left(\sqrt{1 + G^2} - G\right) \tag{7.16}$$

where

$$G = \frac{C - A}{2B} = \frac{\sum_i \exp\left(-\frac{r_i^2}{2\sigma^2}\right)\sin(2\theta_i)}{\sum_i \exp\left(-\frac{r_i^2}{2\sigma^2}\right)\cos(2\theta_i)} \tag{7.17}$$

for $B \neq 0$. If $B = 0$, then the orientation is $\phi = \pi/2$ if $C > A$ and $\phi = -\pi/2$ if $A > C$. If $A = C$ and $B = 0$ then T is isotropic - a multiple of the identity matrix, and has no dominant orientation.

7.2.2 Magnitude

There are two remaining components of the tensor. They can be used to describe the *magnitude* and *anisotropy* of the tensor. The magnitude relates to how much

'information' has contributed to the tensor - under the tensor method this corresponds to a combination of the number of points and the size of their weights $w_i = \exp(-r_i^2/2\sigma^2)$. In particular, it is independent of the tensor orientation ϕ and also the direction of the individual point-to-point vectors v^i.

A natural choice for a measure of the magnitude of our tensor is the trace, $\mathrm{tr}(T)$. For the tensor T in Equation (7.15) this quantity is simply $\sum_i \exp\left(-\frac{r_i^2}{2\sigma^2}\right)$, a weighted sum of the number of points in the pattern, quantifying the amount of information contributing to the tensor.

An alternative measure of magnitude, based on the correspondence between ellipses and tensors is mentioned in the following section.

7.2.3 Measure of Anisotropy

The anisotropy of a tensor describes the extent to which the weighted vectors $\tilde{v}^i$ are concentrated around the dominant orientation. This is a measure of how much the tensor deviates from an isotropic tensor exhibiting no determinable orientation; it literally means 'not the same in all directions'.

Various measures of anisotropy have been suggested for tensors, mostly in literature focusing on diffusion tensor imaging where they are also called diffusion anisotropy indices. For the analysis of the robustness of tensors in Section 7.3, we will use a modified version of the squared Fractional Anisotropy, but first we shall consider other alternatives. The focus of the section is on anisotropy measures for 2-dimensional tensors, but n-dimensional extensions are also given where appropriate.

Note that if the model is extended to n dimensions, more components are introduced. In particular, there are n orthogonal orientations (eigenvectors), each of which has an associated eigenvalue indicating the amount of evidence supporting that orientation as the dominant orientation. For example, in the tensor method the magnitude of the eigenvalue describes the number of data points near to the point q (at which the tensor is calculated) in the corresponding direction. The measure of anisotropy, typically a scalar variable, summarises the extent to which k of these n eigenvalues dominate the others in magnitude, for some k, where $0 < k < n$. In two dimensions we are restricted to $k = 1$, but for higher dimensions there is a decision as to which value of k to use.

Consider the 3-dimensional example, if $k = 1$ a high anisotropy indicates that most of the orientations are aligned in the same direction, whereas a high anisotropy measure using $k = 2$ tells us that most of the local orientations fall within a planar surface

in 3-dimensional space. In this application we are studying point patterns generated from fibres (1-dimensional curves), so anisotropy measures should summarise the extent to which the first eigenvalue dominates the others (i.e. $k = 1$).

The measures of anisotropy described here are split into three categories - those based on ellipses (or ellipsoids in higher dimensions), measures described by the ratios of various measures of magnitude, and measures equal to the distance from the tensor in question to the *nearest* isotropic tensor for a given choice of distance metric. Finally we propose a new measure of anisotropy adapted from the Fractional Anisotropy for easier calculations on tensors calculated from the tensor method.

Measures Based on Ellipsoids

There is a bijective relationship between n-dimensional tensors and n-dimensional ellipsoids (or ellipses in two dimensions) centred at the origin, which is why ellipsoids are often used to depict tensors graphically. The bijection most commonly used maps the tensor T to an ellipsoid with orthogonal axes in the directions of the eigenvectors. The lengths of the semi-axes (half the lengths of the axes) are then related to the eigenvalues, $\lambda_1, \lambda_2, ...,$ of the tensor. Here we take the length of the semi-axes to be equal to the eigenvalues, although it is also common to use the square roots of the eigenvalues (e.g. Dryden et al., 2009). Properties of an ellipsoid can be used to describe measures of magnitude and anisotropy of the associated tensor.

In two dimensions, the natural quantities to measure are the shape and size of the ellipse. The size of an ellipse could be represented by its area, given by $\pi\lambda_1\lambda_2$. As a measure of the magnitude of a tensor the elliptical area has the drawback that it is not independent of angles θ_i. However, the elliptical area is proportional to the determinant of the corresponding tensor $\pi\lambda_1\lambda_2 = \pi\det(T)$, and therefore it shares a useful property: when calculating the mean of m tensors in the log-Euclidean metric, the determinant of the mean tensor, is equal to the scalar geometric mean of the determinants of the m original tensors. This property naturally extends to higher dimensions.

The shape of an ellipsoid relates to the anisotropy of the tensor. Pierpaoli and Basser [1996] mention the *volume ratio* as a possible anisotropy measure in three dimensions. It is defined as the ratio of the volumes of the ellipsoid and a sphere with radius equal to the average length of the ellipsoids axes (i.e. the isotropic equivalent). For a general n-dimensional tensor with eigenvalues λ_i $(i = 1, ..., n)$ it is given by

$$\text{Volume Ratio} = \frac{\prod_{i=1}^{n} \lambda_i}{\bar{\lambda}^n} \quad \text{where } \bar{\lambda} = \frac{\sum_{i=1}^{n} \lambda_i}{n} \tag{7.18}$$

The *area ratio*, the equivalent measure of anisotropy for a 2-dimensional tensor with eigenvalues λ_1, λ_2 is

$$\text{Area Ratio} = \frac{4\lambda_1\lambda_2}{(\lambda_1 + \lambda_2)^2}. \tag{7.19}$$

Another measure based on the geometry of ellipses is the eccentricity (also called the first eccentricity), given in two dimensions by

$$\sqrt{1 - \frac{\lambda_2^2}{\lambda_1^2}}, \tag{7.20}$$

where $\lambda_1 > \lambda_2$.

Ratio-Based Anisotropy Measures

A very simple measure in two dimensions is the ratio of the eigenvalues λ_1/λ_2 ($\lambda_1 > \lambda_2$), given in Basser et al. [1994] which gives a value ≥ 1; a measure of 1 indicates isotropy. However this measure is unbounded, has little intuitive meaning, and in higher dimensions there is no standard way to extend it.

Bahn [1999] suggests taking the ratio of different measures of tensor magnitude. Three such magnitude measures applicable to two dimensions are:

$$A = \frac{\lambda_1 + \lambda_2}{2} \qquad \text{(arithmetic mean)}$$

$$J = \sqrt{\lambda_1\lambda_2} \qquad \text{(geometric mean)}$$

$$K = \frac{2\lambda_1\lambda_2}{\lambda_1 + \lambda_2} \qquad \text{(harmonic mean).}$$

From these we calculate two more measures of anisotropy,

$$J/A = 2\frac{\sqrt{\lambda_1\lambda_2}}{\lambda_1 + \lambda_2} \tag{7.21}$$

$$K/A = \frac{4\lambda_1\lambda_2}{(\lambda_1 + \lambda_2)^2}. \tag{7.22}$$

Note that K/A is equal to the area ratio, see Equation (7.18), and is also the square of J/A. Both these measures of anisotropy *decrease* from 1 to 0 as the anisotropy increases.

The three magnitude measures are easily extended to n dimensions:

$$A_n = \frac{\sum_i \lambda_i}{n} \qquad \text{(arithmetic mean)}$$

$$J_n = \sqrt[n]{\prod_i \lambda_i} \qquad \text{(geometric mean)}$$

$$K_n = \frac{n}{\sum_i \lambda_i^{-1}} \qquad \text{(harmonic mean).}$$

The two measures of anisotropy become

$$J_n/A_n = n\frac{\sqrt[n]{\prod_i \lambda_i}}{\sum_i \lambda_i} \tag{7.23}$$

$$K_n/A_n = \frac{n^2}{\left(\sum_i \lambda_i^{-1}\right)\left(\sum_i \lambda_i\right)}. \tag{7.24}$$

Distance-Based Anisotropy Measures

The measure of anisotropy most commonly used in diffusion tensor imaging is the Fractional Anisotropy (or FA) index (Basser and Pierpaoli, 1996), which in 2 dimensions is given by

$$FA = \frac{|\lambda_1 - \lambda_2|}{\sqrt{\lambda_1^2 + \lambda_2^2}}. \tag{7.25}$$

The FA is defined as the distance from the tensor to the nearest isotropic tensor, where the Euclidean metric is used to calculate distances. In n dimensions, the Fractional Anisotropy generalises to

$$FA = \sqrt{\frac{n}{n-1}\frac{\sum_{i=1}^n (\lambda_i - \overline{\lambda})^2}{\sum_{i=1}^n \lambda_i^2}}, \tag{7.26}$$

where $\overline{\lambda}$ is the mean of the eigenvalues. This is equal to the ratio of the standard deviation of the eigenvalues and $\sqrt{\mathbb{E}(\lambda^2)}$ up to a normalising constant.

The geodesic anisotropy or GA (Batchelor et al., 2005) is an alternative to the FA based on the log-Euclidean metric rather than the Euclidean metric. It is given in 2 dimensions by

$$GA = \frac{1}{\sqrt{2}}\left|\log\left(\frac{\lambda_1}{\lambda_2}\right)\right|, \tag{7.27}$$

and extends to n dimensions as

$$GA = \sqrt{\sum_{i=1}^{n} (\log \lambda_i - \overline{\log \lambda})^2}, \tag{7.28}$$

where $\overline{\log \lambda}$ is the mean logarithm of the eigenvalues, $\frac{1}{n}\sum_{i=1}^{n} \log \lambda_i$. The geodesic anisotropy is not bounded above, but has a minimum value of 1 which, when attained, indicates complete isotropy.

An alternative extension of the Fractional Anisotropy is the Procrustes Anisotropy (PA) given in Dryden et al. [2009]. This has a similar form to the FA, except that it uses the full Procrustes shape distance rather than the Euclidean distance, so that the eigenvalues λ_i are replaced by $\sqrt{\lambda_i}$. In 2 dimensions this becomes $\frac{\sqrt{\lambda_1}-\sqrt{\lambda_2}}{\sqrt{\lambda_1}+\lambda_2}$. The n-dimensional generalisation is

$$PA = \sqrt{\frac{n}{n-1} \sum_{i=1}^{n} \left(\sqrt{\lambda_i} - \overline{\sqrt{\lambda}}\right)^2 / \sum_{i=1}^{n} \lambda_i} \tag{7.29}$$

where $\overline{\sqrt{\lambda}} = \frac{1}{n}\sum_i \sqrt{\lambda_i}$.

Modified Square Fractional Anisotropy

A final measure that we propose here is a modification of the squared Fractional Anisotropy given by

$$msFA = \frac{n}{n-1} \frac{\sum_{i=1}^{n} (\lambda_i - \overline{\lambda})^2}{\left(\sum_{i=1}^{n} \lambda_i\right)^2}. \tag{7.30}$$

Like the Fractional Anisotropy, this measure lies in the interval $[0, 1]$ with 0 indicating isotropy and a value of 1 indicating anisotropy, reached if only one eigenvalue is non-zero. Compare this with the Fractional Anisotropy given in Equation (7.26). Note that, as $\sum \lambda_i^2 < (\sum \lambda_i)^2$ and the Fractional Anisotropy is bounded above by one, the modified square Fractional Anisotropy is always less than the Fractional Anisotropy. It can also be considered as the variance of the normalised eigenvalues,

$$msFA = \frac{n}{n-1} \mathrm{Var}(X_i) \quad \text{where } X_i = \frac{\lambda_i}{\sum_{i=1}^{n} \lambda_i}. \tag{7.31}$$

What makes the modified Fractional Anisotropy appealing is that properties such as its mean and variance can be estimated for tensors created by the tensor method.

In two dimensions it reduces to

$$msFA = \left(\frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2}\right)^2. \tag{7.32}$$

In terms of the tensor components of Equation (7.15), the new modified Fractional Anisotropy is written

$$\begin{aligned} msFA &= \frac{(A-C)^2 + 4B^2}{(A+C)^2} \\ &= \frac{\left(\sum_{i=1}^n w_i \cos(2\theta_i)\right)^2 + \left(\sum_{i=1}^n w_i \sin(2\theta_i)\right)^2}{\left(\sum_{i=1}^n w_i\right)^2}, \end{aligned} \tag{7.33}$$

where $w_i = \exp(-r_i^2/2\sigma^2)$.

Under the assumption that the point pattern is isotropic and hence the angles θ_i are identically and independently Uniformly distributed over $[0, 2\pi)$, we can calculate the expected $msFA$ of tensor

$$T = \sum_{i=1}^{\infty} W_i \begin{pmatrix} \cos^2\theta_i & \cos\theta_i \sin\theta_i \\ \cos\theta_i \sin\theta_i & \sin^2\theta_i \end{pmatrix}, \tag{7.34}$$

conditional on weights $W_i = w_i$ for $i = 1, ..., n$:

$$\begin{aligned} &\mathbb{E}\left[msFA | W_1 = w_1, ..., W_n = w_n\right] \\ &= \mathbb{E}\left[\frac{1}{\left(\sum_{i=1}^n w_i\right)^2}\left(\left(\sum_{i=1}^n w_i \cos(2\theta_i)\right)^2 + \left(\sum_{i=1}^n w_i \sin(2\theta_i)\right)^2\right)\middle| W_1 = w_1, ..., W_n = w_n\right] \\ &= \mathbb{E}\left[\frac{1}{\left(\sum_{i=1}^n w_i\right)^2}\left(\left(\sum_{i=1}^n w_i\right)^2 - 2\sum_{i=1}^n \sum_{\substack{j=1 \\ j\neq i}}^n w_i w_j \sin^2(\theta_i - \theta_j)\right)\middle| W_1 = w_1, ..., W_n = w_n\right] \\ &= \frac{1}{\left(\sum_{i=1}^n w_i\right)^2}\left(\left(\sum_{i=1}^n w_i\right)^2 - \sum_{i=1}^n \sum_{\substack{j=1 \\ j\neq i}}^n w_i w_j\right) \\ &= \frac{\sum_{i=1}^n w_i^2}{\left(\sum_{i=1}^n w_i\right)^2}. \end{aligned} \tag{7.35}$$

The second equality can be seen by expanding the two squared terms involving θ, and comparing with $\left(\sum_{i=1}^n w_i\right)^2$.

Furthermore, the second moment can also be calculated:

$$\mathbb{E}\left[msFA^2 | W_1 = w_1, ..., W_n = w_n\right] \tag{7.36}$$

$$= \mathbb{E}\left[\frac{1}{(\sum_{i=1}^n w_i)^4}\left(\left(\sum_{i=1}^n w_i\right)^2 - 2\sum_{\substack{i,j=1,...,n \\ j\neq i}} w_i w_j \sin^2(\theta_i - \theta_j)\right)^2 \middle| W_1 = w_1, ..., W_n = w_n\right]$$

$$= \frac{1}{(\sum_{i=1}^n w_i)^4}\mathbb{E}\left[\left(\sum_{i=1}^n w_i\right)^4 - 4\left(\sum_{i=1}^n w_i\right)^2\left(\sum_{\substack{i,j=1,...,n \\ j\neq i}} w_i w_j \sin^2(\theta_i - \theta_j)\right)\right.$$

$$\left. + \ 4\left(\sum_{\substack{i,j=1,...,n \\ j\neq i}} w_i w_j \sin^2(\theta_i - \theta_j)\right)^2 \middle| W_1 = w_1, ..., W_n = w_n\right]$$

$$= 1 - \frac{2}{(\sum_{i=1}^n w_i)^2}\sum_{\substack{i,j=1,...,n \\ j\neq i}} w_i w_j$$

$$+ \mathbb{E}\left[\frac{4}{(\sum_{i=1}^n w_i)^4}\left(\sum_{\substack{i,j=1,...,n \\ j\neq i}} w_i w_j \sin^2(\theta_i - \theta_j)\right)^2 \middle| W_1 = w_1, ..., W_n = w_n\right]$$

$$= 1 - \frac{2}{(\sum_{i=1}^n w_i)^2}\sum_{\substack{i,j=1,...,n \\ j\neq i}} w_i w_j$$

$$+ \frac{1}{(\sum_{i=1}^n w_i)^4}\left(\left(\sum_{\substack{i,j=1,...,n \\ j\neq i}} w_i w_j\right)^2 + \sum_{\substack{i,j=1,...,n \\ j\neq i}} w_i^2 w_j^2\right)$$

$$= \left(1 - \frac{1}{(\sum_{i=1}^n w_i)^2}\sum_{\substack{i,j=1,...,n \\ j\neq i}} w_i w_j\right)^2 + \frac{1}{(\sum_{i=1}^n w_i)^4}\sum_{\substack{i,j=1,...,n \\ j\neq i}} w_i^2 w_j^2$$

$$= \left(\frac{\sum_{i=1}^n w_i^2}{(\sum_{i=1}^n w_i)^2}\right)^2 + \frac{1}{(\sum_{i=1}^n w_i)^4}\sum_{\substack{i,j=1,...,n \\ j\neq i}} w_i^2 w_j^2.$$

Thus the conditional variance is

$$\mathrm{Var}(msFA|W_1 = w_1, ..., W_n = w_n) = \frac{1}{(\sum_{i=1}^{n} w_i)^4} \sum_{\substack{i,j=1,...,n \\ j \neq i}} w_i^2 w_j^2 \tag{7.37}$$

These theoretical properties can be used to test if a point pattern is isotropic at different scales by varying σ.

7.2.4 Comparison of Anisotropy Measures

The anisotropy measures are compared by applying them to the mean tensor calculated at a point q in a homogeneous Poisson process, conditional on there being a second point at $(r_0, \theta_0 = 0)$ (in polar coordinates centred at q). The mean tensor, derived in Section 7.1.2, is

$$\begin{pmatrix} \alpha + \beta & 0 \\ 0 & \alpha \end{pmatrix}, \tag{7.38}$$

where $\alpha = \rho\pi\sigma^2$ and $\beta = \exp\left(-r_0^2/2\sigma^2\right)$. Parameters are fixed: $\sigma = 1$ and $\rho = 1/(4\pi)$ so that $\alpha = 1/4$, and β is varied from near zero to 1, (ranging over all the possible distances r_0), and plots are generated for different measures of anisotropy. These plots are shown in Figure 7.1.

It is particularly evident from these plots that there are very few rules governing how a measure of anisotropy should behave as the eigenvalues vary. While they are all monotonic for $\beta \in [0, \infty)$, two decrease from 1 to 0 (J/A and the area ratio), the nearest eigenvector anisotropy increases from 0.5 to 1, the geodesic anisotropy increases from 0 but is unbounded, and the remaining four increase from 0 to 1.

7.3 Robustness of the Tensor Method

This section investigates how different types of noise in the point pattern can effect the tensor created in Equation (4.5), page 46, in particular how well it estimates the local fibre orientation. There are three common types of noise that can occur in signal point data:

(A) Additional noise: random noise points are superimposed on the signal point pattern,

(B) Subtractive noise: random thinning where points are removed from the signal point pattern,

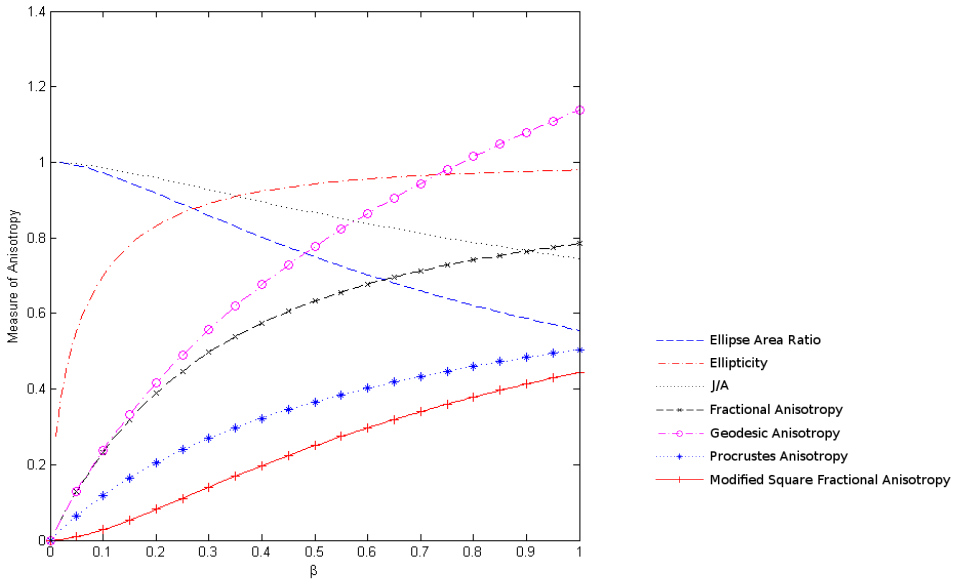


Figure 7.1: Plot of various measures of anisotropy based on a tensor with eigenvalues α and $\alpha + \beta$. The smaller eigenvalue, α, is fixed at 0.25, and β ranges from 0 to 1. The measures of anisotropy shown are: the ellipse area ratio (Equation (7.19), also equal to K/A in Equation (7.22)); the ellipticity (Equation (7.20)); the ratio of magnitude measures J/A (Equation (7.21), also equal to the square of the ellipse area ratio); the Fractional Anisotropy (Equation (7.26)); the Geodesic Anisotropy (Equation (7.27)); the Procrustes Anisotropy (Equation (7.29)); and the modified square Fractional Anisotropy (Equation (7.32)).

(C) Random displacement: signal points are randomly perturbed.

The effects of the first two types of noise are analysed by applying them to basic fibre-generated point process models and estimating properties of the tensor resulting from the tensor method.

The focus in this section is primarily on the fingerprint pore data and for this reason most of the fibre models are based on a structure of multiple parallel lines, each equidistant from the neighbouring lines. Due to the close proximity of fingerprint ridge lines, tensors calculated at signal points in the fingerprint data are generally more isotropic as they include a rather high weighting of points from adjacent ridges. This makes the tensor calculated at a point particularly susceptible to noise; in general the more anisotropic a tensor, the more robust it is to noise.

However, we begin with a more general model, of points randomly located along a single fibre.

7.3.1 Linear Fibre Model

In the first basic model, the set of fibres is approximated by a single linear fibre of infinite length. The tensor method is a local estimate of orientations and as such, approximations made at a larger scale are unlikely to have a significant effect on the tensor produced. The infinite length of the fibre allows us to ignore the possibility that the point q (at which the tensor is calculated) is near an end point. Restricting the type of fibre to a straight line is a reasonable approximation for smooth fibres. The question of how robust the tensor method is around the end of a curvilinear point cluster, or near a fibre with high curvature have not been addressed here.

The signal points are assumed to lie directly on the fibre; in the language of Chapter 3, we let the dispersion parameter σ_{FO} governing the deviation of points from fibres tend to 0. The points are then distributed along the fibre according to a 1-dimensional Poisson process of rate λ.

Background noise is superimposed as a 2-dimensional homogeneous Poisson process of rate ρ, so that the mean tensor calculated at point q on the fibre is

$$\mathbb{E}[T] = \begin{pmatrix} \pi\rho\sigma^2 + \lambda\sqrt{2\pi}\sigma & 0 \\ 0 & \pi\rho\sigma^2 \end{pmatrix} \tag{7.39}$$

where the coordinate system is chosen so that the fibre lies parallel to the first axis.

The modified square Fractional Anisotropy for this model is

$$msFA = \frac{1}{\left(1 + \sqrt{2\pi}\rho\sigma/\lambda\right)^2}. \tag{7.40}$$

This suggests, as we would expect, that increasing the intensity of the background noise will make the tensor more isotropic. Similarly, thinning the signal point process, equivalent to decreasing λ, will reduce the anisotropy of the tensor.

It is informative to compare the effects of subtractive and additive noise as, in some types of data - most noticeably the fingerprint pore data, the data extraction process requires a choice of parameters. Varying these parameters may lead to not only an increase in the number of signal points identified, but also the number of artefacts or background noise. It is clear that if the anisotropy measure $msFA$ is fixed, parameters λ and ρ are proportional. This suggests that if a change in the extraction parameters halves both the number of background noise points and signal points identified, then it will have little effect on the anisotropy of the resultant tensor.

An alternative approach to assessing the robustness of the tensor method is to study the distribution of the orientation of the tensor calculated. Unfortunately, the formula for the orientation of a tensor (Equation (7.17)) is not simple and therefore calculations of the distribution of the orientation are unfeasibly complicated. For this reason we estimate the orientation of a tensor calculated at point q by the direction of the nearest point - signal or noise. While this does not relay much information about the robustness of the tensor, it does provide an approach for identifying what intensity of background noise will cause most methods for estimating local orientations to break down.

The probability that the nearest point to q is signal and lies on the fibre, rather than being a background noise point is given by

$$\begin{aligned} &\int_0^\infty \int_0^x 4\lambda\pi\rho y \exp\left(-\pi y^2\rho - 2\lambda x\right) \,\mathrm{d}y\,\mathrm{d}x \qquad (7.41)\\ &= 1 - \exp\left(\frac{\lambda^2}{\pi\rho}\right)\frac{\lambda}{\rho}\operatorname{erfc}\left(\frac{\lambda}{\sqrt{\pi\rho}}\right), \end{aligned}$$

where erfc() is the complementary error function. A contour plot of this probability for $(\lambda, \rho) \in [0, 10]^2$ is shown in Figure 7.2. Note that if $\rho \propto \lambda^2$ this probability is constant. In particular if $\lambda^2/\rho < 1.7$ then the probability that the nearest point is a signal point is greater than 0.5.

7.3.2 Parallel Linear Fibres Model: Poisson Distributed Points

The general linear model is extended to an infinite set of multiple linear fibres each lying parallel and an equal distance d to each of its neighbouring fibres. Points are distributed at random along the parallel fibres as independent and identically distributed Poisson processes of rate λ.

With homogeneous background noise of intensity ρ, the mean tensor calculated at an arbitrary point q on a fibre is

$$\mathbb{E}[T] = \pi\rho\sigma^2\mathbf{I}_2 + \lambda\sum_{i=-\infty}^{\infty}\int_{-\infty}^{\infty}\exp\left(-\frac{i^2d^2+x^2}{2\sigma^2}\right)\frac{1}{i^2d^2+x^2}\begin{pmatrix} x^2 & idx \\ idx & i^2d^2 \end{pmatrix}\mathrm{d}x \qquad (7.42)$$

$$= \pi\rho\sigma^2\mathbf{I}_2 + \sum_{i=-\infty}^{\infty}\lambda\begin{pmatrix} \sqrt{2\pi}\sigma\exp\left(-\frac{i^2d^2}{2\sigma^2}\right) - di\pi\operatorname{erfc}\left(\frac{|id|}{\sqrt{2}\sigma}\right) & 0 \\ 0 & di\pi\operatorname{erfc}\left(\frac{|id|}{\sqrt{2}\sigma}\right) \end{pmatrix},$$

where erfc() is the complementary error function and $\mathbf{I}_2$ denotes the 2×2 identity

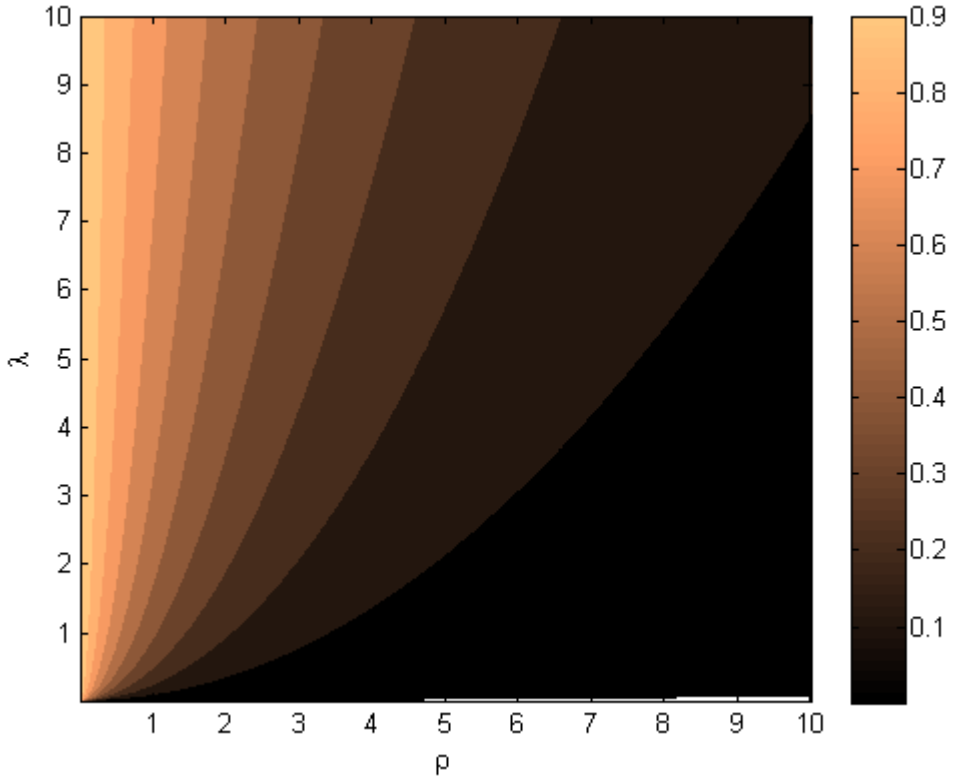


Figure 7.2: Linear Fibre Model: Contour plot for the probability that the nearest point to a signal point is also a signal point.

matrix. The modified square Fractional Anisotropy is then given by

$$msFA = \left(\frac{\sum_{i=-\infty}^{\infty} \sqrt{2\pi}\sigma \exp\left(-\frac{i^2 d^2}{2\sigma^2}\right) - 2di\pi \operatorname{erfc}\left(\frac{|id|}{\sqrt{2}\sigma}\right)}{2\pi\rho\sigma^2/\lambda + \sum_{i=-\infty}^{\infty} \sqrt{2\pi}\sigma \exp\left(-\frac{i^2 d^2}{2\sigma^2}\right)} \right)^2 . \tag{7.43}$$

Due to the infinite summation the $msFA$ can only be estimated computationally.

It is clear from Equation (7.43) that for a fixed value of the $msFA$, the two intensities (λ of the signal process and ρ of the background noise) are proportional. A corresponding result was found for the single linear fibre model. However, the relation between the inter-fibre distance d and the $msFA$ is less clear. Figure 7.3 shows a contour plot of the $msFA$ for different values of the ratio of intensities ρ/λ and the distance between fibres, d. It appears that, once the distance d between two fibres exceeds a threshold value, further changes in d have little effect. In this case, where $\sigma = 1$, the threshold is around $d = 3$. Furthermore, if the distance d falls below another threshold (around 1 in this case), the mean tensor appears to

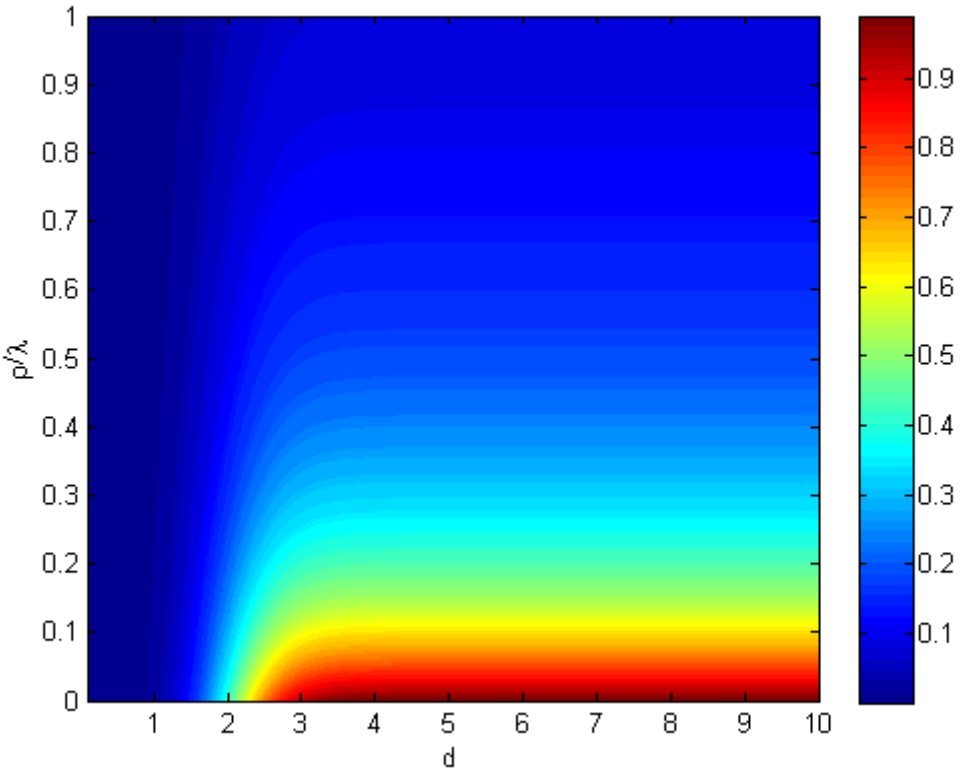


Figure 7.3: Contour plot of the $msFA$ of the mean tensor based on the parallel lines model, for a range of values of the inter-fibre distance d, and the ratio of the noise and signal intensities ρ and λ. The parameter σ was fixed equal to 1.

be isotropic. This is because points on adjacent fibres are close enough to q that their weighted contribution to the tensor cancels with the contributions of points on the same fibre as q. As we would expect, increasing the intensity of the background noise relative to the intensity of the Poisson process of points along fibres, decreases the anisotropy.

The following section considers the cosine Poisson process defined by the intensity function $\tau(x, y) = \frac{\gamma}{2\pi}(1 + \cos(x))$. For a direct comparison of the parallel lines model with the cosine Poisson process model, the distance between the parallel lines is now fixed at $d = 2\pi$. A plot of the $msFA$ for varying intensity ratio ρ/λ, and varying parameter σ is presented in Figure 7.4. Evidently the anisotropy will be maximised by choosing parameter σ as small as possible, indeed as $\sigma \to 0$ the anisotropy $msFA \to 1$, no matter how high the intensity of background noise. It is suspected that the reason for the parameter $\sigma \to 0$ optimising the anisotropy of the tensor, is that the signal points in the parallel lines model all lie directly on the fibre. This is not the case for the cosine Poisson process model, as we shall see in

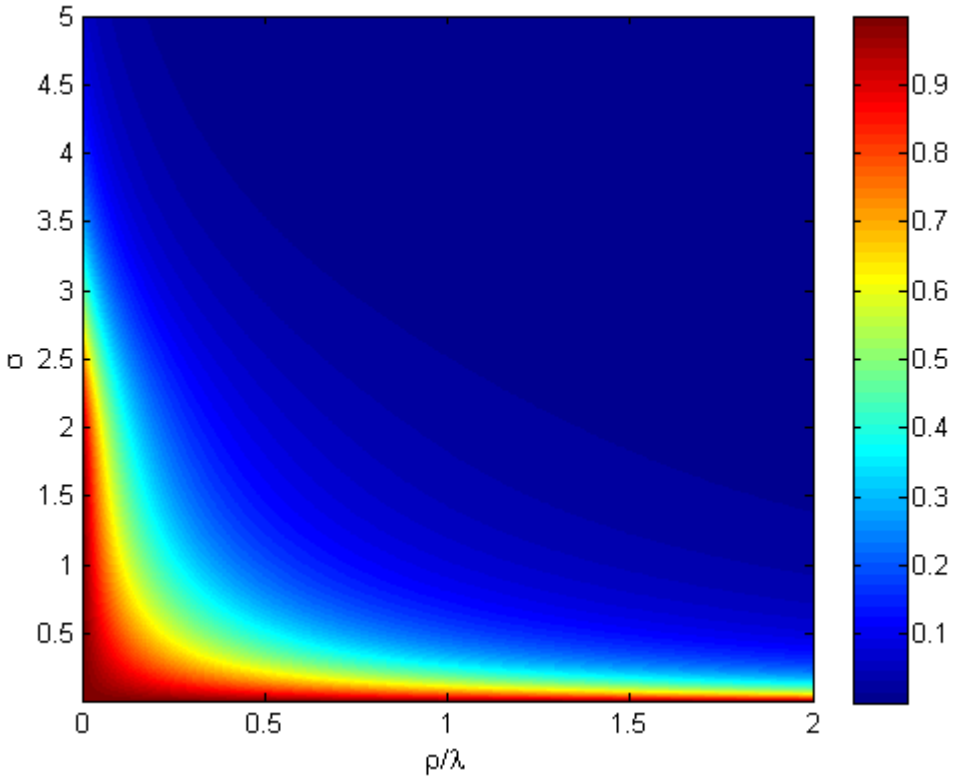


Figure 7.4: Contour plot of the $msFA$ of the mean tensor based on the parallel lines model. The msFA is calculated for a range of values of the parameter σ and the ratio of the noise and signal intensities ρ and λ. The distance between adjacent fibres is fixed equal to 2π.

the next section.

7.3.3 Cosine Poisson Process

Displacement noise can be included in the parallel linear fibres model as Gaussian perturbations of the points. However, calculating the distribution of the distance a point is perturbed by an isotopic kernel generally leads to complicated calculations. As an alternative, we study the effects of noise through the calculation of a tensor on the cosine Poisson process described in Section 7.1.3. This is both an approximation for the point-fibre relationship where each signal point is displaced from an anchor point on a fibre, and a model for displacement noise of points under the assumption that signal points lie exactly on the fibres.

Recall that the intensity of the cosine Poisson process is $\tau(x, y) = \frac{\gamma}{2\pi}(1 + \cos(x))$, for some constant $\gamma > 0$, using Cartesian coordinates $(x, y) \in \mathbb{R}^2$. With background

noise modelled by a homogeneous Poisson process of rate ρ superimposed on the cosine Poisson process, the mean tensor at the origin $(0,0)$ is given by

$$\mathbb{E}[T] = \left(\frac{\gamma\sigma^2}{2} + \pi\rho\sigma^2\right)\mathbf{I}_2 + \gamma \begin{pmatrix} (1+\sigma^2)\exp\left(-\frac{\sigma^2}{2}\right) - 1 & 0 \\ 0 & 1 - \exp\left(-\frac{\sigma^2}{2}\right) \end{pmatrix}. \tag{7.44}$$

This has an modified square Fractional Anisotropy measure of

$$msFA = \left(\frac{2 - (1+\frac{2}{\sigma^2})\exp\left(-\frac{\sigma^2}{2}\right)}{1 + 2\pi\rho/\gamma + \exp\left(-\frac{\sigma^2}{2}\right)}\right)^2. \tag{7.45}$$

It is clear that if there is no background noise ($\rho = 0$) then the constant of proportionality γ in the intensity of the signal process has no effect on the $msFA$.

For a fixed ratio of intensities γ/ρ, it is possible to find the parameter σ that maximises the $msFA$. Figure 7.5 shows a contour plot of the $msFA$ for values of $\sigma \in [0,4]$ and $\gamma/\rho \in [0,20]$. Evidently, a value of σ around 2 produces a relatively high anisotropy for this model, where parallel fibres lie a distance of 2π apart. The $msFA$ is still very low, even for the optimum σ, indicating that the cosine Poisson process is locally more isotropic than the parallel lines model. This suggests that displacement noise can have a significant effect on the anisotropy of the tensor calculated in the tensor method.

7.4 Applications of Anisotropy Measures

This chapter concludes with a discussion of some of the ways in which the anisotropy could be used, both to enhance the inference of the fibre distribution given an instance of a fibre-generated point process, and more generally for the identification of anisotropy in point patterns.

The anisotropy of a point pattern can be estimated locally by using the tensor method to construct a tensor at each point. This provides the basis of a simple test of anisotropy in a point process. For example the null hypothesis that a point pattern arises from an isotropic point process might be rejected if more than a certain proportion of the tensors have a measure of anisotropy greater than c for some $c \in (0,1)$. However, this requires careful consideration as correlation of the tensors across the point process is expected. For example, if two tensors calculated at nearby points both appear to indicate the point pattern is anisotropic, this should not be considered twice the supporting evidence of a single anisotropic tensor in the

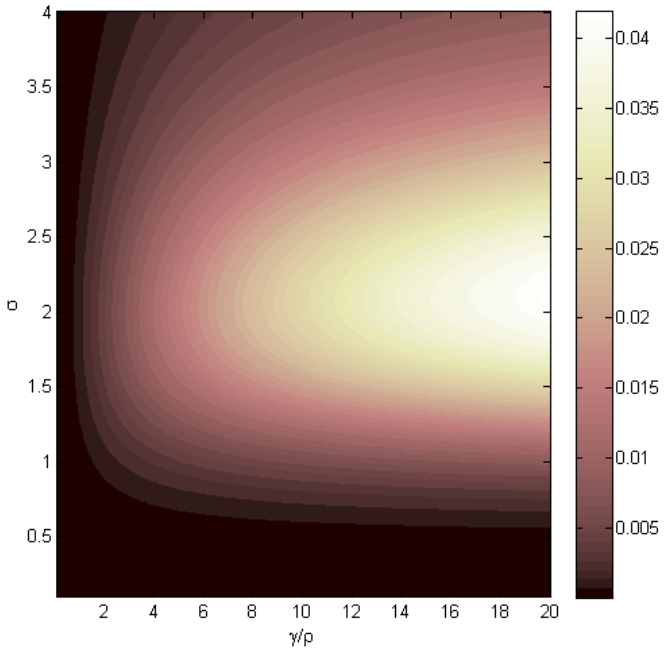


Figure 7.5: The modified square Fractional Anisotropy for the cosine Poisson process for a range of values of σ and the ratio γ/ρ

proposed rejection of the null hypothesis. This is a known problem in multiple testing, see for example Miller [1980].

A high anisotropy of a tensor at a point suggests evidence of a filamentary cluster. Therefore the empirical Bayes approach could be extended to include this information into the prior for the signal probabilities $\boldsymbol{\epsilon}$.

An example of the local anisotropy of a point pattern estimated using the tensor method is shown in Figure 7.6. There are collections of points in both of the two main clusters that are surrounded by a highly anisotropic cluster of points. Compare with the equivalent results on a homogeneous, isotropic point pattern in Figure 7.7.

It is perhaps more informative to calculate a field of anisotropy over the window of observation, averaging point estimates of the anisotropy. The field of tensors calculated as described in Chapter 4, by interpolating the initial tensors, gives rise to a field of anisotropies as well as a field of orientations. Figure 7.8 shows a map of the anisotropy field for Stanford and Raftery's data set (Stanford and Raftery, 2000). It is clear that the two main semi-circular clusters have been identified as areas of high anisotropy.

The empirical Bayes step could be extended to directly use the field of anisotropies.

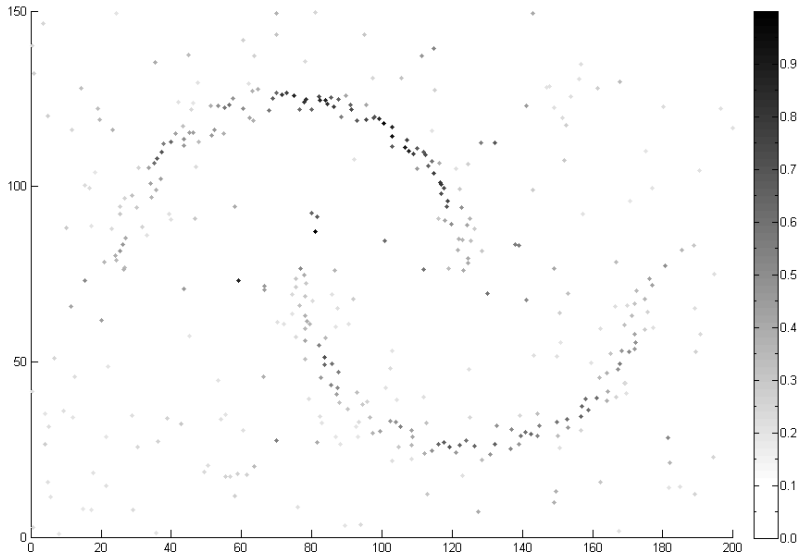


Figure 7.6: Anisotropy plot of data points based on the initial tensors calculated using the tensor method at each point. The data are from Stanford and Raftery [2000] and consist of 200 signal points and 200 homogeneous noise points. Darker points indicate more anisotropic tensors. The measure of anisotropy used was the modified square Fractional Anisotropy, normalised by subtracting the theoretical mean and dividing by theoretical standard deviation (both mean and standard deviation are conditional on the distance-dependent weights). The parameter $\sigma = 13$ was used.

One possible extension is to adapt the prior density function of fibres such that fibres that lie through regions of high anisotropy are favoured over those that lie in regions of low anisotropy. This has the additional benefit of reducing the number of fibres that pass near singularities; such fibres tend to have undesirable 'kinks' where the orientation changes rapidly. Singularities in the tensor field, where the orientation is undefined correspond to zeros in the field of anisotropies. Due to the continuity of the field of anisotropies, the anisotropy in a neighbourhood of a singularity will be near zero, and so fibres in this neighbourhood would be less probable. Also, adapting the birth density, possibly to the new prior density, may be beneficial in increasing the number of fibres proposed in regions of high anisotropy.

7.5 Conclusions

In this chapter we have shown that a tensor can provide more information than just an orientation. Most notably it also describes the anisotropy - a measure of how much variation there is from the dominant orientation. Using the tensor method to construct tensors enables us to measure the anisotropy of a point pattern.

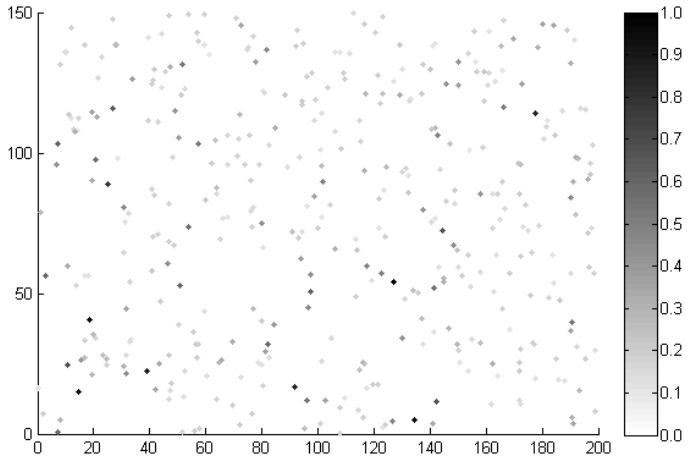


Figure 7.7: Anisotropy plot of data points based on the initial tensors calculated using the tensor method at each point. The data are 400 points uniformly distributed across a 200×150 window. Darker points indicate more anisotropic tensors. The measure of anisotropy used was the modified square Fractional Anisotropy, normalised by subtracting the theoretical mean and dividing by theoretical standard deviation (both mean and standard deviation are conditional on the distance-dependent weights). The parameter $\sigma = 13$ was used.

A new measure of anisotropy has been introduced, the modified square Fractional Anisotropy ($msFA$), adapted from the Fractional Anisotropy (FA) for easier calculations on tensors resulting from the tensor method.

The robustness of the tensor method has been analysed by considering its application to some basic models and calculating the mean tensor. By finding the $msFA$ of the mean tensor we have begun to identify the conditions under which the tensor method breaks down, i.e. fails to provide a reliable estimate of the dominant orientation. We have also briefly described how parameters such as σ can be chosen so as to optimise the anisotropy measured by the $msFA$.

Finally, a few possible ways in which the anisotropy could be used to enhance the fibre-generated point process model of Chapter 3 have been suggested. There is plenty of scope in this area for further work, as is evident from the final section on applications of the measure of anisotropy.

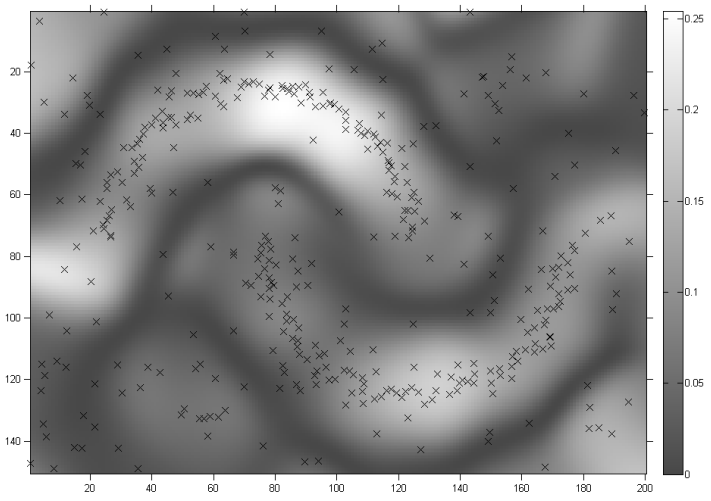


Figure 7.8: Anisotropy field for Stanford and Raftery data. The data are from Stanford and Raftery [2000] and consist of 200 signal points and 200 homogeneous noise points. Lighter areas indicate a higher estimate of local anisotropy. The anisotropy is measured using the unnormalised modified square Fractional Anisotropy over the tensor field. The tensor field was calculated by interpolating tensors created by the tensor method. The log-Euclidean metric was used for interpolation and parameters were chosen at $\sigma = 13$ and $h = 10$. Crosses indicate data points.

Chapter 8

Conclusions

8.1 Discussion

In this thesis we have presented a new model for fibre processes and for point processes generated from a fibre process. Further, we have shown how Monte Carlo methods can be used to sample from the posterior distribution of a fibre process that is instrumental in generating a point process. A methodology such as this, that allows distributional inferences on both the fibres and other parameters to be made, is new to this area of research. Our approach is particularly novel as it produces informative statistics on the generating fibre process for a variety of types of curvilinear clustering. Most notably, we can now reliably estimate the locations of fingerprint ridges from sweat pore data, an achievement which is not reasonably possible with existing methods, and may lead to the reconstruction of fingerprints given pore locations.

As well as fingerprint pore data and earthquake epicentres that cluster around fault-lines, we have proved that our approach can produce informative results on galaxy data, estimating the cosmic structure of the universe. Other data sets for which this model would be suitable include the locations of land mines found in reconnaissance imagery of minefields. Reconnaissance images identify land mines, typically deployed in linear strips, as well as a number of miscellaneous objects. Being able to locate minefields before and during assaults is a matter of great importance to the armed forces, and has led to the USA government's investment in the Coastal Battlefield Reconnaissance and Analysis program or COBRA (Witherspoon et al., 1995).

This process fits non-parametric curves to point patterns with just two limitations on the nature of the curves: curves must not intersect, and must also be 'sufficiently'

smooth (indeed, all curves are C^∞). The two restrictions, which arise as the curves are identified as integral curves of a field of orientations, lead to a very flexible state space of fibres and are generally appropriate for the data considered. The non-intersection property may be less desirable but, at some computational cost, the model could be generalised to allow each fibre to integrate a different field of orientations, see Section 8.2.2 for further details. Certainly, our approach hugely benefits from being formulated around a Bayesian hierarchical model which makes extensions and adaptations readily feasible. For example, extending the model to allow the fibres to vary in 'width' by varying the parameter σ_{disp} (governing the displacement of points from fibres) between fibres could be implemented easily.

The main limitations of our model arise from the issue of non-intersecting fibres, and the constraints on the similarity of fibres. Fibres are assumed to be of the same width (the displacement of points from the fibres is independent of the fibre), and have the same mean points per unit fibre length. These are not always reasonable assumptions, as is evidenced by the earthquake data set. Possible extensions to the model that eliminate this issue are suggested in Section 8.2.2.

It is crucial that the field of orientations is integrated by fibres that produce high likelihoods given the point pattern data, as otherwise the MCMC will not traverse the peaks of the posterior distribution. Through an empirical Bayes approach, calculating point estimates of local orientations within the point pattern and interpolating these estimates, we have shown how to produce such orientation fields. Again, our approach produces good estimates of the field of orientations (i.e. satisfying the criterion that they are integrated by fibres that produce high likelihoods), for a variety of clustered data types: dense or sparse clusters, clusters of varying curvature, the parallel and very sparse clusters of the fingerprint pore data, etc. Our approach even extends to 3 dimensions with the single issue of the scaling problems that are inherent problem in higher dimensional analysis.

A further strength of our model is that it fits the noise-signal and cluster allocations implicitly, in contrast to other work where the clustering must often be predetermined. The advantage is that we can produce reliability estimates for these clustering and noise allocations and explore more potential clustering configurations, and hence more fibre structures.

The complexity of the model, considering the infinite dimensionality of the field of orientations, raises the question of whether or not the Markov chain adequately explores the sample space. Our examples indicate that, whilst the sample space of fields of orientations is not explored particularly well, the space of fibre configurations is well explored and the field of orientations varies enough to explore a wide space of fibre configurations. However, as the density of fibres increases, so the

BDMCMC algorithm requires a longer runtime to overcome these issues.

Our approach performs consistently well, both accurately estimating properties of simulated data and providing an explanatory summary of unknown properties of the earthquake and fingerprint data. It is most notable that it does significantly better than other available techniques on the fingerprint data where a large number of densely-packed fibre clusters account for most of the data.

The remainder of this chapter describes some of the main issues with this approach and, where applicable, suggested solutions to the problems. Also proposed are a number of areas of work that could benefit from further analysis, and may open up new fields of interesting research.

8.2 Issues and Further Work

8.2.1 Edge Effects

It is necessary to bear in mind the ramifications of edge effects in the model and subsequently the BDMCMC algorithm. As we are sampling from a bounded subset $W \subseteq \mathbb{R}^2$ the omission of potential points and fibres just outside W induces a bias on distance-related measures. These edge effects appear in multiple stages of the process, for example:

- The field of orientations will have a bias at the edge favouring orientations parallel to the sides of a rectangular window W. There is slight evidence of this bias in the orientation of the initial tensors of the full fingerprint data, shown in Figure 4.4, page 50. Much of this bias is reduced through the smoothing step in the orientation field estimation.

- Fibres are created by sampling a random reference point from the field and integrating the field of orientations from that point. However the reference point cannot be sampled from outside W, and fibres that extend past the boundary of W are ill-defined as the field of orientations is only defined over the window. We have reduced this edge bias by assigning a probability of 0 to the collection of fibres that do not lie completely within W.

- The model for the displacement of points from fibres does not account for edge effects. There is little evidence of this having a significant effect in the examples of Chapter 6, but it should be borne in mind that there will be a biasing effect on fibres that lie near the boundary of W.

Most of these biases would be significantly decreased by creating a wide border around W and completing the analysis over the whole area. However this would come at a large computational cost.

8.2.2 The Fibre and Point Model

As the model is hierarchical, is it extremely easy to introduce further complexity.

A few extensions of the model are suggested here.

Signal Point Process

As described in Chapter 3, signal points are modelled as normally dispersed points from random anchor points which lie on the fibres. However, other distributions may be preferable.

In certain data, such as minefield locations, it may be believed that the points are uniformly dispersed. This can be included by modelling signal points associated to the fibre F as a uniformly distributed point process over the region $\{x \in W : \|x - y\| < d \text{ for some } y \in F\}$. The maximum distance d of points from the fibre determines the dispersion parameter. This is similar to the Candy model (Stoica et al., 2005) which models points as being uniformly distributed across a collection of connected cylinders.

Sensitivity of Parameters

In Section 6.1.1 we suggested how hyperparameters could be chosen given some prior knowledge of the fibre process. Our experience is that the model is reasonably robust to changes in hyperparameters. For example, our experiments have found that the hyperparameters κ governing the number of fibres, and λ governing the lengths of fibres, can be doubled (or halved) with little effect on the posterior distribution. The model is also fairly robust to changes in the density parameter η, although a very low value can lead to a bias on the fibres, favouring a higher total fibre length, and often leading to extrapolation of fibres into areas of noise. Similarly, a poor choice of the parameter α_{Dir}, which determines the amount of regularity in the prior distribution of the anchor points, will not greatly affect posterior statistics, but higher values can hamper the mixing properties. This is because the prior distribution becomes highly modal with large α_{Dir}, and yet no distributional qualities of this prior are

included in the proposal distributions for the anchor points, leading to proposed moves that are frequently rejected.

The model is somewhat more sensitive to σ_{disp} which governs the deviation (or dispersion) of points from the fibres. If chosen too large the resulting posterior distribution will favour too few fibres with a sizeable error in their locations. If chosen too small, fibre clusters may be split into multiple parallel smaller clusters. However, our experience suggests that small changes in σ_{disp}, i.e. $\pm 30\%$, will have no significant effect on the main characteristics of the posterior distribution of fibres.

Extending the Fibre Process

Two restrictions that we have imposed on the fibres are that the fibres do not intersect, and that they share similar properties; fibres are assumed to be of the same width (the displacement of points from the fibres is independent of the fibre), and have the same mean points per unit fibre length. Following the analysis of the earthquake data set in Section 6.2.3, where the curvilinear point clusters vary in width and density, we suggested the extension of the model, allowing hyperparameters of fibres to vary.

The model can be easily adapted by replacing hyperparameters such as λ (half the mean length), and σ_{disp} (the parameter governing the dispersion of points from fibres) by indexed parameters $\lambda_1, .., \lambda_k$ and $\sigma_{1,\text{disp}}, ..., \sigma_{k,\text{disp}}$. A hyperprior on these parameters should be introduced, and they may be updated as part of the BDMCMC. Otherwise, with different fixed hyperparameters (or different hyperpriors) for each cluster, a fixed labelling would be imposed on the fibre process. Hyperparameters or hyperpriors would need to be identified for any possible number of fibres k, which would rather complicate the process.

The model could conceivably be extended to assign a different orientation field to each fibre. This would permit fibre processes containing meeting or crossing fibres. For each fibre, a field of orientations must be estimated which is integrated by fibres that produce a high likelihood. A natural extension to the empirical Bayes estimation of the field of orientations for a single fibre cluster is to restrict the data sets to the points associated to the fibre. However, it is likely that this approach would be sensitive to the initial clustering, similar to the principal curves approach of Stanford and Raftery [2000]. An additional issue with this approach is that any move that reassigns points to fibres would require the orientation fields to be recalculated, and hence also the fibres. It is possible that the auxiliary (signal probability) variable $\boldsymbol{\epsilon}$ could also be extended to include the probability that each

point belongs to a particular fibre cluster, and orientation field estimates could be based on these weightings. Further investigation is required to prove whether such an approach would be successful and whether sufficiently mixing BDMCMC chains could be constructed.

The question of how to include the crossing fibres that integrate a principal eigenvector orientation field, is one that has been given substantial consideration in the diffusion tensor imaging literature. One proposed solutions is tensor deflection (Lazar et al., 2003) where the entire tensor determines the direction of integration, rather than the principal eigenvector orientation. A second solution is regularisation, reposing the problem as the minimisation of some functional, see for example Deriche and Descoteaux [2007]. With this approach the orientation of integration in regions of the tensor field which exhibit low anisotropy are effectively adjusted in favour of fibres with particular features (e.g. continuous smooth fibres). Based on these ideas, we could adapt the notion of integration of the field of orientations, so that rather than forcing the fibre tangent to *agree* with the field of orientations, the two orientations may simply need to be close. The measure of *how close* these two orientations are, could be determined by the field of anisotropies. Hence areas of low anisotropy (e.g. regions near singularities), are more likely to contain fibres which intersect or cross.

Another extension that could be made to the fibre process is to impose restrictions on where fibres may lie relative to each other. For example we may have a prior belief that there is a threshold on the minimum distance between pairs of fibres, or that they roughly lie in parallel, for example the ridge lines on a fingerprint. The fibre process could also be varied by imposing restrictions on the locations of fibres in W by extending the empirical Bayes approach and favouring fibres in areas of high point density or high anisotropy, as estimated from the data.

A further extension would be to include isotropic point clusters. Isotropic features appear in the galaxy data, see for example the large cluster on the left-most corner in Figure 6.10, 124. These could be easily included, for example as a Poisson cluster process (Neyman and Scott, 1958) producing a richer collection of models.

Noise Points

Lund et al. [1999] describe three types of noise that may occur in spatial point patterns: superposition of 'ghost points' (or additive noise), thinning (or subtractive noise) and random displacement (or perturbation noise). In the model, we include noise as an additive background homogeneous Poisson process. This could be extended by permitting variability in the background noise, by modelling it simply

as an *inhomogeneous* Poisson process. We briefly suggest how perturbation and thinning noise may also be included the model.

Perturbation noise, is included, to an extent, in the dispersion of points from their anchor points. If perturbation noise is believed to have an additional effect on the model, for example by inaccuracies in the data collection, then this could be included in the model by choosing a different distribution for the location of a signal point given the corresponding anchor point. For example, one could take the convolution of the Gaussian kernel with a second kernel describing the perturbation due to noise.

A rather more challenging problem is to model subtractive or thinning noise, where some data were unobserved. Note that the aim is not to reconstruct the data given a sample of fibres and associated variables, but to directly include subtractive noise in the model. If the anchor points are modelled as a Poisson point process along the fibres (i.e. the Dirichlet parameter $\alpha_{\text{Dir}} = 1$), then subtractive noise is included by scaling the density parameter η. However, in our approach the distance between adjacent anchor points on a fibre is modelled by a general Gamma distribution (as described in Section 3.2.1), rather than an exponential distribution. Subtractive noise is often modelled as the thinning of a point process whereby every point is independently removed from the complete (noise-free) point pattern with some small probability $\rho \in (0, 1)$. The distances between adjacent anchor points in a point process that has undergone independent thinning, are conditionally Gamma distributed with size parameter determined by a geometric distribution. Before subtractive noise in the signal process may be implemented, further investigation into how this distribution behaves and how it effects acceptance probability calculations in the BDMCMC is required.

8.2.3 Estimation of the Orientation Field

Our approach relies on the estimation of a field of orientations that is well suited to the data. By this we mean the field of orientations should be integrated by curves that produce high likelihoods. The calculation of an orientation field estimate must be reasonably efficient despite the substantial size of the field of orientations, as the BDMCMC methods require repeated estimations of the orientation field. Bearing in mind both of these issues, we have chosen to estimate the field of orientations using tensors as described in Chapter 4.

Some possible extensions are suggested below.

The Tensor Method

The tensor method, as described in Section 4.3.2 estimates effectively the local orientations in a pattern, however there is still potential for further analysis and possibly some improvement. For example, the Gaussian transformation applied to the point-to-point vectors in Equation (4.4) on page 46 in the first stage of the calculation of the initial tensor, was chosen because it happens to be smooth and decreases with distance, and is based on a well understood function. Alternatives have been suggested such as a transformation based on the functional version of the Cauchy density. Further investigation may tell if an alternative transformation is preferable, although it is believed that (rather like kernel choice in kernel smoothing) the choice of transformation has little effect on the orientation estimation.

Another extension that has been considered, is whether to allow σ_{FO} to vary across the data. If a tensor is calculated at a point y_j with the nearest point y_i situated at a distance greater than $2\sigma_{\mathrm{FO}}$ then this point will usually dominate the tensor calculation, regardless of the perceived isotropy of the point pattern. It can prove informative to allow σ_{FO} to increase when the empirical density of the point pattern around y_j is low. This is particularly effective when using tensors to measure anisotropy, as described in Chapter 7, because different scales of anisotropy can be measured corresponding to the varying density of the point pattern. Experiments using a Delaunay triangulation to estimate the local density appear to perform well.

Extending Steven's Method

Ideas from the tensor method described in Section 4.3.2 and Steven's method (Section 2.1.3), could be combined to provide an alternative tensor estimation of the local orientation. In Steven's method, for each point y, all arcs connecting pairs of points in a neighbourhood of y (a disc of radius r centred at y) are collected. Rather than following Steven's method to create a histogram of the orientations of these arcs, a tensor could be constructed from the orientations, with the contribution of each arc weighted by a function of its length. There is also potential to dispense with the cutoff radius r used in Steven's method and replace it with a smooth weighting function.

By considering the arcs connecting any pair of points in a radius of y, rather than just those with an end point at y, Steven's method induces more smoothing than the tensor method. This means that it is more likely that estimations within a close proximity will have similar orientations. This is natural for an algorithm

which was designed to mimic human visualisation, but it is unknown whether the additional smoothing would improve the orientation estimation or just increase the smoothing bias. Indeed, it is likely that it would only be well suited to certain types of data.

Curvature Bias

The phenomenon of curvature bias, caused by the interpolation and smoothing step in the field of orientations estimation, is described in Section 4.5 and its effects on the estimation of parameters has been commented on in Chapter 6. Possible corrections to the bias are suggested in Section 4.6. The consequence of the curvature bias is that sampled fibres may not fit the data well (may not produce high likelihoods). While this bias has had little significant effect on the examples of Chapter 6, it is much more evident in the full fingerprint data set, where small discrepancies in the field of orientations have a greater effect on the long fibres. This is noted as a fruitful area for future research.

Comparison of Orientation Fields

It can be convenient to measure the discrepancy between the estimated field of orientations and a known field of orientations (where it exists). For example, the field of orientations estimated over the fingerprint pore data can be compared to a direct estimate of ridge line orientations found using gradient-based approaches over the original fingerprint image. This provides an alternative quantification of the curvature bias, when measuring the displacement of singularities is not possible.

Experiments show that as a measure, the mean squared difference in orientation calculated over the window W, highlights the central region in the fingerprint exhibiting curvature bias, and the areas at the edges where there is insufficient information to accurately estimate the ridge orientation. Further investigation and inclusion of ideas from directional statistics (see Mardia and Jupp, 1999) may provide a more suitable measure of discrepancy between orientation fields.

Full Orientation Field Density

In the model described in this thesis, an empirical Bayes approach is used to estimate an instance of the field of orientations dependent on the data and an auxiliary variable, $\boldsymbol{\epsilon}$, denoting the signal probabilities. Rather than using a point estimate

of the field of orientations, the model could be extended to include a prior distribution of orientation fields with parameters estimated through an empirical Bayes step.

A simple extension would be to use a Gaussian random field (Adler and Taylor, 2007), modulo π, with a mean orientation field estimated by the current empirical Bayes method, described in Chapter 4. This would decrease the bias of the orientation field on the fibre set at the expense of an increase in variance, necessitating the exploration of a larger state space. However, the run-time of the BDMCMC could be reduced if the mean orientation field were calculated only once, reducing the number of tensor calculations.

This orientation field density could be extended by using other properties of the field of tensors such as the field of anisotropies (mentioned in Section 7.4) to model the variance of the field of orientations at that point.

An alternative approach is to use a random field of *tensors*, with mean estimated through the empirical Bayes step. One approach may be developed from taking the exponential of the symmetric 2×2 matrix whose three elements are Gaussian random fields. However, some careful consideration is required to ensure that the field of orientations corresponding to an instance of a tensor field from this density satisfies the desired properties, such as smoothness, a finite number of singularities, and that the corresponding orientation field is generally integrated by fibres with high likelihoods.

Singularities in the Field of Orientations

Singularities define the overall shape of the field of orientations and have potential uses in identifying bifurcating or crossing fibres. However integration of the field of orientations can be unreliable around a singularity, as briefly mentioned Chapter 2. Integration of a field of orientations is ill-defined at a singularity as there is no dominant orientation. In itself this is a minor issue as, by estimating the orientation field discretely, we ensure that the probability of evaluating the field at a singularity is 0. However, orientations can vary rapidly around singularities and therefore the discretisation required to integrate the orientation field is subject to large errors. It may prove informative to further study the effect of the singularity on the integration of an orientation field.

Correspondence with Other Mathematical Objects

The subject of the mathematical object that we have termed a *field of orientations* has not been given much attention in the literature. However, there is substantial literature on other mathematical objects that may provide further insight into these orientation fields.

By construction, the orientation field calculated in Chapter 4 corresponds to a tensor field. There is substantial literature on tensor fields due to their frequent usage in engineering, physics, image reconstruction and diffusion tensor imaging.

There is also an obvious similarity between vector fields and orientation fields: any continuous vector field can be transformed into a continuous orientation field by taking the direction of each vector, modulo π. However, the converse is not necessarily true as, for example, if the field of orientations determined by the arch model described in Theorem 1 (page 58), were transformed to a unit vector field by including a direction of orientation, there would be infinitely many singularities. An alternative mapping from a field of orientations $\upsilon_{\mathrm{FO}} : W \to [0, \pi)$ to a vector field $\upsilon_{\mathrm{VF}} : W \to [0, 2\pi)$ is to simply double the angles $\upsilon_{\mathrm{VF}}(x) = 2\upsilon_{\mathrm{FO}}(x)$. Further investigation into the correspondence between vector fields and orientation fields could prove beneficial to further research. In particular, it may prove informative to study how the theory of stationary flows and dynamical systems (see for example Irwin, 1980) relates to properties of the field of orientations such as the effect of noise on integration and singularities. For example, there is a dynamical systems theorem that states sufficient conditions for the integral curve of a slightly perturbed vector field to be continuous. We can also determine properties of integral curves, such as how many times they may be differentiated, given similar properties of the vector field.

An alternative approach is to consider a scalar field with the property that if a point x can be reached by integrating the field of orientations from a point y, then the scalar field evaluates to equal values at x and y. A continuous scalar field that satisfies this property should exist if the orientation field has no singularities. The gradient and other properties of the scalar field may assist in the understanding and reduction of length-biased fibre processes. The scalar field could also provide a neat alternative to identifying fibres, removing the discretisation error that arises when integrating the field of orientations. Indeed, if each streamline of the field of orientations corresponds to a unique value in the scalar field, then the iso-contours of the scalar field identify the streamlines of the orientation field.

8.2.4 Birth-Death Markov Chain Monte Carlo

The birth-death MCMC as described in Chapter 5, provides a method for sampling from the posterior distribution of fibres given an instance of a fibre process. However, there exist alternatives to continuous-time birth death Markov chain Monte Carlo methods. One such alternative, as mentioned at the beginning of Chapter 5, is the random jump Markov chain Monte Carlo.

The two main issues of the BDMCMC: the long run-times and the question of sufficient mixing are discussed here.

Computational Issues

The run-times on the 2-dimensional data sets presented in Chapter 6 range from under 2 to 3.5 days, using an execution node with 3.16GHz processors and 2GB RAM. While this is long in data analysis terms it is not an unworkable length of time. Indeed similar uses of MCMC algorithms, such as in the reconstruction of a Voronoi tessellation from a point process (Skare et al., 2007) can take over 4 days to run. We remind the reader that, as this work is in the developmental stage, the algorithm and implementation have not yet been fully optimised.

One way in which the BDMCMC could be improved is to adapt the birth proposal density (given in Section 5.2.1) so that fibres which produce high likelihoods are more likely to be proposed. For example, the density could be changed so that fibres are only proposed if they pass near a data point. A natural extension to the current proposal density is to favour reference points that lie near data points. Such a proposal density for the reference points could be found by applying a kernel smoothing to the point data.

The other principal way to change the proposal density of fibres given a field of orientations is to vary the lengths of fibres. This could be achieved by imposing a location-dependent stopping probability on the integration of the field of orientations. For example, this could depend on the density of data points, or the associated anisotropy field of the field of orientations. However these adaptations result in a fibre process that is increasingly difficult to describe, and discussion of edge effects becomes equally complicated.

The run-time of the BDMCMC may also be reduced by implementing a more efficient algorithm. Some of steps taken to improve the efficiency were described in Section 6.1.2. Implementing the code in a more powerful language such as C would likely also decrease the run times.

Some consideration has been given as to whether the code could be implemented in parallel, possibly by running separate chains for each disjoint section of the window, and using the approximation that most calculations are based on distances and are therefore local. Alternatively, moves could be proposed simultaneously on different processors. This would only be beneficial with moves that are frequently rejected as moves must be accepted sequentially, and death rates consequently updated. Parallel chains could also be used to improve mixing as described in the following section.

Mixing Issues

A general issue with MCMC algorithms is that it is difficult to confirm whether they are mixing sufficiently well. Sufficient mixing of the chain is necessary in order to reliably draw representative samples from the target distribution. Experience of the implementation of the BDMCMC described in Chapter 5 suggests that data arising from larger sets of fibres produce slower mixing chains.

Generally, sufficient moves have been introduced (such as adjusting the location and length of fibres) to allow each fibre to move about explore nearby clusters. Births, deaths, splits and joins provide frequent moves between different numbers of fibres. Together these moves improve the mixing of the BDMCMC.

A further improvement to the mixing properties of the chain may be found by using a Metropolis-coupled approach. Multiple chains are run with different stationary distributions, one of which is the posterior distribution. Swaps can be proposed between two chains and are accepted according to a Metropolis-Hastings probability, see Gilks and Roberts [1995]. In our model, this could be implemented by gradually varying the dispersion parameter σ_{disp} across chains. When σ_{disp} is larger, there is less restriction on the location of fibres given points, and therefore fibres can move and switch places more frequently. With Metropolis-coupled MCMC the strong mixing properties of the chains with large dispersion parameter σ_{disp} are transferred, to an extent, to the chain with the appropriate target distribution. Simulated tempering, also described in Gilks and Roberts [1995], provides a single chain alternative to this approach, but this requires a longer chain, and therefore longer run-times.

8.2.5 Other Data

We have focused on data sampled over a subset of the 2-dimensional plane, and suggested how this could be extended to 3-dimensional point patterns. Further research possibilities include the fitting of 2 dimensional surfaces in 3 dimensions.

Then new geometric issues need to be taken into account; for example it is not the case that a generic field of tangent planes can be developed into a fibration by surfaces. This would likely provide a challenging area of research.

The model could alternatively be extended to cope with different types of data.

One such type of data is a point pattern sampled over some non-Euclidean manifold such as a sphere or a torus. This would provide an interesting technical challenge, taking care to ensure that distances are well-defined, and introducing an alternative to the log-Euclidean tensor metric.

An alternative extension is to allow marked point processes. The mark may be a magnitude such as the strength of an earthquake or the estimated size of a galaxy, or a confidence probability, for example the prior probability that a data point has been correctly identified as a pore in the pore extraction process of a fingerprint image. Alternatively, the mark may be categorical, such as the shape of a galaxy: elliptical, spiral or irregular.

Confidence probabilities could naturally be included in the auxiliary variable $\boldsymbol{\epsilon}$ which indicates the probability that each data point is signal.

Similarly, a magnitude could be incorporated into the prior probability that each point is signal, so as to favour fibres that 'generate large points'. Alternatively, the prior for the allocation of points to fibres could be extended so that all points associated with a particular fibre must be of a similar magnitude. Depending on the choice of model, this may require the extension of the fibre process to a marked fibre process. If the magnitude describes a height and is comparable to the planar dimensions then the point pattern could simply be treated as a 3-dimensional point pattern, and a 3-dimensional fibre process fitted correspondingly. Another approach is to transform the magnitude so that it is comparable to the planar dimensions and treat the data as a 3-dimensional point pattern.

Similarly, if the marks indicate categories for the data then a marked fibre process could be used, or the data could be split into smaller sets, by category, which would then analysed separately.

8.2.6 Minutiae in Fingerprint Data

In fingerprint identification and verification, the features that are most prominently used for comparison and identification are minutiae: ridge endings, bifurcations, etc. It would be of interest to study whether the minutiae can be determined from pore data. Bifurcations are equivalent to the meeting point of two fibres and are therefore not possible in our fibre model. Evidence of the location of a potential

bifurcation may be suggested by areas of low anisotropy, or by comparison of two independent samples of fibre sets, however they can be indistinguishable from ridge endings.

8.2.7 Reconstruction of Missing Data

Reconstruction of missing data has been briefly mentioned already in Section 6.2.4. If point pattern data is missing over a sub-region of W, then in principle the generating fibre process and other parameters could be estimated, from which the missing data could be reconstructed. This idea of reconstructing data is particularly relevant in the fingerprint pore data, where a smudge on the fingerprint often results in undetected pores. Estimation of the location of missing pores may assist in comparison or identification of fingerprints.

8.2.8 Direct Clustering from Field of Orientations

A rather quick, though less informative way to estimate the fibre process conditional on a field of orientations, is to compare the collection of fibres found by integrating the orientation field from each data point. Depending on prior knowledge of the fibre process, either full integral curves (streamlines) may be chosen or the field could be integrated to a length λ in each direction. This results in a collection of non-intersecting integral curves. An appropriate clustering approach may facilitate the partitioning of this set into clusters of integral curves, each corresponding to a single fibre of the fibre process. This approach ignores the hierarchical Bayesian structure of the model and provides limited inference, however, it may be useful in identifying an appropriate initial state for slow-mixing MCMC algorithms.

8.3 Summary

In conclusion, this thesis has introduced a new model for fibre-generated point processes. Fibres are modelled as integral curves of a field of orientations, and as such any smooth, non-intersecting collection of fibres may arise from the random fibre process. A birth-death MCMC algorithm has been designed, allowing samples to be drawn from the posterior distribution of fibres given an instance of the point process. From these samples, numerous properties of the fibres and other parameters can be estimated. Furthermore, our approach is flexible to different types of fibre-generated point processes, as well as being readily extended to higher dimensions.

As discussed in this final chapter, this work opens up a number of interesting and challenging areas for further research.

Appendix A

Table of Notation

Table A.1: Table of notation used throughout thesis

Symbol	Usage	Page
Π	Random point process.	28
W	Planar window.	28
$\mathbb{R}^2$	Two-dimensional Euclidean space.	28
$\mathbf{y} = y_1, ..., y_m$	Data (locations of points in W).	28
m	Number of data points.	28
Υ_{FO}	Random field of orientations.	30
υ_{FO}	Instance of a random field of orientations.	30, 5
F or $F_j = F_j(\omega_j, l_j, \upsilon_{\mathrm{FO}})$	Instance of a random fibre.	30
$\boldsymbol{\omega} = \{\omega_1, ..., \omega_k\}$	Reference points of fibres.	30
$\mathbf{F} = \{F_1, ..., F_k\}$	Set of fibres.	31
k	Number of fibres.	31
$l_j = (l_{j,1}, l_{j,2})$	Arc lengths from the 2 end points of a fibre to its reference point.	31
$\mathbf{l} = l_1, ..., l_k$	Set of fibre length pairs.	32
$l_{j,T} = l_{j,1} + l_{j,2}$	Total length of a fibre.	32
$F_j \subset W$	Fibre that lies completely within the window.	33
$\mathbf{p} = p_1, ..., p_m$	Anchor points.	34

Symbol	Usage	Page
σ_{disp}	Parameter governing the perturbation of data points from anchor points.	34
$\mathbf{I}_n$	$n \times n$ identity matrix.	34
$\text{MVN}(\mu, \sigma^2\mathbf{I}_n)$	Multivariate normal distribution with mean μ and covariance matrix $\sigma^2\mathbf{I}_n$.	34
$\mathbf{X} = X_1, ..., X_m$	Indicator variable of which fibre a point is allocated to.	34
$x \in F$	Point x lies on fibre F.	34
$\mathbf{Z} = Z_1, ..., Z_m$	Indicator variable for noise/signal allocation of points.	34
$\boldsymbol{\varepsilon} = \varepsilon_1, ..., \varepsilon_m$	Probability a point is signal.	34
μ_{total}	Mean total number of points.	34
μ_{signal}	Mean number of signal points.	34
ρ	Hyperparameter indicating proportion of points that are noise points (also used as general intensity parameter).	34
η	Density of signal points (per unit length of fibre).	35
κ	Poisson parameter for prior on number of fibres, k.	36
λ	Exponential rate for prior on fibre lengths $l_{1,j}, l_{2,j}$.	36
$\mathbf{1}_{[...]}$	Indicator function.	36
$\alpha_{\text{signal}}, \beta_{\text{signal}}$	Hyperparameters of the probability a point is signal, ε_i.	36
α_{Dir}	Parameter for Dirichlet distribution of of proportion of fibre arc length between anchor points.	36
$D(\mathbf{q}(\mathbf{p}), \alpha_{\text{Dir}})$	Component of Dirichlet prior on anchor points.	36
$L(\ldots \mid \mathbf{y})$	Likelihood function.	37
$\text{dist}(x, y)$	Euclidean distance between point locations x and y.	37

Symbol	Usage	Page
$\|W\|$	Lebesgue measure of window W.	37
$v^i = (v_1^i, v_2^i)$	The Euclidean vector from initial point y_j to terminal point y_i.	46
$\tilde{v}^i = (\tilde{v}_1^i, \tilde{v}_2^i)$	The resultant vector of a Gaussian transformation on the length of v^i.	46
σ_{FO} or σ	Scaling parameter in Gaussian transformation of the length of v^i.	46
$T_0(y_j)$	Initial tensor calculated at point y_j using the tensor method.	46
$T_{h_{\mathrm{FO}}}(x)$	Tensor field calculated at $x \in W$ by applying Gaussian kernel smoothing with parameter h_{FO} to initial tensors.	52
h_{FO} or h	Tensor field smoothing parameter.	52
$f_{h_{\mathrm{FO}}}(t)$	Gaussian kernel in tensor smoothing.	52
$r_{\boldsymbol{\varepsilon}}, r_{\mathbf{l}}, r_{\boldsymbol{\omega}}, r_{\mathbf{Z}}, r_{\mathrm{S}}, r_{\mathrm{J}}, r_{\mathrm{o}}$	Rates of moves.	73
β	Birth rate of fibres.	73
$\boldsymbol{\delta} = \delta_1, ..., \delta_k$	Death rates of k fibres.	74
t	Algorithmic time.	74
$b(F_j, \omega_j, l_j, \mathbf{Z}, \mathbf{X}, \mathbf{p})$	Birth density.	74
$Q_{\mathrm{sig,birth}}(\mathbf{Z}'\|\cdot)$, $Q_{\mathrm{sig,death}}(\mathbf{Z}'\|\cdot)$	Proposal density of signal/noise allocations following a birth or death.	75
$Q_{\mathrm{aux,birth}}(\mathbf{X}', \mathbf{p}'\|\cdot)$, $Q_{\mathrm{aux,death}}(\mathbf{X}', \mathbf{p}'\|\cdot)$	Proposal density of auxiliary variables $\mathbf{X}'$ and $\mathbf{p}'$ following a birth of death.	75
l_T	Total length of all fibres.	77
$Q_{\mathrm{aux}}(\mathbf{X_i}', \mathbf{p_i}'\|\cdot)$	Proposal density of auxiliary variables X_i' and p_i' following a general move.	78
$N(y_i, \mathbf{F}', \sigma^2_{\mathrm{disp}}, Z_i' = 1)$	Normalising constant in proposal of $\mathbf{X}$, $\mathbf{p}$.	78
ϕ_{μ,σ^2} or $\phi_{\mu,\sigma^2\mathbf{I}_2}$	Univariate or bivariate normal density function with mean μ and variance σ^2.	85
$\tau(x, y)$	Intensity of cosine Poisson process.	131
T	Tensor calculated through the tensor method.	132

Symbol	Usage	Page
ϕ	Orientation of tensor.	132
w_i	Weight of contribution of the i-th data point to tensor method calculation.	133
$\lambda_1, \lambda_2, ...$	Eigenvalues of a tensor.	134
$msFA$	Modified square Fractional Anisotropy measure.	137

Appendix B

Proofs of Theorems on the Extent of Curvature Bias

B.1 Bias Calculation - Arch Model

Theorem 1 (Singularity Location Bias: Arch Model)**.** *Let $T_0^{arch} : \mathbb{R}^2 \to [0, \pi)$ be a tensor field with constant eigenvalues $\lambda_1 > \lambda_2 > 0$, and principal eigenvectors that agree with the arch model pictured in Figure 4.9. Explicitly, using a Cartesian coordinate system $(x_1, x_2) \in \mathbb{R}^2$ with origin $(0, 0)$ at the singularity, eigenvectors are tangent to a circle centred at $(0, 0)$ if $x_2 > 0$ and equal to $(0, 1)$ if $x_2 \leq 0$. Denote by T_h^{arch} the result of applying a kernel smoothing in the log-Euclidean metric to T_0^{arch}, using a Gaussian kernel with parameter h. Then T_h^{arch} has a singularity at $(x_1, x_2) = (0, hc)$. The constant c is the solution of*

$$\begin{aligned}
&\int_0^c \left(\int_{\cos^{-1}\left(\frac{c}{r'_x}\right)-\frac{\pi}{2}}^{\frac{3\pi}{2}-\cos^{-1}\left(\frac{c}{r'_x}\right)} \cos^2\theta_0 - \sin^2\theta_0 \; d\theta_x \right. \\
&\left. + \, 2\cos^{-1}\left(\frac{c}{r'_x}\right) \right) \frac{\exp(-(r'_x)^2/2)}{2\pi} \, dr'_x \\
&+ \int_c^\infty \int_0^{2\pi} \cos^2\theta_0 - \sin^2\theta_0 \; d\theta_x \frac{\exp(-(r'_x)^2/2)}{2\pi} \, dr'_x = 0
\end{aligned} \tag{B.1}$$

where

$$\theta_0 = \theta_0(r_x, \theta_x) = \tan^{-1}\left(\tan\theta_x + \frac{\beta}{r\cos\theta_x}\right). \tag{B.2}$$

Proof. A singularity in tensor field T_h^{arch} is defined to be a point $x \in \mathbb{R}^2$ such that the two eigenvalues of $T_h^{\text{arch}}(x)$ are equal, or equivalently the two eigenvalues

of $\log(T_h^{\text{arch}}(x))$ are equal. As the tensor field has constant eigenvalues and the principal eigenvector field is symmetric about $x_1 = 0$ the location of the singularity must lie on the line $x_1 = 0$, therefore we need only evaluate the tensor field for $(0, x_2)$.

In polar coordinates (r_x, θ_x) centred at $x = (0, x_2)$, the tensor field evaluated at x, $T_h^{\text{arch}}(x)$, is defined by

$$\log(T_h^{\text{arch}}(x)) \tag{B.3}$$
$$= \int_0^{\infty} \left(\int_0^{2\pi} \log\left(T_0^{\text{arch}}(r_x \cos\theta_x, x_2 + r_x \sin\theta_x) \right) \mathrm{d}\theta_x \right) r_x \frac{\exp(-r_x^2/2h^2)}{2\pi h^2} \mathrm{d}r_x.$$

Note that the inner integral over a circle of radius r_x centred at $x = (0, x_2)$ is independent of the Gaussian weights.

Consider a second polar coordinate representation, $(r_0, \theta_0) = (r_0(r_x, \theta_x), \theta_0(r_x, \theta_x))$, the polar coordinates centred at the origin. The two polar coordinate systems are related by the following pair of simultaneous equations,

$$r_x \cos\theta_x = r_0 \cos\theta_0$$

$$r_x \sin\theta_x + x_2 = r_0 \sin\theta_0.$$

Then the initial tensor field $T_0^{\text{arch}}(r_0, \theta_0)$ is written

$$T_0^{\text{arch}}(r_0, \theta_0) = \begin{pmatrix} \lambda_2 \cos^2\theta_0 + \lambda_1 \sin^2\theta_0 & (\lambda_2 - \lambda_1)\sin\theta_0\cos\theta_0 \\ (\lambda_2 - \lambda_1)\sin\theta_0\cos\theta_0 & \lambda_2 \sin^2\theta_0 + \lambda_1 \cos^2\theta_0 \end{pmatrix} \text{ for } \theta_0 \in [0, \pi] \tag{B.4}$$

and

$$T_0^{\text{arch}}(r_0, \theta_0) = \begin{pmatrix} \lambda_2 & 0 \\ 0 & \lambda_1 \end{pmatrix} \text{ for } \theta_0 \in [\pi, 2\pi]. \tag{B.5}$$

Writing r_0 and θ_0 in terms of r_x and θ_x we have

$$r_0(r_x, \theta_x) = \sqrt{r_x^2 + 2r_x x_2 \sin\theta_x + x_2^2} \quad \text{and} \tag{B.6}$$

$$\theta_0(r_x, \theta_x) = \tan^{-1}\left(\tan\theta_x + \frac{x_2}{r\cos\theta_x} \right). \tag{B.7}$$

Recall that to calculate the logarithm of a tensor we preserve the eigenvectors, but take the logarithm of the eigenvalues. Expanding the inner (bracketed) integral of

equation (B.3) we get

$$\int_0^{2\pi} \log\left(T_0^{\text{arch}}(r_x\cos\theta_x, x_2 + r_x\sin\theta_x)\right)\,\mathrm{d}\theta_x =$$

$$\int_{\cos^{-1}\left(\frac{x_2}{r_x}\right)-\frac{\pi}{2}}^{\frac{3\pi}{2}-\cos^{-1}\left(\frac{x_2}{r_x}\right)} \begin{pmatrix} \log\lambda_2\cos^2\theta_0 + \log\lambda_1\sin^2\theta_0 & \log\left(\frac{\lambda_2}{\lambda_1}\right)\sin\theta_0\cos\theta_0 \\ \log\left(\frac{\lambda_2}{\lambda_1}\right)\sin\theta_0\cos\theta_0 & \log\lambda_2\sin^2\theta_0 + \log\lambda_1\cos^2\theta_0 \end{pmatrix}\mathrm{d}\theta_x$$

$$+\int_{\frac{3\pi}{2}-\cos^{-1}\left(\frac{x_2}{r_x}\right)}^{\frac{3\pi}{2}+\cos^{-1}\left(\frac{x_2}{r_x}\right)} \begin{pmatrix} \log\lambda_2 & 0 \\ 0 & \log\lambda_1 \end{pmatrix}\mathrm{d}\theta_x \tag{B.8}$$

if $r_x \geq x_2$ (i.e. the circle of radius r_x centred at $(0, x_2)$ intersects the horizontal axis), and

$$\int_0^{2\pi} \log\left(T_0^{\text{arch}}(r_x\cos\theta_x, x_2 + r_x\sin\theta_x)\right)\,\mathrm{d}\theta_x =$$

$$\int_0^{2\pi} \begin{pmatrix} \log\lambda_2\cos^2\theta_0 + \log\lambda_1\sin^2\theta_0 & \log\left(\frac{\lambda_2}{\lambda_1}\right)\sin\theta_0\cos\theta_0 \\ \log\left(\frac{\lambda_2}{\lambda_1}\right)\sin\theta_0\cos\theta_0 & \log\lambda_2\sin^2\theta_0 + \log\lambda_1\cos^2\theta_0 \end{pmatrix}\mathrm{d}\theta_x \tag{B.9}$$

if $r_x < x_2$.

The second term of equation (B.8) is easily integrated:

$$\int_{\frac{3\pi}{2}-\cos^{-1}\left(\frac{x_2}{r_x}\right)}^{\frac{3\pi}{2}+\cos^{-1}\left(\frac{x_2}{r_x}\right)} \begin{pmatrix} \log\lambda_2 & 0 \\ 0 & \log\lambda_1 \end{pmatrix}\mathrm{d}\theta_x = 2\cos^{-1}\left(\frac{x_2}{r_x}\right)\begin{pmatrix} \log\lambda_2 & 0 \\ 0 & \log\lambda_1 \end{pmatrix}. \tag{B.10}$$

A singularity at $(0, x_2)$ corresponds to a tensor $T_h^{\text{arch}}(0, x_2)$ with equal eigenvalues. By consideration of the eigendecomposition of a 2×2 positive definite matrix it is clear that the tensor has equal eigenvalues if and only the diagonal elements are equal and the off-diagonal elements are 0.

The off-diagonal elements evaluate to 0 for all points $x = (0, x_2)$. Equating the

difference of the diagonal elements of $T_h^{\text{arch}}(x)$ to 0 implies that x_2 must satisfy

$$0 = \int_0^{x_2} \left(\int_{\cos^{-1}\left(\frac{x_2}{r_x}\right)-\frac{\pi}{2}}^{\frac{3\pi}{2}-\cos^{-1}\left(\frac{x_2}{r_x}\right)} \log\left(\frac{\lambda_2}{\lambda_1}\right) cos^2\theta_0 + \log\left(\frac{\lambda_1}{\lambda_2}\right) \sin^2\theta_0 \; \mathrm{d}\theta_x \right.$$
$$\left. + \, 2\cos^{-1}\left(\frac{x_2}{r_x}\right) \log\left(\frac{\lambda_2}{\lambda_1}\right) \right) r_x \frac{\exp(-r_x^2/2h^2)}{2\pi h^2} \, \mathrm{d}r_x$$
$$+ \int_{x_2}^{\infty} \left(\int_0^{2\pi} \log\frac{\lambda_2}{\lambda_1} \cos^2\theta_0 + \log\frac{\lambda_1}{\lambda_2} \sin^2\theta_0 \; \mathrm{d}\theta_x \right) r_x \frac{\exp(-r_x^2/2h^2)}{2\pi h^2} \, \mathrm{d}r_x. \quad \text{(B.11)}$$

Finally, dividing Equation (B.11) by $\log\frac{\lambda_2}{\lambda_1}$, substituting $r'_x = r_x/h$ and setting $c = x_2/h$ gives

$$0 = \int_0^{c} \left(\int_{\cos^{-1}\left(\frac{c}{r'_x}\right)-\frac{\pi}{2}}^{\frac{3\pi}{2}-\cos^{-1}\left(\frac{c}{r'_x}\right)} \cos^2\theta_0 - \sin^2\theta_0 \; \mathrm{d}\theta_x + 2\cos^{-1}\left(\frac{c}{r'_x}\right) \right) \frac{\exp(-(r'_x)^2/2)}{2\pi} \, \mathrm{d}r'_x$$
$$+ \int_c^{\infty} \left(\int_0^{2\pi} \cos^2\theta_0 - \sin^2\theta_0 \; \mathrm{d}\theta_x \right) \frac{\exp(-(r'_x)^2/2)}{2\pi} \, \mathrm{d}r'_x. \quad \text{(B.12)}$$

□

B.2 Bias Calculation - Parabolic Model

The proof of the lemma follows from the definition of a parabolic field.

Lemma 1. *Let $e(x_1, x_2) = (e_1(x_1, x_2), e_2(x_1, x_2))$ denote the principal eigenvector of the tensor $T_0^{para}(x_1, x_2)$. Then $e(x_1, x_2)$ is perpendicular to $e(-x_1, -x_2)$ for all $(x_1, x_2) \neq (0, 0)$.*

Proof. We begin by noting that, by construction, for $x_1 = 0$ the desired property holds.

Otherwise, for $x_1 \neq 0$, it suffices to show that $e_2(x_1, x_2)/e_1(x_1, x_2)$ is equal to $-e_1(-x_1, -x_2)/e_2(-x_1, -x_2)$. Recall that $(e_1(x_1, x_2), e_2(x_1, x_2))$ is tangent to the parabola $x_2 = \frac{1}{4a} - ax_1^2$, and is therefore proportional to $(1, -2ax_1)$. The constant a is determined by

$$a = \frac{-x_2 + \sqrt{x_1^2 + x_2^2}}{2x_1^2}, \quad \text{(B.13)}$$

so the result follows:

$$\begin{aligned} e_2(x_1,x_2)/e_1(x_1,x_2) &= \frac{x_2-\sqrt{x_2^2+x_1^2}}{x_1} \\ &= \frac{-x_1\left(x_2-\sqrt{x_2^2+x_1^2}\right)}{-x_1^2} \\ &= \frac{-x_1}{x_2+\sqrt{x_2^2+x_1^2}} \\ &= -\left(\frac{-x_2-\sqrt{x_2^2+x_1^2}}{-x_1}\right)^{-1} \\ &= -e_1(-x_1,-x_2)/e_2(-x_1,-x_2). \end{aligned} \tag{B.14}$$

Equality of the second and third lines can be seen by multiplying both numerator and denominator by $\left(x_2+\sqrt{x_2^2+x_1^2}\right)$. □

Theorem 2 (Singularity Location Bias: Parabolic Model)**.** *Let $T_0^{para}(x_1,x_2)$ be a parabolic tensor field with constant eigenvalues $\lambda_1 > \lambda_2$, and let $T_h^{para}(x_1,x_2)$ be the result of applying a convolution in the log-Euclidean metric to this tensor field with a Gaussian kernel and smoothing parameter h. Then $T_h^{para}(x_1,x_2)$ contains a singularity located at the origin $(0,0)$.*

Proof. Consider the convolution of this tensor field in the log-Euclidean metric evaluated at the origin in polar coordinates (centred at the origin),

$$\log(T_h^{\text{para}}(x_1,x_2)) = \int_0^\infty \left(\int_0^{2\pi} \log\left(T_0^{\text{para}}(r\cos\theta, r\sin\theta)\right)\ \mathrm{d}\theta\right) r\frac{\exp(-r^2/2h^2)}{2\pi h^2}\,\mathrm{d}r. \tag{B.15}$$

The inner integral, over a circle of radius r, evaluates to a multiple of the identity matrix. We show this by considering the tensor field T_0^{para} evaluated at antipodal points on the circle and proving that, as the eigenvectors of these tensors are perpendicular, the terms will cancel in the integration.

The inner integral is

$$\begin{aligned} &\int_0^{2\pi} \log\left(T_0^{\text{para}}(r\cos\theta, r\sin\theta)\right)\ \mathrm{d}\theta \\ &= \int_0^{\pi} \log\left(T_0^{\text{para}}(r\cos\theta, r\sin\theta)\right) + \log\left(T_0^{\text{para}}(r\cos(\theta+\pi), r\sin(\theta+\pi))\right)\ \mathrm{d}\theta. \end{aligned} \tag{B.16}$$

Suppose the tensor $T(r\cos\theta, r\sin\theta)$ has principal eigenvector $(\cos\phi, \sin\phi)$, then

$$\log(T(r\cos\theta, r\sin\theta)) \tag{B.17}$$

$$= \begin{pmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{pmatrix} \begin{pmatrix} \log(\lambda_1) & 0 \\ 0 & \log(\lambda_2) \end{pmatrix} \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix}$$

$$= \begin{pmatrix} \log(\lambda_1)\cos^2\phi + \log(\lambda_2)\sin^2\phi & (\log(\lambda_2) - \log(\lambda_1))\cos\phi\sin\phi \\ (\log(\lambda_2) - \log(\lambda_1))\cos\phi\sin\phi & \log(\lambda_1)\sin^2\phi + \log(\lambda_2)\cos^2\phi \end{pmatrix}. \tag{B.18}$$

By Lemma 1, the principal eigenvector of $T(r\cos(\theta+\pi), r\sin(\theta+\pi))$ is perpendicular to $(\cos\phi, \sin\phi)$ and hence is equal to $(-\sin\phi, \cos\phi)$. So

$$\log(T(r\cos(\theta+\pi), r\sin(\theta+\pi)))$$

$$= \begin{pmatrix} \log(\lambda_1)\sin^2\phi + \log(\lambda_2)\cos^2\phi & (\log(\lambda_1) - \log(\lambda_2))\cos\phi\sin^2\phi \\ (\log(\lambda_1) - \log(\lambda_2))\cos\phi\sin\phi & \log(\lambda_1)\cos\phi + \log(\lambda_2)\sin^2\phi \end{pmatrix} \tag{B.19}$$

Substituting the tensor logarithms of Equations (B.18) and (B.19) into Equation (B.16), we get

$$\int_0^\pi (\log(\lambda_1) + \log(\lambda_2))\mathbf{1}_2 \; \mathrm{d}\theta = \pi(\log(\lambda_1) + \log(\lambda_2))\mathbf{1}_2 \tag{B.20}$$

where $\mathbf{1}_2$ is the identity matrix. This is independent of r so Equation (B.15) evaluates to

$$\log(\tilde{S}_0) = c\pi(\log(\lambda_1) + \log(\lambda_2))\mathbf{1}_2 \tag{B.21}$$

for some constant c. The eigenvalues of $\log(T_0^{\text{para}}(0,0))$ are equal and therefore the eigenvalues of $(T_0^{\text{para}}(0,0))$ are equal, indicating that there is a singularity at $(0,0)$ in the tensor field T_h^{para}. □

Bibliography

R. J. Adler and J. E. Taylor. *Random Fields and Geometry*. Springer, Boston, 2007.

D. Allard and C. Fraley. Nonparametric Maximum Likelihood Estimation of Features in Spatial Point Processes Using Voronoi Tessellation. *Journal of the American Statistical Association*, 92(440):1485–1493, 1997.

E. Arias-Castro, D. L. Donoho, X. Huo, and C. A. Tovey. Connect the Dots: How Many Random Points can a Regular Curve Pass Through? *Advances in Applied Probability*, 37:571–605, 2005.

V. Arsigny, P. Fillard, X. Pennec, and N. Ayache. Log-Euclidean Metrics for Fast and Simple Calculus on Diffusion Tensors. *Magnetic Resonance in Medicine*, 56 (2):411–421, 2006.

D. R. Ashbaugh. *Quantitative-Qualitative Friction Ridge Analysis: An Introduction to Basic and Advanced Ridgeology*. CRC Press, 1999.

A. J. Baddeley, R. A. Moyeed, C. V. Howard, and A. Boyde. Analysis of a Three-Dimensional Point Pattern with Replication. *Journal of the Royal Statistical Society: Series C*, 42(4):641–668, 1993.

M. M. Bahn. Comparison of Scalar Measures Used in Magnetic Resonance Diffusion Tensor Imaging. *Journal of Magnetic Resonance*, 139:1–7, 1999.

J. D. Barrow, S. P. Bhavsar, and D. H. Sonoda. Minimal Spanning Trees, Filaments and Galaxy Clustering. *Royal Astronomical Society, Monthly Notices*, 216:17–35, 1985.

P. J. Basser and C. Pierpaoli. Microstructural and Physiological Features of Tissues Elucidated by Quantitative-Diffusion-Tensor MRI. *Journal of Magnetic Resonance, Series B*, 111(86):209–219, 1996.

P. J. Basser, J. Mattiello, and D. LeBihan. MR Diffusion Tensor Spectroscopy and Imaging. *Biophysical Journal*, 66(1):259–267, 1994.

P. J. Basser, S. Pajevic, C. Pierpaoli, J. Duda, and A. Aldroubi. In Vivo Fiber Tractography Using DT-MRI Data. *Magnetic Resonance in Medicine*, 44:625–632, 2000.

P. G. Batchelor, M. Moakher, D. Atkinson, F. Calamante, and A. Connelly. A Rigorous Framework for Diffusion Tensor Calculus. *Magnetic Resonance in Medicine*, 53:221–225, 2005.

T. Bayes. An Essay Towards Solving a Problem in the Doctrine of Chances. *Philosophical Transactions*, 53:370–418, 1763.

P. Brémaud. *Point Processes and Queues: Martingale Dynamics*. Springer Series in Statistics. Springer, 1981.

A. Brix and W. S. Kendall. Simulation of Cluster Point Processes Without Edge Effects. *Advances in Applied Probability*, 34(2):267–280, 2002.

S. Brooks and P. Giudici. Convergence Assessment for Reversible Jump MCMC Simulations. *Bayesian Statistics*, 6, 1998.

S. Brooks, A. Gelman, G. L. Jones, and X-L. Meng. *Handbook of Markov Chain Monte Carlo*. Handbooks of Modern Statistical Methods. Chapman & Hall/CRC, 2011.

S. P. Brooks, P. Giudici, and G. O. Roberts. Efficient Construction of Reversible Jump Markov Chain Monte Carlo Proposal Distributions. *Journal of the Royal Statistical Society: Series B*, 65(1):3–55, 2003.

S. Byers and A. E. Raftery. Nearest Neighbor Clutter Removal for Estimating Features in Spatial Point Processes. *American Statistical Association*, 93(442): 577–584, 1998.

O. Cappé, C. P. Robert, and T. Rydén. Reversible Jump MCMC Converging to Birth-and-Death MCMC and More General Continuous Time Samplers. *Journal of the Royal Statistical Society: Series B*, 65:679–700, 2003.

B. P. Carlin and T. A. Louis. *Bayesian Methods for Data Analysis*. Chapman & Hall/CRC, 2008.

J. M. Castelloe and D. L. Zimmerman. Convergence Assessment for Reversible Jump MCMC Samplers. Technical Report 313, Department of Actuarial Sciences, University of Iowa, 2002.

G. Celeux, D. Chauveau, and J. Diebolt. On Stochastic Versions of the EM Algorithm. Technical Report 2514, INRIA, 1995.

S. Chanraud, N. Zahr, E. V. Sullivan, and A. Pfefferbaum. MR Diffusion Tensor Imaging: A Window into White Matter Integrity of the Working Brain. *Neuropsychology Review*, 20(2):209–225, 2010.

M. Colless, G. Dalton, S. Maddox, W. Sutherland, C. Collins, W. Couch, K. Deeley, R. De Propris, G. Efstathiou, R. S. Ellis, C. S. Frenk, K. Glazebrook, C. Jackson, O. Lahav, D. Madgwick, J. A. Peacock, B. A. Peterson, M. Seaborne, K. Taylor, I. J. Lewis, S. L. Lumsden, and I. A. Price. The 2dF Galaxy Redshift Survey: Spectra and Redshifts. *Monthly Notices of the Royal Astronomical Society*, 328: 1039, 2001.

N. Cressie and L. B. Collins. Patterns in Spatial Point Locations: Local Indicators of Spatial Association in a Minefield with Clutter. *Naval Research Logistics*, 48 (5):333–347, 2001.

A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society: Series B*, 39(1):1–38, 1977.

R. Deriche and M. Descoteaux. Splitting Tracking through Crossing Fibers: Multidirectional Q-Ball Tracking. In *4th IEEE Symposium on Biomedical Imaging: From Nano to Macro*, pages 756–759, 2007.

P. J. Diggle. *Statistical Analysis of Spatial Point Patterns*. Academic Press, 1983.

I.L. Dryden, A. Koloydenko, and D Zhou. Non-Euclidean Statistics for Covariance Matrices, with Applications to Diffusion Tensor Imaging. *The Annals of Applied Statistics*, 3(3):1102–1123, 2009.

B. S. Everitt, S. Landau, M. Leese, and D. Stahl. *Cluster Analysis*, volume 848 of *Wiley Series in Probability and Statistics*. John Wiley and Sons, Ltd, Fourth edition, 2011.

P.T. Fletcher and S. Joshi. Riemannian Geometry for the Statistical Analysis of Diffusion Tensor Data. *Signal Processes*, 87(2):250–262, 2007.

C. Fraley and A. E. Raftery. How Many Clusters? Which Clustering Method? Answers Via Model-Based Cluster Analysis. *The Computer Journal*, 41(8):578–588, 1998.

A. Gelman. Objections to Bayesian Statistics. *Bayesian Analysis*, 3(3):445–450, 2008.

A. Gelman and D. B. Rubin. Inference from Iterative Simulation using Multiple Sequences. *Statistical Science*, 7:457–511, 1992.

C. R. Genovese, M. Perone-Pacifico, I. Verdinelli, and L. Wasserman. On the Path Density of a Gradient Field. *Annals of Statistics*, 67, 2009.

J. Geweke. Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments. Research Department Staff Report 148, Federal Reserve Bank of Minneapolis, 1991.

W. R. Gilks and G. O. Roberts. Strategies for Improving MCMC. In W. R. Gilks, S. Richardson, and D. Spiegelhalter, editors, *Markov Chain Monte Carlo in Practice: Interdisciplinary Statistics (Chapman & Hall/CRC Interdisciplinary Statistics Series)*, chapter 6. Chapman & Hall/CRC, 1995.

W. R. Gilks, S. Richardson, and D. Spiegelhalter. *Markov Chain Monte Carlo in Practice: Interdisciplinary Statistics (Chapman & Hall/CRC Interdisciplinary Statistics Series)*. Chapman & Hall/CRC, First edition, 1995.

L. Glass. Moiré Effect from Random Dots. *Nature*, 223:578–580, 1969.

M. Goldstein. Subjective Bayesian Analysis: Principles and Practice. *Bayesian Analysis*, 1(3):403–420, 2006.

P. J. Green. Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination. *Biometrika*, 82(2):711–32, 1995.

Y. Guan, M. Sherman, and J. A. Calvin. Assessing Isotropy for Spatial Point Patterns. *Biometrics*, 62:119–125, 2006.

T. Hastie and W. Stuetzle. Principal Curves. *Journal of the American Statistical Association*, 84(406):502–516, 1989.

A. R. Hildebrand, M. Pilkington, M. Connors, C. Ortiz-Aleman, and R. E. Chavez. Size and Structure of the Chicxulub Crater Revealed by Horizontal Gravity Gradients and Cenotes. *Nature*, 376:415–417, 2002.

B. J. Hill, W. S. Kendall, and E. Thönnes. Fibre-Generated Point Processes and Fields of Orientations (submitted), 2011.

C. Howson and P. Urbach. Bayesian Reasoning in Science. *Nature*, 350:371–374, 1991.

M. Huber. Spatial Point Processes. In S. Brooks, A. Gelman, G. L. Jones, and X-L. Meng, editors, *Handbook of Markov Chain Monte Carlo*, chapter 9. Chapman & Hall/CRC, 2011.

J. Illian, A. Penttinen, H. Stoyan, and D. Stoyan. *Statistical Analysis and Modelling of Spatial Point Patterns*. Statistics in Practice. Wiley, 2008.

M. C. Irwin. *Smooth Dynamical Systems*. Pure and Applied Mathematics. Academic Press, 1980.

A. Jasra, C. C. Holmes, and D. A. Stephens. MCMC and the Label Switching Problem in Bayesian Mixture Models . *Statistical Science*, 20:50–67, 2005.

M. Kass and A. Witkin. Analyzing Oriented Patterns. *Computer Vision, Graphics and Image Processing*, 37(3):362–385, 1987.

M. Lazar, D. M. Weinstein, J. S. Tsuruda, K. M. Hasan, K. Arfanakis, M. E. Meyerand, B. Badie, H. A. Rowley, V. Haughton, A. Field, and A. L. Alexander. White Matter Tractography using Diffusion Tensor Deflection. *Human Brain Mapping*, 18(4):306–321, 2003.

S. C. Lee, D. Y. Yang, J. Ko, and J. R. Youn. Effect of Compressibility on Flow Field and Fiber Orientation During the Filling Stage of Injection Molding. *Journal of Materials Processing Technology*, 70:83–92, 1997.

J. Lund, A. Penttinen, and M. Rudemo. Bayesian Analysis of Spatial Point Patterns From Noisy Observations. In *Report, Department of Mathematics and Physics, The Royal Veterinary and Agricultural University.*, 1999.

P. K. H. Maguire, G. D. Mackenzie, P. Denton, A. Trejo, R. Kind, and Members of the Chicxulub Working Group. Preliminary Results from a Passive Seismic Array over the Chicxulub Impact Structure in Mexico. *Geological Society, London, Special Publications*, 140:177–193, 1998.

D. Maltoni, A. K. Jain, D. Maio, and S. Prabhakar. *Handbook of Fingerprint Recognition*. Springer Professional Computing. Springer, 2003.

J. Mardia, A. J. Baczkowski, X. Feng, and T. J. Hainsworth. Statistical Methods for Automatic Interpretation of Digitally Scanned Finger Prints. *Pattern Recognition Letters*, 18:1197–1203, 1997.

K. V. Mardia and P. E. Jupp. *Directional Statistics*. Wiley Series in Probability and Statistics. John Wiley and Sons, Ltd, 1999.

J. S. Maritz and T. Lwin. *Empirical Bayes Methods*. Monographs on Statistics and Applied Probability 35. Chapman and Hall, 1989.

F. H. C. Marriott. *The Interpretation of Multiple Observations*. Academic Press, 1974.

V.J. Martínez and E. Saar. *Statistics of the Galaxy Distribution*. Chapman & Hall/CRC, 2002.

G. J. McLachlan and D. Peel. *Finite Mixture Models*. Wiley Series in Probability and Statistics. John Wiley and Sons, Ltd, 2000.

R. E. Miles. On the Elimination of Edge Effects in Planar Sampling. In *Stochastic Geometry: A Tribute to the Memory of Rollo Davidson*. John Wiley and Sons, Ltd, 1974.

G. M. Miller. *Simultaneous Statistical Inference*. Springer Series in Statistics. Springer-Verlag, 1980.

Z. Minier and L. Csató. Kernel pca based clustering for inducing features in text categorization. In *European Symposium on Artificial Neural Networks Proceedings*, pages 349–354, 2007.

J. Møller and R. P. Waagepetersen. *Statistical Inference and Simulation for Spatial Point Processes*. Monographs on Statistics and Applied Probability 100. Chapman & Hall/CRC, 2004.

S. Mori, B. J. Crain, V. P. Chacko, and P. C. M. van Zijl. Three-Dimensional Tracking of Axonal Projections in the Brain by Magnetic Resonance Imaging. *Annals of Neurology*, 45(2):265–269, 2001.

C. N. Morris. Parametric Empirical Bayes Inference: Theory and Applications. *Journal of the American Statistical Association*, 78(381):47–55, 1983.

J. Neyman and E. L. Scott. Statistical Approach to Problems of Cosmology. *Journal of the Royal Statistical Society: Series B*, 20(1):1–43, 1958.

D. Novikov, S. Colombi, and O. Doré. Skeleton as a Probe of the Cosmic Web: the Two-Dimensional Case. *Monthly Notices of the Royal Astronomical Society*, 336 (4):1201–1216, 2008.

J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, 1988.

X. Pennec, P. Fillard, and N. Ayache. A Riemannian Framework for Tensor Computing. *International Journal of Computer Vision*, 66(1):41–66, 2006.

C. Pierpaoli and P. J. Basser. Toward a Quantitative Assessment of Diffusion Anisotropy. *Magnetic Resonance in Medicine*, 36(6):893–906, 1996.

C. J. Preston. Spatial Birth-and-Death Processes. *Bulletin of the International Statistical Institute*, 46:371–391, 1977.

N. K. Ratha, S. Chen, and A. K. Kain. Adaptive Flow Orientation-Based Feature Extraction in Fingerprint Images. *Pattern Recognition*, 28(11):1657–1672, 1995.

N. K. Ratha, K. Karu, S. Chen, and A. K. Kain. A Real-Time Matching System for Large Fingerprint Databases. *Geological Society, London, Special Publications*, 140:177–193, 1998.

S. Richardson and P. J. Green. On Bayesian Analysis of Mixtures with an Unknown Number of Components. *Journal of the Royal Statistical Society: Series B*, 59(4): 731–792, 1997.

B. Ripley. *Spatial Satistics*. Wiley Series in Probability and Statistics. John Wiley and Sons, Ltd, 1981.

B. D. Ripley. Simulating Spatial Patterns: Dependent Samples from a Multivariate Density. *Journal of the Royal Statistical Society: Series C*, 28(1):109–112, 1979.

H. Robbins. An Empirical Bayes Approach to Statistics. In J. Neyman, editor, *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, pages 157–163. University of California Press, 1956.

H. Robbins. The Empirical Bayes Approach to Statistical Decision Problems. *The Annals of Mathematical Statistics*, 35(1):1–20, 1964.

K. Roeder and L. Wasserman. Contribution to the discussion of paper by Richardson and Green, 1997.

B. Schölkopf, A. Smola, and K-R. Müller. Kernel Principal Components Analysis. In W. Gerstner, A. Germond, M. Hasler, and J.-D. Nicoud, editors, *Artificial Neural Networks - ICANN'97*, volume 1327 of *Lecture Notes in Computer Science*, pages 583–588. Springer, 1997.

B. Sherlock and D. Monro. A Model for Interpreting Fingerprint Topology. *Pattern Recognition*, 26(7):1047–1095, 1993.

W. Shi, Y-F. Guo, and X. Xue. International joint conference on neural networks. In *European Symposium on Artificial Neural Networks Proceedings*, pages 2908–2913, 2009.

Ø. Skare, J. Møller, and E. B. V. Jensen. Bayesian Analysis of Spatial Point Processes in the Neighbourhood of Voronoi Networks. *Statistical Computing*, 17: 369–379, 2007.

D. C. Stanford and A. E. Raftery. Finding Curvilinear Features in Spatial Point Patterns: Principal Curve Clustering with Noise. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22:601–609, 2000.

M. Stephens. Contribution to the discussion of paper by Richardson and Green. *Journal of the Royal Statistical Society, Series B*, 59:731–792, 1997.

M. Stephens. Bayesian Analysis of Mixture Models with an Unknown Number of Components - an Alternative to Reversible Jump Methods. *The Annals of Statistics*, 28(1):40–74, 2000a.

M. Stephens. Dealing with Label Switching in Mixture Models. *Journal of the Royal Statistical Society: Series B*, 62(4):795–809, 2000b.

K. A. Stevens. Computation of Locally Parallel Structure. *Biological Cybernetics*, 29:19–28, 1978.

R. S. Stoica, V. J. Martìnez, and E. Saar. Detection of Cosmic Filaments Using the Candy Model. *Astronomy and Astrophysics*, 434:423–432, 2005.

R. S. Stoica, V. J. Martìnez, and E. Saar. A Three-Dimensional Object Point Process for Detection of Cosmic Filaments. *Applied Statistics*, 56(Part 4):459–477, 2007.

D. Stoyan, W. S. Kendall, and J. Mecke. *Stochastic Geometry and its Applications*. Wiley Series in Probability and Statistics. Wiley, Second edition, 1995.

J. Su. *A Tensor Approach to Fingerprint Analysis*. PhD thesis, University of Warwick, 2009.

J. Su, B. J. Hill, W. S. Kendall, and E. Thönnes. Inference for Point Processes with Unobserved One-Dimensional Reference Structure. CRiSM Working Paper No. 8-10, University of Warwick, 2008.

L. Vincent. *Stochastic Geometry: Likelihood and Computation*, chapter 6: Mathematical Morphology. Monographs on Statistics and Applied Probability 80. Chapman & Hall/CRC, 1999.

M. P. Wand and M. C. Jones. *Kernel Smoothing*, volume 60 of *Monographs on Statistics and Applied Probability*. Chapman & Hall/CRC, 1995.

C. Watson. NIST Special Database 30: Dual Resolution Images from Paired Fingerprint Cards. National Institute of Standards and Technology, Gaithersburg, 2001.

N. H. Witherspoon, J. H. Holloway, K. S. Davis, R. W. Miller, and A. C. Dubey. Coastal Battlefield Reconnaissance and Analysis Program for Minefield Detection. In A. C. Dubey, I. Cindrich, J. M. Ralston, and k. a. Rigano, editors, *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 2496 of *Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference*, pages 500–508, 1995.

W. Zhang and J. Deng. Topology-Driven Streamline Seeding for 2D Vector Field Visualisation. *Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics*, 2009.